

An ensemble approach for short-term load forecasting by extreme learning machine

Song Li^{a,*}, Lalit Goel^a and Peng Wang^a

^a School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore

* Corresponding author at: Blk S2-B7c-05, 50 Nanyang Avenue, Nanyang Technological University, 639798 Singapore. Tel: +65 98628450. E-mail: sli5@e.ntu.edu.sg.

Abstract

This paper proposes a novel ensemble method for short-term load forecasting based on wavelet transform, extreme learning machine (ELM) and partial least squares regression. In order to improve forecasting performance, a wavelet-based ensemble strategy is introduced into the forecasting model. The individual forecasters are derived from different combinations of mother wavelet and number of decomposition levels. For each sub-component from the wavelet decomposition, a parallel model consisting of 24 ELMs is invoked to predict the hourly load of the next day. The individual forecasts are then combined to form the ensemble forecast using the partial least squares regression method. Numerical results show that the proposed method can significantly improve forecasting performance.

Keywords: Ensemble method, extreme learning machine, partial least squares regression, short-term load forecasting, wavelet transform.

1. Introduction

Short-term load forecasting (STLF) is a basic requirement in the operation of power systems. The primary application of STLF is to provide load predictions for generation scheduling, such as unit commitment and economic dispatch [1]. For example, if the load demand is known in advance, we can operate the generators at the lowest possible cost. It is shown in [2] that a small increase in forecasting accuracy would save millions of dollars in operation costs. A second application of STLF is for power system security assessment [3]. The forecast results are essential in the detection of future conditions under which the system may be vulnerable. In addition, energy purchasing and bidding also require STLF [4, 5].

Various methods have been proposed for STLF in power systems. In the early stages, statistical methods such as regression models and time series methods [6, 7], etc., were extensively employed. Nowadays, artificial intelligence techniques such as artificial neural networks (ANN), expert systems and support vector machines [8-12] have been applied to solve the STLF problem. In addition, many hybrid forecasting models have also been developed, which can make use of the advantages of each technique involved. In [13], the forecasting problem was presented in state space form, where the model parameters were fine-tuned by a particle swarm optimizer (PSO). In [14], a hybrid forecasting model combining differential evolution (DE) and support vector regression (SVR) was proposed for load forecasting, where the DE algorithm was used to choose the appropriate parameters for SVR.

As an emerging class of ANN, extreme learning machine (ELM) has a fast learning speed for its iterative-free learning mechanism [15]. In ELM, the input weights and hidden biases are randomly initialized, while the output weights are

directly calculated by a least squares method. ELM has been successfully used in load and price forecasting [16, 17]. But there is still room for improvement. Like other ANN-based models, the random initialization of weight parameters may introduce inferior values, leading to unreliable result. Moreover, ELM usually contains hundreds of hidden neurons and the output weights are directly estimated from the training data. Therefore, ELM might suffer from the overtraining problem [18, 19].

In this paper, a neural network ensemble model is proposed to alleviate the above problems. An ensemble model consists of a number of ELMs, which are trained for the same load forecasting problem. It has been shown that ensemble models have strengths with respect to single models in terms of increased accuracy and robustness. The reason is that the diverse errors from individual ELMs can cancel out in the aggregating process [20]. Several ensemble methods have been developed for STLF [21-23]. For example, in [23], the regularized negative correlation learning method was used to enhance the prediction capability of the network ensemble.

Wavelet transform has been widely used to capture the inner load features and improve the forecast accuracy [24, 25]. However, there is no fixed criterion to select the wavelet parameters: mother wavelet and number of decomposition levels. In this paper, a novel ensemble strategy is introduced, which uses wavelet transform to create the collection of ELM-based predictors. The individual predictors are featured with different combinations of mother wavelet and number of decomposition levels. The rationale is that different wavelet parameters would generate different inputs for the ELM-based predictors, which can promote the diversity of ensemble. In such a way, wavelet parameter selection is avoided without sacrificing the forecast accuracy.

Simple averaging is usually used to combine the individual forecasts [22, 26]. But in practice, some predictors in the ensemble are more accurate than others. The ensemble output should be a weighted average of the individual outputs. In this paper, the weight factors are obtained by the partial least squares regression (PLSR) method. Moreover, PLSR can tackle the high degree of correlation between the individual forecasts and therefore generate accurate ensemble output [27].

This paper proposes a novel ensemble method for STLF, in which wavelet transform, extreme learning machine and partial least squares regression are integrated. The proposed wavelet-based ensemble strategy uses different wavelet specifications to generate different input features for the ELM-based predictors. For each sub-component from the wavelet transform, a parallel model of 24 ELMs is allocated to predict the hourly data of the next day. To establish an accurate ensemble forecast, PLSR is used to combine the outputs of individual predictors.

The main contributions of this paper are as follows:

- 1) A new hybrid forecast method is proposed to deal with the STLF problem in electric power systems, which can alleviate the overtraining and uncertainty problems and improve the forecasting accuracy.
- 2) A novel ensemble strategy based on the wavelet transform is employed in the proposed method. It can avoid the trivial process of wavelet parameter selection. Moreover, the complementary information contained in each set of wavelet parameters can be utilized to improve the forecasting accuracy.
- 3) Weighted averaging is used to combine the individual forecast outputs, which can consider the relative accuracy between them. PLSR is adopted to calculate the optimal weight factors.

Although this paper focuses on STLF, the proposed method can be easily extended to electricity price forecasting, which is another important topic in power systems. Note that the input variables should be altered accordingly.

The rest of this paper is organized as follows. Section 2 presents the relevant theories and describes the implementation of the proposed method. Section 3 provides the numerical results and comparisons with other well-established methods. Section 4 outlines the conclusions.

2. Proposed methodology

2.1 Extreme learning machine

ELM accomplishes the learning process in two steps. First, the input weights and hidden biases are initialized with random numbers. Second, the output weights are calculated through an inverse operation on the hidden layer output matrix. The idea of ELM lies in that the neural network learning is turned into a least squares problem, which can be easily solved with much less time [15].

Given a training set of N samples $(\mathbf{x}_i, \mathbf{d}_i)$, the single-hidden layer feedforward network (SLFN) in Fig. 1 can be modeled by

$$\sum_{j=1}^n \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{o}_i, \quad i = 1, \dots, N. \quad (1)$$

where \mathbf{x}_i is the input pattern, \mathbf{d}_i is the desired output, n is the number of hidden nodes, $g(\cdot)$ is the activation function, \mathbf{o}_i is the actual output, \mathbf{w}_j is the input weight, b_j is the hidden bias and β_j is the output weight.

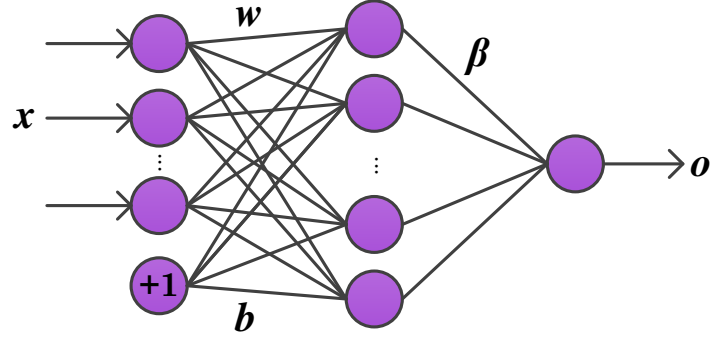


Fig. 1 Single-hidden layer feedforward network.

If the training error is zero, we can say that there exist β_j , w_j and b_j such that

$$\sum_{j=1}^n \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{d}_i, \quad i = 1, \dots, N. \quad (2)$$

The compact matrix form of (2) can be given by $\mathbf{H}\boldsymbol{\beta}=\mathbf{D}$, where $\boldsymbol{\beta}=[\beta_1, \dots, \beta_n]^T$, $\mathbf{D}=[\mathbf{d}_1, \dots, \mathbf{d}_N]^T$ and \mathbf{H} is called the hidden layer output matrix.

Since the number of hidden nodes is usually less than the number of training samples, \mathbf{H} is not square and ELM cannot approach the zero training error. To train the SLFN, ELM assigns uniformly distributed random numbers for the input weights w_j and hidden biases b_j . Then the network becomes an over-determined linear system and the output weights $\boldsymbol{\beta}$ can be determined by a least squares method. ELM provides a special solution by $\boldsymbol{\beta}^* = \mathbf{H}^\dagger \mathbf{D}$, where \mathbf{H}^\dagger is the Moore-Penrose (MP) generalized inverse of \mathbf{H} .

ELM presents many important properties, which make it an appealing learning algorithm for SLFN. First, ELM exhibits a fast learning speed as it determines the network parameters without iterative adjustments. Second, ELM can achieve good performance because the minimum training error $\|\mathbf{H}\boldsymbol{\beta}-\mathbf{D}\|$ is approached with a least

squares method. Third, ELM can avoid many problems faced by traditional learning methods, such as local minima, stopping criterion and learning rate.

2.2 Wavelet transform

Wavelet transform is able to provide local spectral and temporal information of a signal simultaneously, which is suitable for capturing the multiscale load features [28]. Given a mother wavelet $\psi(t)$, the load signal $s(t)$ can be represented by

$$s(t) = \sum_{j,k} W(j,k) 2^{j/2} \psi(2^j t - k) \quad (3)$$

where t is the time index, j and k are integer variables for scaling and translation, and $W(j,k)$ is called the discrete wavelet transform (DWT).

The multiresolution analysis is derived based on the DWT. With a scaling function $\varphi(t)$, the multiresolution analysis decomposes the load signal $s(t)$ by

$$s(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi(2^{j_0} t - k) + \sum_k \sum_{j=j_0}^{\infty} v_j(k) 2^{j/2} \psi(2^j t - k) \quad (4)$$

where j_0 is the predefined scale of interest and $c_{j_0}(k)$ and $v_j(k)$ are the coefficients, respectively. The first term on the right of (4) provides an approximation component of $s(t)$, while the second term refers to a set of detail components.

A demonstration of two-level decomposition for load series is given by

$$s(t) = A_1(t) + D_1(t) = A_2(t) + D_2(t) + D_1(t). \quad (5)$$

The load signal $s(t)$ is cut up into an approximation component A_2 related to the low frequency and two detail components D_2 and D_1 related to the high frequency. The subseries present a better behavior (more stable variances and fewer outliers) than the original load series and therefore can be predicted more accurately [29]. The reason

behind is the filtering effect of wavelet transform, which can extract irregular information from the original signal [24].

2.3 Wavelet-based ensemble strategy

The type of mother wavelet and number of decomposition levels are two essential parameters in wavelet transform. The selection of these two parameters has a significant impact on the forecasting performance. So far, there is no fixed method to choose the parameters. Many papers have tested various wavelet specifications and selected the most appropriate setting according to the results. However, this kind of selection process has two shortcomings. First, the testing of various wavelet specifications would take too much time. Only the best setting is selected and the others are discarded, which is actually a kind of waste. Second, the fixed specification may not always provide a good representation for load series. As the forecast model is usually built on the rolling window, the fixed specification obtained from one load segment may not be optimal for another segment.

To tackle these problems, this paper proposes a wavelet-based ensemble strategy, which uses the wavelet transform to create the population of individual predictors. The individual members are derived using different combinations of mother wavelet and number of decomposition levels. In STLF, Daubechies (db) wavelets have been widely used [30-32]. In the experiment, it is found that Coiflets (coif) can also give good representation. Moreover, the order of mother wavelet typically ranges from 2 to 5 and the number of levels is less than 4. Therefore, in this paper, 8 wavelet functions: db2–db5 and coif2–coif5 are used and the level for decomposition is from 1 to 3. Overall, there are 24 wavelet parameter combinations to build the ensemble model.

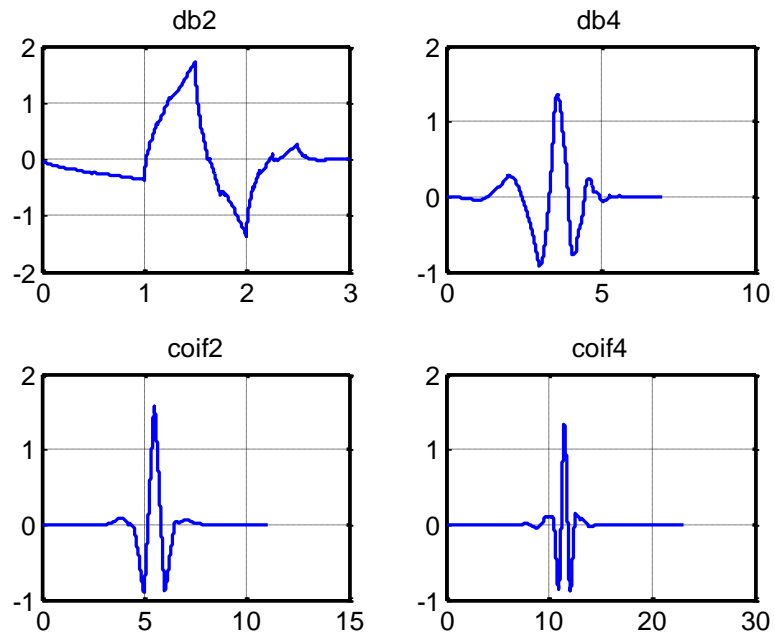


Fig. 2 Four mother wavelet functions.

Four mother wavelets (db2, db4, coif2 and coif4) are shown in Fig. 2.

Technically, a wavelet function has null mean and only oscillates in a short time. As they are in different shapes, the corresponding subseries will also be different. For example, a load signal is decomposed into three levels using the above four wavelets.

The resulting approximation components are shown in Fig. 3.

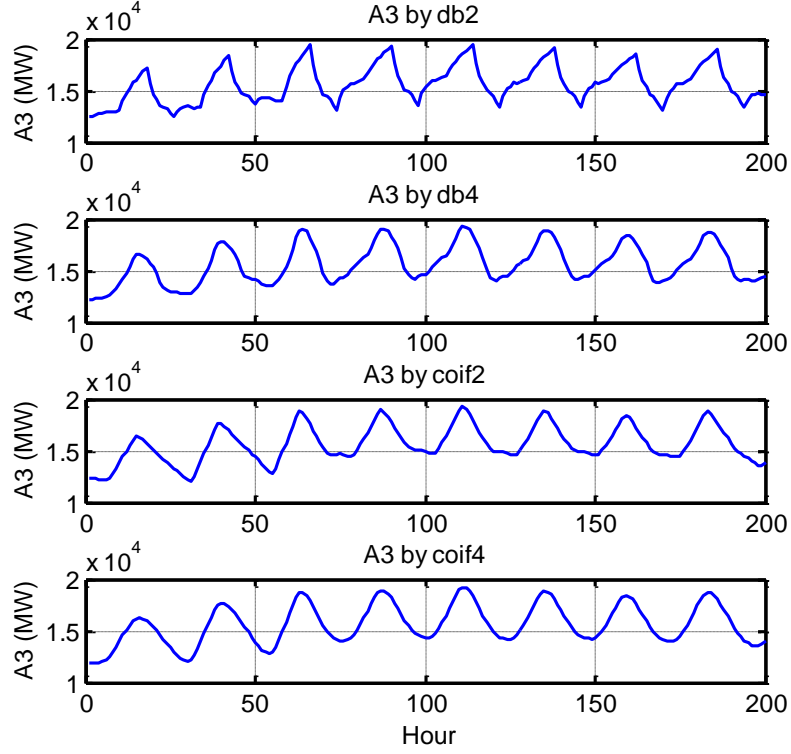


Fig. 3 Approximation components obtained with db2, db4, coif2 and coif4.

The wavelet-based ensemble strategy will produce various input features for the individual forecasters, which can promote the degree of diversity in the ensemble. Every forecast in the ensemble would contain some independent information. The ensemble strategy can make use of the complementary information to improve the generalization capability. Moreover, the ensemble strategy will avoid the selection process for optimal wavelet parameters and reduce the biases introduced by fixed wavelet specification.

2.4 Partial least squares regression (PLSR)

Simple averaging is a commonly used method to combine the individual forecasts [22]. The ensemble output is defined by

$$z_e = \frac{1}{\gamma} \sum_{i=1}^{\gamma} z_i \quad (6)$$

where z_e is the ensemble output, z_i is the individual output and γ is the number of individual outputs. Simple averaging is easy and effective, but it neglects the fact that some individual predictors are more accurate than others. The relative accuracy of individual predictors is considered in the weighted averaging method by

$$z_e = \sum_{i=1}^{\gamma} \alpha_i z_i \quad (7)$$

where α_i is the combining weight. In this paper, PLSR is selected to find the optimal combining weight factors.

Given two blocks of variables \mathbf{X} and \mathbf{Y} , PLSR is going to predict \mathbf{Y} from \mathbf{X} by means of latent variables. In this paper, \mathbf{X} represents the individual outputs and \mathbf{Y} refers to the ensemble output. In the beginning, \mathbf{X} and \mathbf{Y} are preprocessed through mean-centering and rescaling, which are denoted by \mathbf{X}_c and \mathbf{Y}_c , respectively. For each column of \mathbf{X} , mean-centering involves subtracting its column average from the column data. The scaling of a matrix makes each column has standard deviation 1 [33]. Then, PLSR decomposes \mathbf{X}_c and \mathbf{Y}_c by the outer relations [27]:

$$\begin{aligned} \mathbf{X}_c &= \mathbf{TP}^T + \mathbf{E} = \sum_{j=1}^J \mathbf{t}_j \mathbf{p}_j^T + \mathbf{E} \\ \mathbf{Y}_c &= \mathbf{UQ}^T + \mathbf{F} = \sum_{j=1}^J \mathbf{u}_j \mathbf{q}_j^T + \mathbf{F} \end{aligned} \quad (8)$$

where \mathbf{T} and \mathbf{U} are the score matrices (i.e. latent variables), \mathbf{P} and \mathbf{Q} are the loading matrices, and \mathbf{E} and \mathbf{F} are the residual matrices. The superscript ‘‘T’’ means transposition and J is the number of latent components. PLSR aims to search for latent vectors \mathbf{t}_j and \mathbf{u}_j to maximize the covariance between \mathbf{X}_c and \mathbf{Y}_c with the condition that the residuals \mathbf{E} and \mathbf{F} are reduced [34].

As shown in (8), the variables \mathbf{X}_c and \mathbf{Y}_c can be replaced by the new ones \mathbf{T} and \mathbf{U} , which have more compact sizes, better properties (orthogonality) and also span the column spaces of \mathbf{X}_c and \mathbf{Y}_c . PLSR tries to establish an inner relation between \mathbf{T} and \mathbf{U} so as to further link the two blocks \mathbf{X}_c and \mathbf{Y}_c . The inner relation is given by

$$\mathbf{U} = \mathbf{T}\mathbf{B} + \mathbf{U}_E \quad (9)$$

where \mathbf{B} is the matrix of regression coefficients, and \mathbf{U}_E is an error term similar to \mathbf{E} and \mathbf{F} . If these error terms are minimized, we can estimate \mathbf{Y}_c by

$$\hat{\mathbf{Y}}_c = \mathbf{U}\mathbf{Q}^T = \mathbf{T}\mathbf{B}\mathbf{Q}^T. \quad (10)$$

The block of independent variables \mathbf{X} is finally related to the block of dependent variables \mathbf{Y} via a series of intermediate variables. In PLSR, the solutions to the above equations can be obtained by the SIMPLS method [33]. In our experiments, we can directly use the built-in function *plsregress* in Matlab to carry out PLSR.

In PLSR, the number of components J in (8) is an important parameter. Simply using a large number of components is not reliable. The reason is that some latent components with small scores only describe the noise, which leads to overtraining [27]. Hence, cross validation is adopted to choose J based on the training and validation datasets, which can reduce the chance of overtraining. The number that gives a minimal mean squared prediction error is selected.

To determine the combining weights, the multiple linear regression (MLR) method may have matrix inversion problems due to the high degree of correlation between the individual outputs. PLSR is developed as a remedy to tackle the problems in MLR. The regression is not implemented on the original variables but on the latent variables, which have better properties. Thus, PLSR can handle the collinearity and

provide reliable estimates for the combining weights. Moreover, PLSR allows the selection of the number of components used in the regression model. The components with smaller scores can be eliminated to avoid overtraining [27].

2.5 Proposed forecast model

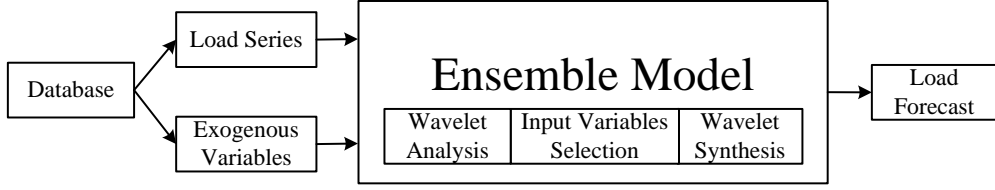
The structure of the proposed STLF model is shown in Fig. 4, which consists of data acquisition, wavelet analysis and synthesis, input variables selection, wavelet-based ensemble model and ELM-based parallel model.

As depicted in Fig. 4a, data are collected and prepared for the ensemble model. In our model, the following candidate input variables are considered:

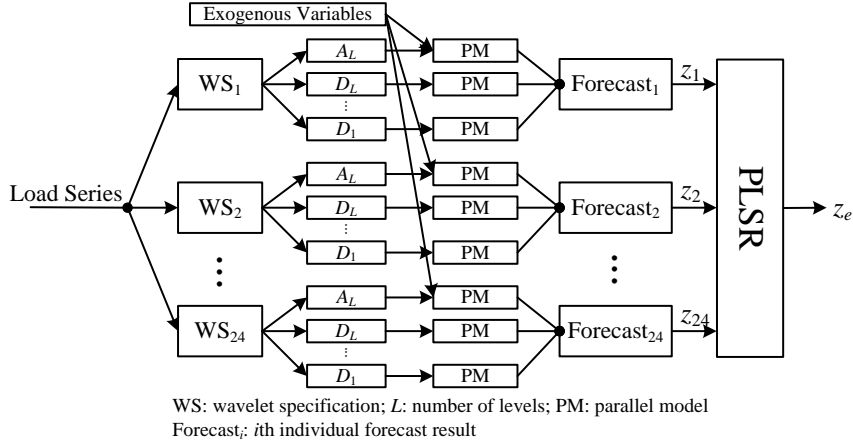
- 1) Past loads: Load behavior exhibits short-run trend and daily and weekly periodicities. Therefore, the past load values up to 200-hour ago are tested. In this paper, correlation analysis is used to select the most relevant ones as input data for the ELM.
- 2) Temperature: Temperature is the most widely used weather factor in load predictors. In our model, the temperature values at time h , $h-1$, $h-2$ and $h-24$ are selected as input data, where h is the forecast hour [16].
- 3) Calendar variables. Day of the week is marked by the integer numbers from 1 to 7. For example, 1 is used for Monday and Sunday is denoted by 7. Moreover, weekends and weekdays are identified by 1 and 0, respectively. All the public holidays are considered as weekends, and marked by 1.

The selected input variables for each subseries are shown in Table 1. In this table, $Dry(h)$ refers to the dry bulb temperature, $Dayofweek(h)$ represents the day of the week and $weekend(h)$ is the weekday, weekend or holiday index. The exogenous variables like temperature and calendar variables are only considered in the

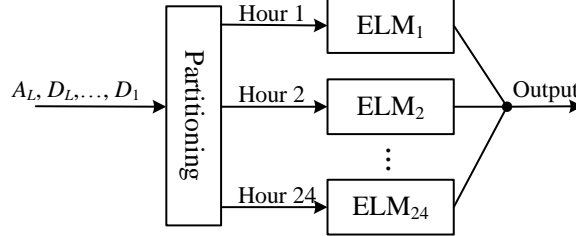
approximation component (i.e. A_i) forecasting.



(a) Model overview



(b) Ensemble model



(c) Parallel model

Fig. 4 Structure of the proposed ensemble method.

Table 1 Selected input variables for each subseries

Level	Subseries	Selected input variables
1, 2, 3	A_i	$A_i(h-1), A_i(h-2), A_i(h-22), A_i(h-23), A_i(h-24), A_i(h-25), A_i(h-26), A_i(h-47), A_i(h-48), A_i(h-49), A_i(h-71), A_i(h-72), A_i(h-73), A_i(h-95), A_i(h-96), A_i(h-97), A_i(h-119), A_i(h-120), A_i(h-121), A_i(h-143), A_i(h-144), A_i(h-145), A_i(h-167), A_i(h-168), A_i(h-169), A_i(h-192), \text{Dry}(h-0), \text{Dry}(h-1), \text{Dry}(h-2), \text{Dry}(h-24), \text{Dayofweek}(h), \text{Weekend}(h)$
	D_i	$D_i(h-24), D_i(h-48), D_i(h-72), D_i(h-96), D_i(h-120), D_i(h-144), D_i(h-168), D_i(h-192)$

The wavelet-based ensemble model is presented in Fig. 4b. In order to capture

the multiscale load features, load series is cut up into a group of approximation and detail components. As discussed in Section 2.3, there are totally 24 combinations of wavelet parameters ($WS_1, WS_2, \dots, WS_{24}$) used for load decomposition. For each sub-component (A_L, D_L, \dots, D_1) obtained from the wavelet decomposition, a parallel model (Fig. 4c) consisting of 24 ELMs is allocated to predict the hourly load of the next day. For each ELM, the sigmoid and linear activation functions are used in the hidden and output layers, respectively. The number of hidden nodes is decided through extensive test, while the number of output nodes is 1.

As shown in Fig. 4c, the load value at each hour is predicted by a separate ELM. Hour i is the hour index of the next day. For example, ELM_1 and ELM_{24} produce the forecasted load values for Hour 1 and 24, respectively. The output of the parallel model is the forecasted load sub-component (e.g. forecasted D_1). If we use a single ELM to recursively predict the loads of next day, the training set will be significantly larger than that of the ELM in a parallel model. To obtain a good forecast result, the single recursive model requires much more hidden neurons than parallel model. Therefore, we can say that the ELM predictors in parallel model are relatively small (i.e. less hidden neurons), and they are not likely to be overtrained in the learning process [35].

It is obvious that the number of individual forecasters is equal to the number of combinations of wavelet parameters. The 24 individual outputs are combined using the PLSR method. Note that the number of latent components involved in the PLSR model is selected by cross validation, which has been discussed in Section 2.4.

3. Results

In this section, the proposed method is tested using actual load and temperature

data. In order to confirm the effectiveness, the proposed method is compared with other state-of-the-art methods. Three error metrics are used to measure the forecasting performance: mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE). They are defined by

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{M} \sum_{i=1}^M \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \\
 \text{MAE} &= \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i| \\
 \text{RMSE} &= \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2}
 \end{aligned} \tag{11}$$

where M is the number of load points, y_i is the actual load and \hat{y}_i is the forecasted load.

Case 1: In this case, the hourly load and temperature data from ISO New England [36] were used. Two typical months have been selected as the testing periods. The first one refers to January 2010, which is a winter month. The second one refers to July 2010, which is a summer month. For the winter month, the training set is from January 1, 2009 to December 31, 2009. For the summer month, the training set is from July 1, 2009 to June 30, 2010. The validation set is the last month of training set, which is used to fine-tune the model parameters.

Four different forecast models are chosen for the purpose of comparison:

- 1) **M1:** Wavelet transform is used to decompose the load series and each sub-component is modeled by a single ELM.
- 2) **M2:** The sub-components in M1 are forecasted by the parallel model, which is made up of 24 ELMs.
- 3) **M3:** The wavelet-based ensemble strategy is applied to generate individual predictors, which are then combined using the simple averaging method.

- 4) **M4 (proposed):** The PLSR method is used to combine the individual predictors in M3.

The 1-hour and 24-hour ahead forecast results are shown in Tables 2 and 3, respectively. The key findings are summarized as follows:

Table 2 Results for 1-hour ahead load forecasting

	Winter month			Summer month		
	MAPE	MAE	RMSE	MAPE	MAE	RMSE
M1	0.4637	71	103	0.7566	139	187
M2	0.4549	70	101	0.6909	125	168
M3	0.3482	55	74	0.5586	109	144
M4	0.2736	43	59	0.5227	103	134

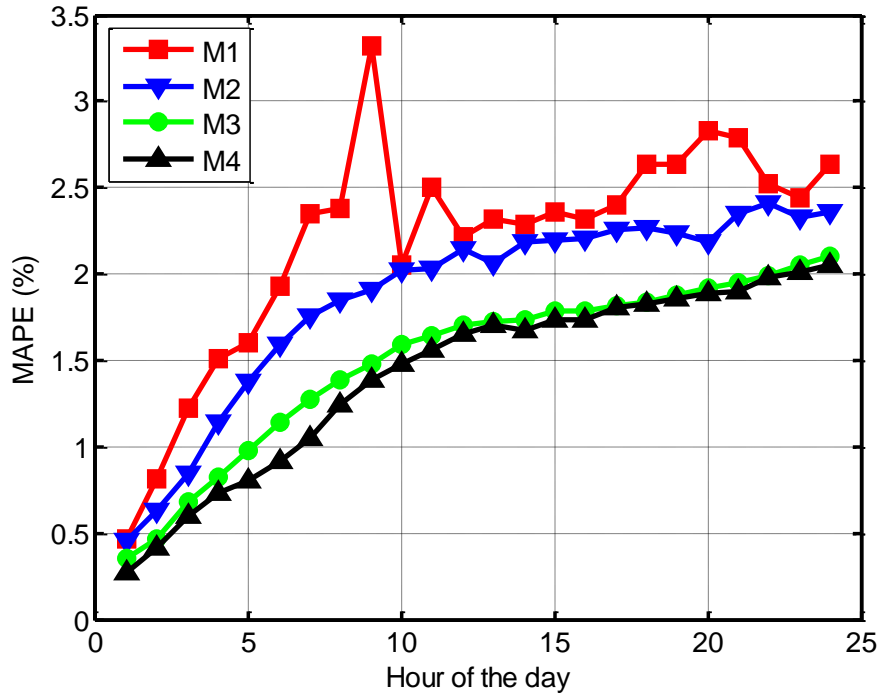
Table 3 Results for 24-hour ahead load forecasting

	Winter month			Summer month		
	MAPE	MAE	RMSE	MAPE	MAE	RMSE
M1	2.1884	332	474	4.2495	775	1058
M2	1.8662	289	402	3.6552	678	957
M3	1.5035	231	322	2.9883	538	749
M4	1.4279	220	303	2.8158	504	708

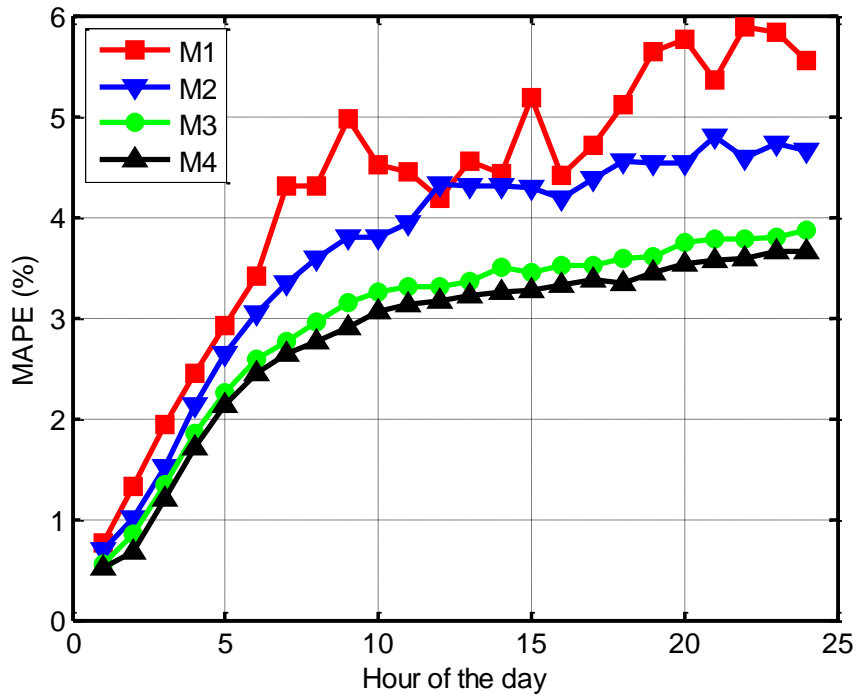
- 1) It can be observed that the forecasting accuracy is improved if the sub-components are forecasted using the parallel model (M2) instead of the single ELM (M1). Furthermore, the improvement in the 1-hour case is smaller than that in the 24-hour case.
- 2) The results indicate that the proposed ensemble strategy is an effective way to enhance the forecasting accuracy. For the winter month, the 1-hour ahead forecast of M3 has obtained an increase of 23.5% (in MAPE) over M2, while the day ahead forecast experiences a 19.4% improvement (in MAPE) from M2 to M3.
- 3) PLSR presents better performance in comparison with simple averaging. Taking the summer month as an example, the forecast error of the 1-hour

case is reduced by 6.4% (in MAPE) from M3 to M4. For the 24-hour case, the forecast accuracy of M3 is 6.1% worse (in MAPE) than that of M4. This implies that PLSR can produce more accurate ensemble forecast.

- 4) The proposed method outperforms other three models in all error measures, which confirms the effectiveness of the proposed method. The forecasting results with different horizons is shown in Fig. 5.



(a) Winter month



(b) Summer month

Fig. 5 Forecast results of M1-M4.

Case 2: This case compares the proposed method to standard neural network (SNN) and similar day-based wavelet neural networks (SIWNN) in [31]. The SNN

method was carried out by using a single neural network, which used weekday index, weather and historical load values as the inputs. The SIWNN method selected similar day load as the input data and used wavelet decomposition and separate neural networks to capture the load features. The comparison is performed on the ISO New England data. The training period is from March 2003 to December 2005. The hourly loads from January 1, 2006 to December 31, 2006 are predicted. The 24-hour ahead forecast results are shown in Fig. 6, where the values of SNN and SIWNN are extracted from [31]. It is clear that the proposed method produces lower forecast errors than other methods in all months. On an average, the proposed method is 26.6% and 12.9% better than the SNN and SIWNN methods, respectively.

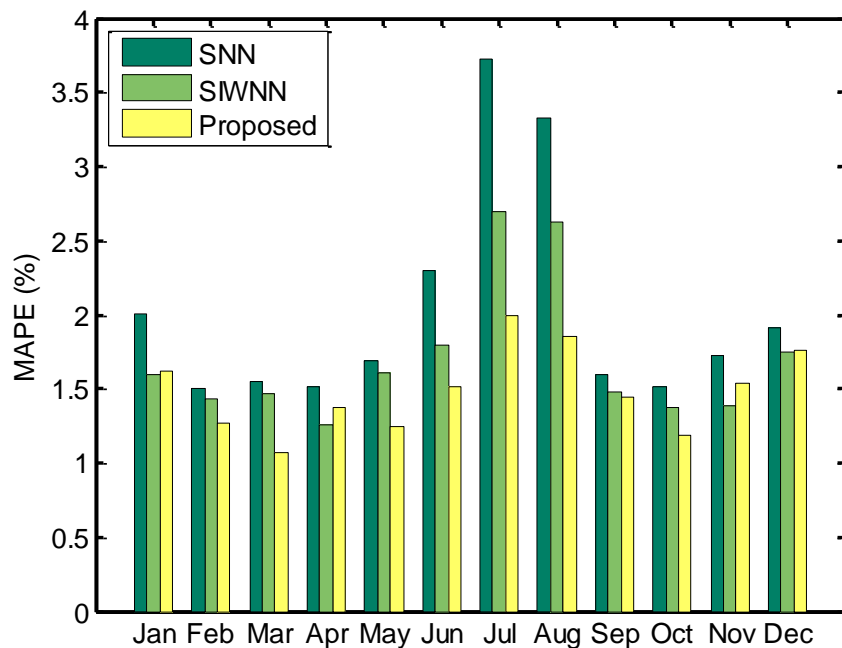


Fig. 6 Monthly results for the models in Case 2.

Case 3: In this case, a comparison is made between the proposed method and the ISO-NE method in [37] and the wavelet neural networks (WNN) method in [38] based on the ISO New England data. The testing period ranges from July 1, 2008 to July 31, 2008. Only the 1-hour ahead load forecast is considered. The forecast results

of the three methods are shown in Table 4. On the given testing month, the proposed method has better performance than the WNN method, about 16.3% better in MAPE and 15.9% better in MAE. Compared to the ISO-NE method, the proposed method shows significant improvements in both MAPE and MAE metrics.

Table 4 Forecast results for the three methods in Case 3

	MAPE (%)	MAE (MW)
ISO-NE [37]	0.81	138.33
WNN [38]	0.49	83.54
Proposed	0.41	70.22

Case 4: This case compares the proposed method to the neural network models in [39]. The ISO New England data from 2004 to 2009 are used for training, and the testing period is from 2010 to 2011. The comparative models are: SVR, ELM, improved second-order (ISO) algorithm, original and modified error correction (ErrCor) algorithms. Only the 24-hour ahead load forecasting is studied. The forecast results are shown in Table 5. The improvement of the proposed method compared to the best available method is given in the last row. It is clear that the proposed method produces better results than other methods in all testing cases.

Table 5 Forecast results of the models in Case 4

	MAPE values	
	2010	2011
SVR	1.79	2.07
ELM	1.83	2.19
ISO	1.95	2.20
ErrCor original	1.80	2.02
ErrCor modified	1.75	1.98
Proposed	1.50	1.80
Improvement	14.3%	9.1%

Case 5: This case examines the proposed method based on the North America electric utility data [40]. The experiment is conducted using the load and temperature data from January 1, 1988 to October 12, 1992. The two-year loads prior to October

12, 1992 are forecasted and the remaining data are used for training.

The proposed method is compared to other state-of-the-art methods reported in [29, 30, 35, 41]. Both 1-hour and 24-hour ahead forecasts are tested. To study the effect of temperature errors on load forecasting, the proposed method is examined using noisy temperatures. The Gaussian noise of zero mean and standard deviation of 0.6 °C is added to the measurements [30]. It is noted that holidays and weekends are the challenging parts in load forecasting, because the load pattern is quite different from that of regular workdays.

The results with and without holidays and weekends are presented in Tables 6 and 7, respectively. The improvement of the proposed method compared to the best available method is given in the last row. It can be observed that the proposed method outperforms other methods in all testing cases on the North American utility data. Moreover, the proposed method can result in encouraging forecasting results in case of noisy temperatures, and holidays and weekends.

Table 6 MAPE (%) values with holidays and weekends

	Actual temperature		Noisy temperature	
	1-hour	24-hour	1-hour	24-hour
SSA-SVR [35]	0.72	1.99	0.73	2.03
ESN [41]	1.14	2.37	1.21	2.533
Proposed	0.59	1.86	0.62	1.90
Improvement	18.1%	6.5%	15.1%	6.4%

Table 7 MAPE (%) values without holidays and weekends

	Actual temperature		Noisy temperature	
	1-hour	24-hour	1-hour	24-hour
Model 1 [30]	2.10	3.58	2.15	4.46
Model 2 [30]	1.10	3.41	1.11	3.64
Model 3 [30]	1.12	3.16	1.14	3.38
Model 4 [30]	1.99	2.64	2.04	2.82
WT-NN-EA [29]	0.99	2.04	n.a.	n.a.
ESN [41]	1.048	2.1174	n.a.	n.a.
Proposed	0.52	1.81	0.57	1.85
Improvement	47.5%	11.3%	48.6%	34.5%

Case 6: In this case, the proposed method is compared to the abductive network model in [42] on the North American utility data. The five-year (1985-1989) data are used for model synthesis and the forecast model is tested over the following year (1990). The forecast results of 1-hour and 24-hour cases are given in Table 8. Significant improvements for both 1-hour and 24-hour ahead load forecasting are observed using the proposed method.

Table 8 MAPE (%) values for the models in Case 6

	1-hour	24-hour
Abductive [42]	1.14	2.66
Proposed	0.54	2.02
Improvement	52.6%	24.1%

4. Conclusions

In this paper, a novel STLF method is proposed based on the ELM. The proposed wavelet-based ensemble strategy employs different wavelet specifications to create an ensemble of individual predictors. For each sub-component obtained from the wavelet decomposition, a parallel forecast model is established. To make an accurate ensemble forecast, the individual outputs are combined using the PLSR method. The proposed STLF method can alleviate many problems, such as overtraining and wavelet parameter determination.

The proposed method has been tested using actual data from two electric utilities. Both 1-hour and 24-hour ahead load forecasts are considered. The results demonstrate that the proposed method can produce better forecasting accuracy than other state-of-the-art models.

5. References

- [1] Masa-Bote D, Castillo-Cagigal M, Matallanas E, Caamaño-Martín E, Gutiérrez A, Monasterio-Huelín F, et al. Improving photovoltaics grid integration through short time forecasting and self-consumption. *Applied Energy*. 2014;125:103-13.
- [2] Bunn DW, Farmer ED. *Comparative Models for Electrical Load Forecasting*. Chichester, NY: Wiley; 1985.
- [3] Gross G, Galiana FD. Short-term load forecasting. *Proceedings of the IEEE*. 1987;75:1558-73.
- [4] Javed F, Arshad N, Wallin F, Vassileva I, Dahlquist E. Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short term multiple loads forecasting. *Applied Energy*. 2012;96:150-60.
- [5] Soliman SA, Alkandari AM. *Electrical Load Forecasting: Modeling and Model Construction*. Burlington, MA: Butterworth-Heinemann; 2010.
- [6] Hagan MT, Behr SM. The time series approach to short term load forecasting. *Power Systems, IEEE Transactions on*. 1987;2:785-91.
- [7] Amjady N. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *Power Systems, IEEE Transactions on*. 2001;16:798-805.
- [8] Metaxiotis K, Kagiannas A, Askounis D, Psarras J. Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. *Energy Conversion and Management*. 2003;44:1525-34.
- [9] Yoo H, Pimmely RL. Short term load forecasting using a self-supervised adaptive neural network. *Power Systems, IEEE Transactions on*. 1999;14:779-84.
- [10] Ho KL, Hsu YY, Chen CF, Lee TE, Liang CC, Lai TS, et al. Short term load forecasting of Taiwan power system using a knowledge-based expert system. *Power Systems, IEEE Transactions on*. 1990;5:1214-21.
- [11] Mohandes M. Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*. 2002;26:335-45.
- [12] Hong WC. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy*. 2011;36:5568-78.
- [13] AlRashidi MR, El-Naggar KM. Long term electric load forecasting based on particle swarm optimization. *Applied Energy*. 2010;87:320-6.
- [14] Wang J, Li L, Niu D, Tan Z. An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy*. 2012;94:65-70.
- [15] Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*. 2012;42:513-29.
- [16] Li S, Wang P, Goel L. Short-term load forecasting by wavelet transform and evolutionary extreme learning machine. *Electric Power Systems Research*. 2015;122:96-103.
- [17] Chen X, Dong ZY, Meng K, Xu Y, Wong KP, Ngan HW. Electricity price forecasting with extreme learning machine and bootstrapping. *Power Systems, IEEE Transactions on*. 2012;27:2055-62.
- [18] Hippert HS, Pedreira CE, Souza RC. Neural networks for short-term load forecasting: a review and evaluation. *Power Systems, IEEE Transactions on*. 2001;16:44-55.
- [19] Liu N, Wang H. Ensemble based extreme learning machine. *Signal Processing Letters, IEEE*. 2010;17:754-7.
- [20] Brown G, Wyatt JL, Ti P. Managing diversity in regression ensembles. *Journal of Machine Learning Research*. 2005;6:1621-50.
- [21] Taylor JW, Buizza R. Neural network load forecasting with weather ensemble predictions. *Power Systems, IEEE Transactions on*. 2002;17:626-32.
- [22] Abdel-Aal RE. Improving electric load forecasts using network committees. *Electric Power Systems Research*. 2005;74:83-94.
- [23] De Felice M, Xin Y. Short-term load forecasting with neural network ensembles: a comparative study. *Computational Intelligence Magazine, IEEE*. 2011;6:47-56.
- [24] Bashir ZA, El-Hawary ME. Applying wavelets to short-term load forecasting using PSO-based neural networks. *Power Systems, IEEE Transactions on*. 2009;24:20-7.

- [25] Moazzami M, Khodabakhshian A, Hooshmand R. A new hybrid day-ahead peak load forecasting method for Iran's National Grid. *Applied Energy*. 2013;101:489-501.
- [26] Zhang R, Dong ZY, Xu Y, Meng K, Wong KP. Short-term load forecasting of Australian national electricity market by an ensemble model of extreme learning machine. *Generation, Transmission & Distribution, IET*. 2013;7:391-7.
- [27] Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*. 1986;185:1-17.
- [28] Daubechies I. *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics; 1992.
- [29] Amjady N, Keynia F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy*. 2009;34:46-57.
- [30] Reis AJR, da Silva APA. Feature extraction via multiresolution analysis for short-term load forecasting. *Power Systems, IEEE Transactions on*. 2005;20:189-98.
- [31] Chen Y, Luh PB, Guan C, Zhao Y, Michel LD, Coolbeth MA, et al. Short-term load forecasting: similar day-based wavelet neural networks. *Power Systems, IEEE Transactions on*. 2010;25:322-30.
- [32] Pandey AS, Singh D, Sinha SK. Intelligent hybrid wavelet models for short-term load forecasting. *Power Systems, IEEE Transactions on*. 2010;25:1266-73.
- [33] de Jong S. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*. 1993;18:251-63.
- [34] Rosipal R, Krämer N. Overview and recent advances in partial least squares. *Subspace, Latent Structure and Feature Selection: Springer Berlin Heidelberg*; 2006. p. 34-51.
- [35] Ceperic E, Ceperic V, Baric A. A strategy for short-term load forecasting by support vector regression machines. *Power Systems, IEEE Transactions on*. 2013;28:4356-64.
- [36] ISO New England Data, <http://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info>.
- [37] Shamsollahi P, Cheung KW, Chen Q, Germain EH. A neural network based very short term load forecaster for the interim ISO New England electricity market system. *Power Industry Computer Applications (PICA 2001)2001*. p. 217-22.
- [38] Guan C, Luh PB, Michel LD, Wang Y, Friedland PB. Very short-term load forecasting: wavelet neural networks with data pre-filtering. *Power Systems, IEEE Transactions on*. 2013;28:30-41.
- [39] Cecati C, Kolbusz J, Rozycki P, Siano P, Wilamowski BM. A novel RBF training algorithm for short-term electric load forecasting and comparative studies. *Industrial Electronics, IEEE Transactions on*. 2015;62:6519-29.
- [40] North American Electric Utility Data, <http://sites.google.com/site/fkeynia/loaddata>.
- [41] Deihimi A, Showkati H. Application of echo state networks in short-term electric load forecasting. *Energy*. 2012;39:327-40.
- [42] Abdel-Aal RE. Short-term hourly load forecasting using abductive networks. *Power Systems, IEEE Transactions on*. 2004;19:164-73.