

Short-term load forecasting by wavelet transform and evolutionary extreme learning machine

Song Li^{a,*}, Peng Wang^a and Lalit Goel^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 83286906. E-mail: sli5@e.ntu.edu.sg.

Abstract

This paper proposes a novel short-term load forecasting (STLF) method based on wavelet transform, extreme learning machine (ELM) and modified artificial bee colony (MABC) algorithm. The wavelet transform is used to decompose the load series for capturing the complicated features at different frequencies. Each component of the load series is then separately forecasted by a hybrid model of ELM and MABC (ELM-MABC). The global search technique MABC is developed to find the best parameters of input weights and hidden biases for ELM. Compared to the conventional neuro-evolution method, ELM-MABC can improve the learning accuracy with fewer iteration steps. The proposed method is tested on two datasets: ISO New England data and North American electric utility data. Numerical testing shows that the proposed method can obtain superior results as compared to other standard and state-of-the-art methods.

Keywords: Artificial bee colony, extreme learning machine, short-term load forecasting, wavelet transform.

1. Introduction

Short-term load forecasting (STLF) is always essential for electric utilities to estimate the load power from one hour ahead up to a week ahead. Load forecasting can be used for power generation scheduling, load switching and security assessment. Accurate forecast results help to improve the power system efficiency, reduce the operating cost and cut down the occurrences of power interruption events. Load forecasting becomes more important because of the development of the deregulated electricity markets and the promotion of the smart grid technologies.

Many statistical methods have been used for STLF, including exponential smoothing [1], Kalman filters [2] and time series methods [3]. These methods are highly attractive because some physical interpretation can be attached to their components. However, they cannot properly represent the nonlinear behavior of the load series. Hence, artificial intelligence techniques have been tried out such as neural networks (NNs), fuzzy logic and support vector machines [4-7]. In particular, NNs have drawn the most attention because of their capability to fit the nonlinear relationship between load and its dependent factors.

Recently, extreme learning machine (ELM) has been proposed to train single-hidden layer feedforward neural networks (SLFNs), which can overcome the drawbacks (e.g. time-consuming and local minima) faced by the gradient-based methods [8]. In ELM, the input weights and hidden biases are initialized with a set of random numbers. The output weights of hidden layer are directly determined through a simple inverse operation on the hidden layer output matrix. ELM has been verified to obtain good performance in many applications, including electricity price and load forecasting [9, 10].

This paper presents a hybrid STLF model based on the ELM. Two improvements are carried out to tackle the two key issues in load forecasting: the nonstationary behavior of load series and the robustness of forecast model [6]. First, the wavelet transform is an efficacious treatment to handle the nonstationary load behavior, because it can provide an in-depth time-frequency representation of the load series [11, 12]. We use wavelets to decompose the load series into a set of different frequency components and each component is then separately forecasted. In such a way, we don't handle all the frequency components by a single forecaster but treat them differently. Second, it is found that ELM may yield unstable performance because of the random assignments of input weights and hidden biases [13]. To alleviate this problem, the modified artificial bee colony (MABC) algorithm is developed to look for the optimal set of input weights and hidden biases. MABC is a swarm-based optimization algorithm, which simulates the intelligent foraging behavior of honey bee swarm [14]. MABC can be easily employed and does not require any gradient information. Furthermore, MABC can probe the unknown regions in the solution space and look for the global best solution. This hybrid learning method can be named as ELM-MABC, which makes use of the merits of ELM and MABC.

The proposed method is tested on two datasets: ISO New England data and North American electric utility data. Section 2 describes wavelet transform, extreme learning machine, artificial bee colony algorithm and the proposed STLF method. Simulations are presented in Section 3. Section 4 provides discussion and Section 5 outlines conclusion.

2. Methodology

2.1 Wavelet transform

The multiple frequency components in load series are always the challenging parts in forecasting [15]. A single forecaster cannot handle them appropriately and we can treat them differently with the help of wavelet transform. Wavelet transform can be used to decompose a load profile into a series of constitutive components [16]. These constitutive components usually have better behaviors (e.g. more stable variance and fewer outliers) and therefore can be forecasted more accurately [15].

Wavelet transform makes use of two basic functions: scaling function $\varphi(t)$ and mother wavelet $\psi(t)$. A series of functions are derived from the scaling function $\varphi(t)$ and the mother wavelet $\psi(t)$ by

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \quad (1)$$

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad (2)$$

where j and k are integer variables for scaling and translating [17]. The wavelet functions $\psi_{j,k}(t)$ and scaling functions $\varphi_{j,k}(t)$ can be used for signal representation. Then a signal $S(t)$ can be expressed by

$$S(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi(2^{j_0} t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2} \psi(2^j t - k) \quad (3)$$

where j_0 is the predefined scale, $c_{j_0}(k)$ and $d_j(k)$ are the approximation and detail coefficients, respectively. It is seen that wavelet decomposition is done to compute the above two sets of coefficients. The first term on the right of (3) gives a low resolution representation of $S(t)$ at the predefined scale j_0 . For the second term, a higher resolution or a detail component is added one after another from the predefined scale j_0 [18].

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

$$S(t) = A_1(t) + D_1(t) = A_2(t) + D_2(t) + D_1(t). \quad (4)$$

The load signal S is broken up into a set of constitutive components. The approximation A_2 reflects the general trend and offers a smooth form of the load signal. The terms D_2 and D_1 depict the high frequency components in the load signal. Specifically, the amplitude of D_1 is very small, which carries information about the noise in the load signal.

Three issues must be considered before using the wavelet transform: type of mother wavelet, number of decomposition levels and border effect. In this paper, the trial and error method is used to choose the mother wavelet and number of decomposition levels. Three popular wavelet families: Daubechies (db), Coiflets (coif) and Symlets (sym) [16] are investigated for decomposing the load signal. The combinations of 12 mother wavelets (db2–db5, coif2–coif5 and sym2–sym5) and 3 decomposition levels (1–3) have been tested. It is found that the combination of coif4 and 2-level decomposition can produce the best forecasting performance. In addition, the border distortion will arise if the transform is performed on finite-length signals, which would degrade the performance. The signal extension method in [19] is adopted in this paper, which appends the previous measured values at the beginning of the load signal and forecasted values at the end of it.

2.2 Modified artificial bee colony (MABC) algorithm

The ABC algorithm, introduced by Karaboga, simulates the intelligent foraging behavior of honey bees [14]. The swarm in ABC is divided into three groups: employed bees, onlookers and scouts. The position of a food source represents a solution to the target problem while the nectar amount stands for the fitness value of that solution. An employed

bee may update its position in case of finding a new food source. If the fitness of the new source is higher than that of the old one, the employed bee chooses the new position over the old one. Otherwise, the old position is retained. After all the employed bees finish search missions, they share the information (i.e. positions and nectar amounts) of the food sources with the onlookers in the hive. An onlooker bee will choose a food source based on the associated probability value p_i , which is given by

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \quad (5)$$

where fit_i is the fitness value of i th food source and SN is the number of food sources.

The basic ABC generates a new solution v_{ij} from the old one u_{ij} by:

$$v_{ij} = u_{ij} + \theta_{ij} (u_{ij} - u_{kj}) \quad (6)$$

where i and k are the solution indices and j is the dimension index. The index k has to be different from i and θ_{ij} is a uniformly random number within the range $[-1, 1]$. The old solution u_{ij} will be replaced by the new one v_{ij} , provided that v_{ij} has a better fitness value.

If a food source cannot be improved for many cycles, this source is abandoned. The number of cycles for abandonment is called *limit*, which is a control parameter in ABC.

The employed bee related to the abandoned food source becomes a scout. The scout discovers the new food position by

$$u_{ij} = u_{\min,j} + rand(0, 1)(u_{\max,j} - u_{\min,j}) \quad (7)$$

where $u_{\min,j}$ and $u_{\max,j}$ are the lower and upper bounds for the dimension j , respectively. The random number in (7) follows the uniform distribution.

It has been pointed out that the search equation given by (6) is good at exploration but poor at exploitation [20]. To balance these two capabilities and improve the convergence performance, a modified search equation is proposed as follows:

$$v_{ij} = w \cdot u_{best,j} + \theta_{ij}(u_{best,j} - u_{ij}) \quad (8)$$

where u_{best} is the best solution in current population, and w is the inertia weight. The search equation (8) uses the information of the best solution to direct the movement of population. The new solution is driven towards the best solution of the previous cycle. The coefficient w controls the impact from the best solution u_{best} . A large weight encourages the global exploration, while a small one speeds up the convergence to optima. In this paper, the inertia weight w is chosen to be 0.1. Hence, the modified equation is able to improve the exploitation capability and accelerate the convergence speed. The search process of MABC will end if a stop criterion is satisfied. Normally, a maximum cycle number (MCN) is used to terminate the algorithm.

2.3 Evolutionary extreme learning machine

2.3.1 Basic ELM

ELM is an emerging learning algorithm for SLFN, which randomly chooses the input weights and hidden biases and determines the output weights directly by a least squares method [21]. Consider a training set of N samples $(\mathbf{x}_i, \mathbf{t}_i)$, the SLFN can be modeled by

$$\sum_{j=1}^n \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{o}_i, \quad i = 1, \dots, N \quad (9)$$

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

If ELM fits all the training samples (\mathbf{x}_i, t_i) with zero error, it can be said that there exist β_j , \mathbf{w}_j and b_j such that

$$\sum_{j=1}^n \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = t_i, \quad i = 1, \dots, N. \quad (10)$$

The compact form of (10) can be given by $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where $\boldsymbol{\beta}=[\beta_1, \dots, \beta_n]^T$, $\mathbf{T}=[t_1, \dots, t_N]^T$ and \mathbf{H} is called the hidden layer output matrix.

In practice, ELM cannot obtain the perfect zero error because the number of hidden nodes n is usually less than the number of training samples N . In ELM, the input weights \mathbf{w}_j and hidden biases b_j are randomly initialized. For settled \mathbf{w}_j and b_j , the SLFN becomes an over-determined linear system and the output weights $\boldsymbol{\beta}$ can be calculated by a least squares method. A special solution is given by $\boldsymbol{\beta}^* = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moore-Penrose (MP) inverse of \mathbf{H} . It is suggested that the singular value decomposition method is well-suited to compute the MP inverse of \mathbf{H} in all cases [22].

ELM tends to have good generalization performance because both the minimum training error and the smallest norm of weights can be directly achieved. The special solution $\boldsymbol{\beta}^*$ is one of the least squares solutions of the linear system $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, which implies that ELM can reach the minimum error of the current system. Moreover, $\boldsymbol{\beta}^*$ has the smallest norm among all the least squares solutions of $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$. It is shown in [23] that the smaller the weights are, the better generalization performance the network tends to have. In addition,

ELM can get rid of many trivial problems faced by traditional methods, such as local minima, stopping criteria and learning rate [24].

2.3.2 Evolutionary ELM

It is observed that the performance of ELM depends highly on the chosen set of input weights and hidden biases. ELM may have worse performance in case of non-optimal parameters. In this paper, the proposed MABC algorithm is used to find the optimal set of input weights and hidden biases for ELM.

In the first step, the initial population is generated and each candidate solution \mathbf{u}_i consists of a set of input weights and hidden biases by

$$\mathbf{u}_i = [w_{11}, w_{12}, \dots, w_{1n}, w_{21}, w_{22}, \dots, w_{2n}, \dots, w_{m1}, w_{m2}, \dots, w_{mn}, b_1, b_2, \dots, b_n] \quad (11)$$

where n is the number of hidden nodes and m is the number of input nodes. All the variables in the individuals are within the range $[-1, 1]$. Secondly, for each individual, the output weights are obtained through calculating the MP inverse. In this paper, the root mean square error (RMSE) is chosen as the fitness function, which is given by

$$Fitness = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\sum_{j=1}^n \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) - t_i \right)^2}, \quad i = 1, \dots, N, j = 1, \dots, n. \quad (12)$$

Thirdly, the population is subjected to the search process of MABC. The optimal input weights and hidden biases are obtained until MABC completes $M CN$ cycles.

The hybrid learning algorithm can take advantage of the merits of ELM and MABC. First, MABC is a global search technique with strong exploitation capability, which allows the learning algorithm to avoid the local minima and converge to the global minimum. Moreover, the optimal parameters from MABC guarantee that ELM has a small training

error. Second, it should be noted that in the conventional neuro-evolution methods [12], all the network weights (i.e. input weights, hidden biases and output weights) will be fine-tuned by the evolutionary algorithm (MABC is used as the evolutionary algorithm for fair comparison in Example 1). However, in ELM-MABC, only parts of weights (i.e. input weights and hidden biases) are adjusted by MABC. The output weights of the hidden layer are not determined by MABC but the least squares method. This difference can lead to many advantages in training the network. The iterative minimization is performed over the set of input weights and hidden biases instead of all weight parameters. The learning process will be accelerated because fewer parameters are estimated. Furthermore, since the output weights are calculated by a least squares method at each iteration, the training error is always at a global minimum with respect to the output weights [25]. The robustness of training process is highly improved.

The procedure of ELM-MABC, shown in Fig. 1, can be described as follows:

- a) Generate the initial population randomly. Each individual (i.e. candidate solution) in the population consists of a set of input weights and hidden biases.
- b) For each individual, calculate the matrix \mathbf{H} and the output weights β .
- c) Evaluate the fitness of each individual and start the search process.
- d) Repeat the search process for MCN cycles.
- e) Output the best solution as the optimal set of input weights and hidden biases.

2.4 Proposed forecast model

The structure of the proposed STLF method is shown in Fig. 2. The step by step procedure can be summarized as follows:

a) Extract the required data and divide them into load series and exogenous variables.

It is noted that the resolution of data is one hour, unless otherwise stated.

b) Use the mother wavelet to break up the load series into three sub-series: A_2 , D_2 and D_1 , as discussed in Section 2.1.

c) Select the input variables for each sub-series. The details of input variables selection are presented in Section 3.1. Test different sets of input variables if necessary.

d) Determine the number of hidden nodes for each ELM. It should be noted that there is litter theoretical basis for determining the number of hidden nodes of a network [26]. In this paper, we have tested a few alternative numbers and selected the one gives the best prediction performance.

e) Produce the optimal input weights and hidden biases using MABC. The control parameters of MABC are determined by heuristics and experience. We have tried various parameters and selected the setting that gives the best performance.

f) Evaluate the model on the validation dataset. If the prediction accuracy is not satisfactory, repeat the steps c) to f). Otherwise, go to step g).

g) Deploy the obtained model to forecast future load.

3. Simulations

3.1 Input variables selection

Input variables selection is a very important preprocess step of load forecasting. There are many variables that can be used in STLF. Generally speaking, more input variables may provide more accurate results. However, excessive variables are prone to cause many problems, such as prolonged training process, unnecessary storage space and the curse of

dimensionality [27]. Therefore, a compact set of input variables is selected for the predictor.

Suppose the forecast hour is time τ , the following candidate variables are considered:

a) Historical load. Correlation analysis is used to select the most relevant historical load values as the load inputs. The 200-hour load data prior to the time τ is considered for selection.

b) Temperature. In most situations, temperature is the key factor to drive the variations of load consumption. The temperature values at time τ , $\tau - 1$, $\tau - 2$ and $\tau - 24$ are used as the temperature inputs in our model.

c) Day of the week. The numbers from 1 to 7 are used to mark the day of the week. For example, 1 is used for Monday and Sunday is marked by 7.

d) Hour of the day. The load series usually exhibits a daily pattern and it is necessary to pass this information to the forecaster. This can be achieved by defining two additional variables to codify the hour of the day. Two variables: $H_a = \sin(2\pi h/24)$ and $H_b = \cos(2\pi h/24)$ are included in our model, where h is the hour in a day (0, 1, ..., 23) [28].

e) Weekend index. The numbers 1 and 0 are used to identify weekdays and weekends: 1 for a weekend and 0 for a weekday. All the holidays are regarded as weekends, too.

3.2 Case studies

In this section, the proposed method is tested using the actual load and temperature data. The proposed method is also compared with other methods based on two publicly available datasets: ISO New England data¹ and North American electric utility data². The two electric utilities are different in size, usage pattern of electricity and weather conditions. All simulations were conducted in Matlab on a personal computer with a 2.66-GHz CPU

¹ Available at <http://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info>

² Available at <http://sites.google.com/site/fkeynia/loaddata>

and 3.25-GB memory. To evaluate the forecasting performance, two error metrics: mean absolute percentage error (MAPE) and mean absolute error (MAE) are used. They are defined by

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{A_i - F_i}{A_i} \right| \times 100\%; \quad MAE = \frac{1}{M} \sum_{i=1}^M |A_i - F_i|. \quad (13)$$

where M is the number of data points, A_i is the actual value and F_i is the forecast value.

Example 1: This example compares the convergence performance between the conventional neuro-evolution method (denoted by NN-MABC) and ELM-MABC. As mentioned in Section 2.3.2, in NN-MABC, all the weight values are tuned by evolutionary algorithm (MABC is used as the evolutionary algorithm for fair comparison). For ELM-MABC, only the input weights and hidden biases are adjusted by MABC. For each learning method, 20 trials have been conducted and the average convergence performance has been reported. Fig. 3 shows the convergence performance of NN-MABC and ELM-MABC, with different numbers of hidden nodes. The training data used are actual measurements from ISO New England.

It is observed that the convergence performance of ELM-MABC is significantly better than that of NN-MABC. After finishing 500 iterations, the error of ELM-MABC is quite smaller than that of NN-MABC. Moreover, the error curve of ELM-MABC is also far below its rival all the time. For 10-node case, the error of NN-MABC at iteration 500 (0.0534) is just slightly smaller than that of ELM-MABC at iteration 1 (0.0573). The gap of performance is wider in terms of the 20-node case. The error of ELM-MABC at iteration 1 (0.0330) is even lower than that of its rival at iteration 500 (0.0535).

The improvement in training SLFN benefits from the special learning mechanism of ELM. ELM-MABC converges faster than NN-MABC because only the input weights and hidden biases are estimated. The output weights of hidden layer are calculated by a least squares method. This implies that the training error of ELM-MABC is always at a global minimum with respect to the output weights. Therefore, ELM-MABC has a much better learning accuracy than NN-MABC, as shown in Fig. 3.

Example 2: In this example, the proposed method is used to perform both 1-hour and 24-hour ahead load forecasting. The hourly load and temperature data are collected from ISO New England. The data from Nov. 2009 to Dec. 2010 are used to run simulations. In order to test the proposed method, for each month in 2010, the third week is selected as the testing week. The one week prior to the testing week is set to be the validation week. The five weeks before the validation week are used as the training weeks. Note that the training data will have a part of data from the previous month. The parameters of the forecast model are adjusted on the validation data. During the simulations, we use actual temperatures as the forecasted values. To evaluate the effect of weather forecasting error, the forecasted temperatures are manually simulated. The Gaussian noise of zero mean and standard deviation of 0.6 °C is added to the actual temperature data, which is advised in [19].

The 1-hour and 24-hour ahead forecast results are tabulated in Table 1. It can be seen that the proposed method can yield satisfactory results with two different time horizons. The average error is 0.5554 for 1-hour ahead case and 1.59 for 24-hour ahead case, respectively. The errors of 1-hour ahead case are much smaller than those of 24-hour ahead case. Furthermore, the proposed method is able to generate encouraging results with

simulated temperatures. Under the circumstance of Gaussian noise, the average forecast error only increases 5.8% and 10.1% for 1-hour and 24-hour ahead forecasts, respectively.

Example 3: This example studies the influence of wavelet transform and MABC algorithm on the forecasting performance. The proposed method (WT-ELM-MABC) is compared with other three methods: ELM, ELM with wavelet transform (WT-ELM) and ELM with MABC algorithm (ELM-MABC). All the ELMs have only one output neuron. The sigmoid and linear functions are adopted in the hidden and output layers, respectively. The testing data are identical with those in Example 2. The results for 1-hour ahead load forecasting are shown in Table 2. The findings are summarized as follows:

a) It can be observed that the performance is greatly improved if wavelet transform is involved. With the help of wavelet transform, WT-ELM has obtained an improvement of 15.2% over ELM. Compared with ELM-MABC, WT-ELM-MABC has also experienced an increase of 15.9% in forecast accuracy.

b) The forecast results of Table 2 indicate that the MABC algorithm is an effective tool to improve the forecast accuracy. Comparing WT-ELM with WT-ELM-MABC, the forecast error is reduced by 12.7% if the input weights and hidden biases are pre-optimized. For ELM and ELM-MABC, the accuracy is 13.5% worse if MABC is not used in the model.

c) It is seen that WT-ELM-MABC presents much better performance as compared to other three approaches. On an average, the enhancements of WT-ELM-MABC are 25.9%, 12.7% and 15.9% with regard to the previous three methods, respectively.

d) It is noteworthy to draw attention to the computational time of above four methods. The first two methods ELM and WT-ELM only take a few seconds to complete the training

and testing process, because ELM has no iterative steps. The time of the latter two methods becomes much longer since MABC is adopted to optimize ELM. To accomplish the task, ELM-MABC and WT-ELM-MABC spend about two and five minutes, respectively.

Example 4: This example compares the proposed method to the ISO-NE method in [29] and the wavelet neural networks (WNN) method in [30] on the ISO New England data. The WNN method used a spike filtering technique to clear the spikes in load series. Then the spike-free load data with other inputs such as time indicators were sent to wavelet neural networks. The forecast range for comparison is from July 1, 2008 to July 31, 2008. Table 3 shows the 1-hour ahead forecasting results of the three methods. The results of ISO-NE and WNN methods are extracted from [30]. On the given testing data, the proposed method outperforms the WNN method, about 8.2% better in MAPE and 10.9% better in MAE. Compared to the ISO-NE method, the proposed method has significant improvements in both metrics. It should be noted that the proposed method employs a hybrid method ELM-MABC as the forecaster, which has better learning capability than the ordinary neural networks used in the ISO-NE and WNN methods.

Example 5: In this example, a comparison is performed between the proposed method and the standard neural network (NN) method and the similar day-based wavelet neural network (SIWNN) in [31]. The SIWNN method selected similar day load as the input data based on correlation analysis and used wavelet neural networks to capture the load features at low and high frequencies. The training period is from March 2003 to December 2005. The proposed method is used to predict the hourly load data from January 1, 2006 to December 31, 2006. Only 24-hour ahead forecasting is considered. The forecast results of the three methods are shown in Table 4. It is clear that the proposed method produces the

best forecast results. More precisely, the proposed method is 27.1% and 13.5% better than the standard NN and SIWNN methods, respectively.

Example 6: This example compares the proposed method to four other methods on the North American electric utility data [15, 19, 32, 33]. In [15], a hybrid forecast method composed of wavelet transform, neural network and evolutionary algorithm was proposed for STLF. Specifically, a two-step correlation analysis was integrated to select the most informative input variables. In [19], to overcome the border distortion problem, a novel load signal extension scheme was proposed, which is also used in our model. Each component from the decomposition was then forecasted separately. In [32], echo state network (ESN) was employed as a stand-alone forecaster to deal with the STLF problem. No lagged load and temperature input variables were involved in the model because of the special property of ESN. In [33], a parallel model consisting of 24 SVMs was proposed to conduct the day-ahead load forecast. The parameters of SVMs were optimized by the particle swarm pattern search method on the validation dataset.

The loads and temperatures from January 1, 1988 to October 12, 1992 are used to run experiments. The hourly loads for the two-year period prior to October 12, 1992 are forecasted. Both the hour ahead and day ahead load forecasts are considered. Moreover, the effect of noisy temperature is also studied. Table 5 compares the forecast results of the proposed method to other four methods proposed in [15, 19, 32, 33]. It can be observed that the proposed method can produce superior results to the other methods in all testing cases. With actual temperature data, the results of the proposed method and the method in [19] at different forecast horizons are shown in Fig. 4. It is clear that, for every horizon, the proposed method outperforms the method of [19].

Example 7: In this example, the effect of temperature forecasting error on STLF is further studied using the North American electric utility data. To cover a large scale of temperature errors, a set of Gaussian noises with different means and standard deviations are considered in 1-hour ahead load forecasting. The MAPE result (0.67) obtained with actual temperatures serves as the reference. Using different noises, the MAPE increments with respect to the reference are shown in Fig. 5. It is seen that the noises with large means or deviations will bring large load forecasting errors. For example, the forecasting error has a 12.2% rise when the noise is with mean 3 and standard deviation 3.6. The forecast results with zero-mean Gaussian noises are presented in Table 6. The associated temperature error ranges are also provided in the table. It can be noted that the proposed method is very robust to temperature errors. The forecasting error only climbs 9.61% when the temperature error varies in the largest interval $[-14.1\text{ }^{\circ}\text{C}, 15.2\text{ }^{\circ}\text{C}]$.

4. Discussion

The proposed method has obtained better forecast results in comparison with other well-established models in the literature on two publicly available datasets. There are several factors that contribute to the improved forecasting accuracy, such as the special learning mechanism of ELM, the integration of wavelet transform, the optimal parameters from MABC and the proper selection of input variables. The proposed method presents many advantages in STLF. Firstly, it can tackle the difficulty induced by the nonstationarity of load series in the electricity market. Secondly, it has strong robustness in terms of large temperature forecasting errors. Thirdly, it can produce accurate load predictions for electric utilities with different sizes and weather conditions. In addition, for the above examples, the

maximum training time of the proposed method is about 38 minutes. In contrast, the testing time of the proposed method only takes several seconds, which can be ignored.

5. Conclusion

This paper proposes a novel hybrid model for STLF based on the ELM. Two auxiliary techniques are developed to assist the ELM based forecasting method. Wavelet transform is used to decompose the load series into a set of different frequency components, which are more predictable. Moreover, a modified ABC algorithm is proposed to choose the optimal set of input weights and hidden biases for ELM. The ELM-MABC algorithm has better convergence performance than the conventional neuro-evolution method, leading to a significant improvement in forecasting accuracy. To confirm the effectiveness, the proposed hybrid method has been tested using the actual data from two public datasets. The simulation results reveal that the proposed method can produce excellent forecasting results beyond other well-established methods.

6. References

- [1] J.W. Taylor, Short-term load forecasting with exponentially weighted methods, *IEEE Trans. Power Syst.* 27 (2012) 458-464.
- [2] D.G. Infield, D.C. Hill, Optimal smoothing for trend removal in short term electricity demand forecasting, *IEEE Trans. Power Syst.* 13 (1998) 1115-1120.
- [3] S.J. Huang, K.R. Shih, Short-term load forecasting via ARMA model identification including non-Gaussian process considerations, *IEEE Trans. Power Syst.* 18 (2003) 673-679.
- [4] M. López, S. Valero, C. Senabre, J. Aparicio, A. Gabaldon, Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study, *Electr. Power Syst. Res.* 91 (2012) 18-27.
- [5] Y.M. Wi, S.K. Joo, K.B. Song, Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment, *IEEE Trans. Power Syst.* 27 (2012) 596-603.
- [6] S. Fan, L. Chen, Short-term load forecasting based on an adaptive hybrid method, *IEEE Trans. Power Syst.* 21 (2006) 392-401.

- [7] K. Kalaitzakis, G.S. Stavrakakis, E.M. Anagnostakis, Short-term load forecasting based on artificial neural networks parallel implementation, *Electr. Power Syst. Res.* 63 (2002) 185-196.
- [8] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst. Man Cybern. Part B* 42 (2012) 513-529.
- [9] X. Chen, Z.Y. Dong, K. Meng, Y. Xu, K.P. Wong, H.W. Ngan, Electricity price forecasting with extreme learning machine and bootstrapping, *IEEE Trans. Power Syst.* 27 (2012) 2055-2062.
- [10] R. Zhang, Z.Y. Dong, Y. Xu, K. Meng, K.P. Wong, Short-term load forecasting of Australian national electricity market by an ensemble model of extreme learning machine, *IET Gener. Transm. Distrib.* 7 (2013) 391-397.
- [11] N. Tai, J. Stenzel, H. Wu, Techniques of applying wavelet transform into combined model for short-term load forecasting, *Electr. Power Syst. Res.* 76 (2006) 525-533.
- [12] Z.A. Bashir, M.E. El-Hawary, Applying wavelets to short-term load forecasting using PSO-based neural networks, *IEEE Trans. Power Syst.* 24 (2009) 20-27.
- [13] Q.Y. Zhu, A.K. Qin, P.N. Suganthan, G.B. Huang, Evolutionary extreme learning machine, *Pattern Recognit.* 38 (2005) 1759-1763.
- [14] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, *J. Glob. Optim.* 39 (2007) 459-471.
- [15] N. Amjady, F. Keynia, Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm, *Energy*, 34 (2009) 46-57.
- [16] I. Daubechies, Ten lectures on wavelets, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
- [17] R.S. Pathak, The wavelet transform, Atlantis Press, Amsterdam, 2009.
- [18] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989) 674-693.
- [19] A.J.R. Reis, A.P.A. da Silva, Feature extraction via multiresolution analysis for short-term load forecasting, *IEEE Trans. Power Syst.* 20 (2005) 189-198.
- [20] G. Zhu, S. Kwong, Gbest-guided artificial bee colony algorithm for numerical function optimization, *Appl. Math. Comput.* 217 (2010) 3166-3173.
- [21] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: *IEEE Int. Jt. Con. on Neural Networks 2004*, pp. 985-990.
- [22] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, *Neurocomputing*, 70 (2006) 489-501.
- [23] P.L. Bartlett, The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network, *IEEE Trans. Inf. Theory* 44 (1998) 525-536.
- [24] R. Rajesh, J.S. Prakash, Extreme learning machines: a review and state-of-the-art, *Int. J. Wisdom Comput.* 1 (2011) 35-49.

- [25] S. McLoone, M.D. Brown, G. Irwin, G. Lightbody, A hybrid linear/nonlinear training algorithm for feedforward neural networks, *IEEE Trans. Neural Networks* 9 (1998) 669-684.
- [26] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, *IEEE Trans. Power Syst.* 16 (2001) 44-55.
- [27] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (2003) 1157-1182.
- [28] I. Drezga, S. Rahman, Input variable selection for ANN-based short-term load forecasting, *IEEE Trans. Power Syst.* 13 (1998) 1238-1244.
- [29] P. Shamsollahi, K.W. Cheung, Q. Chen, E.H. Germain, A neural network based very short term load forecaster for the interim ISO New England electricity market system, in: *22nd IEEE PES Int. Conf. Power Ind. Comput. Appl.* 2001, pp. 217-222.
- [30] C. Guan, P.B. Luh, L.D. Michel, Y. Wang, P.B. Friedland, Very short-term load forecasting: wavelet neural networks with data pre-filtering, *IEEE Trans. Power Syst.* 28 (2013) 30-41.
- [31] Y. Chen, P.B. Luh, C. Guan, Y. Zhao, L.D. Michel, M.A. Coolbeth, P.B. Friedland, S.J. Rourke, Short-term load forecasting: similar day-based wavelet neural networks, *IEEE Trans. Power Syst.* 25 (2010) 322-330.
- [32] A. Deihimi, H. Showkati, Application of echo state networks in short-term electric load forecasting, *Energy*, 39 (2012) 327-340.
- [33] E. Ceperic, V. Ceperic, A. Baric, A strategy for short-term load forecasting by support vector regression machines, *IEEE Trans. Power Syst.* 28 (2013) 4356-4364.

Figure captions

Fig. 1. Flowchart of ELM-MABC.

Fig. 2. Structure of the proposed STLF model.

Fig. 3. Convergence curves of NN-MABC and ELM-MABC.

Fig. 4. Hourly MAPE results of the proposed method and the method in [19].

Fig. 5. MAPE increments due to different Gaussian noises: means= $(-4, -3, -2, -1, 0, 1, 2, 3, 4)$ and standard deviations= $(0, 0.6, 1.2, 1.8, 2.4, 3.0)$.