Original Research

Short-Term Power Load Forecasting Based on a Combination of VMD and ELM

Wei Li*, Congxin Quan, Xuyang Wang, Shu Zhang

Department of Economics and Management, North China Electric Power University, Baoding, China

Abstract

Accurate short-term power load forecasting is becoming more and more important for the stable operation and improved economic benefits of electric power systems. However, when affected by various factors the power load shows non-linear and non-stationary characteristics. In order to forecast power load precisely, we propose an extreme learning machine (ELM) combined with variational mode decomposition (VMD), as a new hybrid time series forecasting model. In the first stage, since decomposed modes and hidden layer nodes have great influence on prediction accuracy, a three-dimensional relationship has been established to determine them in advance. In the second stage, using VMD, the time series of power load is decomposed into predetermined modes that are then used to construct training parts and forecast outputs. Then every individual mode is taken as an input data to the ELM. Eventually, in the third stage, the final forecasted power load data is obtained by aggregating the forecasting results of all the modes. To testify the forecasting performance of the proposed model, a five-minute power load data in Hebei of China is used for simulation, and comprehensive evaluation criteria is proposed for quantitative error evaluation. Simulation results demonstrate that the proposed model performs better than some previous methods.

Keywords: short-term power load forecasting, extreme learning machine (ELM), variational mode decomposition (VMD)

Introduction

In recent years, short-term power load forecasting is playing an increasingly important part in the stable operation of electrical power systems, which is essential for the development of preventive maintenance plans [1-2]. With increasingly market-oriented electric power production and consumption, more accurate short-term power load forecasting is needed for maintaining the secure and stable operation of the electrical power grid, which can also promote the sustainable development of the electricity industry [3]. Therefore, it is indispensable to develop power load forecasting techniques to achieve a more accurate simulation result [4-5].

Thus far, short-term power load forecasting methods mainly include conventional methods, intelligent methods, and hybrid forecasting methods. Conventional methods mainly include many mathematical statistics methods such as multiple linear regression analysis [6], the grey-forecasting model [7], autoregressive integrated moving average (ARIMA) [8-9], state space model [10], box-Jenkins model [11], general exponential smoothing [12], and so on. Although these methods have the advantage of capturing linear characteristics, they cannot accurately forecast short-term power load

---

*e-mail: ncepulw@126.com

showing non-linear and non-stationary characteristics. To capture nonlinear characteristics and obtain accurate prediction results, many intelligent methods such as artificial neural network (ANN) [13], support vector regression (SVM) [14-15], and intelligent evolutionary algorithms [16] have been proposed to forecast short-term power loads. ANN and SVM are well-known models, and they have obvious prediction advantages compared with conventional methods. Taking into account the fact that each single forecasting model has its own prediction shortcomings, an increasing number of researchers are adopting hybrid forecasting methods because they can avoid the shortcomings of each single forecasting model and deal with the complex problem more effectively.

ARIMA and Grey prediction models, as typical representatives of conventional methods, have been widely used and applied to solve many forecasting problems. The main idea of these models is to look for the trend of data changes, according to the trend equation, to forecast future data [17]. They are suitable to smooth and linear time series prediction, not good at grasping the trend of nonlinear sequence changes. With the rapid development of the artificial intelligence method, artificial intelligence forecasting technology has been gradually replacing conventional forecasting technology in the field of power load forecasting. Back propagation (BP) neural network, as a common ANN, has the ability to learn by itself, and it has been proposed by many researchers to apply to the short-term power load forecasting problem [18]. But the BP neural network is easily trapped in local minimum value [19]. Hence, many optimization algorithms have been proposed, but the computation burden is greatly increased. Compared with the BP neural network, the SVM algorithm has advantages of good robustness and strong generalization ability, avoiding falling into local minimum [20]. But the training time of the SVM increases exponentially by increasing the dimension of input vector. A new learning algorithm for the single-hidden-layer feed-forward neural network (SLFN) – called the extreme learning machine (ELM) – has recently been proposed [21]. In the learning process of ELM, since ELM requires only a single-pass training stage without any iteration for weight adjustment, it has very fast learning processing [22]. And it is able to solve the problem of stopping criteria, learning rate, learning epochs, and local minimum [23]. In recent years, the ELM has been applied in various fields and has become a significant method in nonlinear modeling [24-26].

Considering the non-stationary and chaotic properties of short-term power load data, the time series decomposition technique is applied to decompose the original power load sequence into several subsequences. Empirical mode decomposition (EMD), based on Hilbert-Huang transform (HHT), can effectively extract the components of the basic mode from non-linear or non-stationary time series. Owing to its attractive features, it has been applied to many economic and financial issues such as predicting crude oil prices [27], electricity prices [28], and so on. But EMD has many disadvantages, including lack of exact mathematical model, interpolation choice, and sensitivity to both noise and sampling [29]. Then a new multiresolution called variational mode decomposition (VMD) was introduced as an alternative to the EMD algorithm to overcome its limits. VMD has the ability to separate tones of similar frequencies, contrary to EMD [30]. And much literature [31-33] has come to the conclusion that VMD was more effective in modeling and forecasting economic and financial data compared with EMD.

Then in this paper, hybrid ELM-VMD intelligence is proposed for forecasting short-term power load. In particular, the number of modes decomposed by the VMD model makes a great influence on analysis, and the number of hidden layer nodes plays an important role in prediction accuracy. And both of them must be predetermined. So, how to determine the number of modes and the number of hidden layer nodes is a highlight of this paper. The point is specifically made in the remainder of this paper. After determining the number of modes and the number of hidden layer nodes, the VMD is applied to decompose power load data into several modes that have specific sparsity properties. Then ELM is constructed to forecast these modes individually. And all of these forecasting values are aggregated to produce the final power load forecasting results. Eventually, by contrast, the proposed method is compared with other alternative models such as the GM (1,1), ARIMA, single BP, single SVM, single ELM, hybrid BP-EMD, hybrid BP-VMD, and ELM-EMD models. Simulation results demonstrate that the proposed model performs more effectively and obtains more accurate prediction results.

Material and Methods

Variational Mode Decomposition (VMD)

The purpose of the VMD is to decompose the original signal into K discrete number of modes that have specific sparsity properties while producing the main signal. Thus, each mode u_k is required to be mostly compact around a center pulsation ω_k determined along with the decomposition. Compared with EMD, VMD has the ability to separate tones of similar frequencies. Meanwhile, the VMD model searches for a number of modes and their respective center frequencies, such that the band-limited modes reproduce the input signal exactly or in a least-squares sense. The VMD algorithm to assess the bandwidth of a one-dimensional signal is as follows [24]:

1) For each mode u_k, compute the associated analytic signal by means of the Hilbert transform to obtain a unilateral frequency spectrum.

2) For each mode, shift the mode's frequency spectrum to baseband (narrow frequency) by mixing with an exponential tuned to the respective estimated center frequency.

3) Estimate the bandwidth through Gaussian smoothness of the demodulated signal; for example, the squared L2-norm of the gradient.

Then, the decomposition process is realized by solving the following optimization problem [24]:

$$\min_{u_k, \omega_k} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{1}$$

$$\text{Subject to} \quad \sum_k u_k = f \tag{2}$$

...where f is the signal to be decomposed, u is its mode, ω represents the frequency, δ is the Dirac distribution, t represents time script, $*$ denotes convolution, $j^2 = -1$, and k is the number of modes. In this framework, the mode u with high-order k expresses low frequency components.

To address the constrained variational problem noted above, a quadratic penalty factor α and Lagrange multipliers $\lambda(t)$ are introduced. The combination of these two terms benefits from both considerable convergence properties of the quadratic penalty at finite weight and the strict enforcement of the constraint by the Lagrangian multiplier. Therefore, the above optimization problem is changed to an unconstrained one as below [24]:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \right.$$

$$\left. e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{3}$$

The optimization methodology denoted as the alternate direction method of multipliers (ADMM) is then used to obtain the saddle point of the augmented Lagrangian by updating u_k^{n+1}, ω_k^{n+1}, and λ_k^{n+1} alternately. The convergence criterion of the algorithm is $\sum_k (\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2) < \varepsilon$, where ε is the convergence tolerance. So, the final updated equations are given as follows [24]:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} u_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{4}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d_\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d_\omega} \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) = \lambda^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \tag{6}$$

...where n represents the iteration number and τ devotes the time step of the dual ascent. And the concrete procedures of the VMD algorithm are illustrated in Fig. 1.

Extreme Learning Machine

The extreme learning machine (ELM) was a new method for learning the single-layer feed-forward neutral network (SLFNs), which was first introduced by Huang et al. [21]. It tends to provide superior generalization performance and extremely fast learning speed by
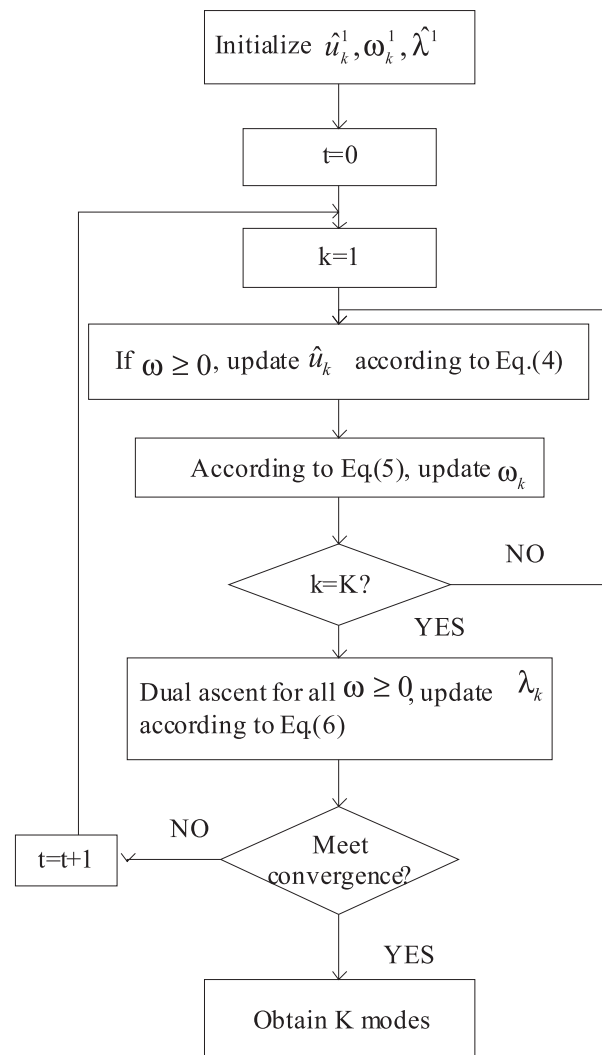

Fig. 1. Concrete procedures of the VMD algorithm.

$$f(W, b, X)$$

$$\eta$$

$$o$$

The input layer

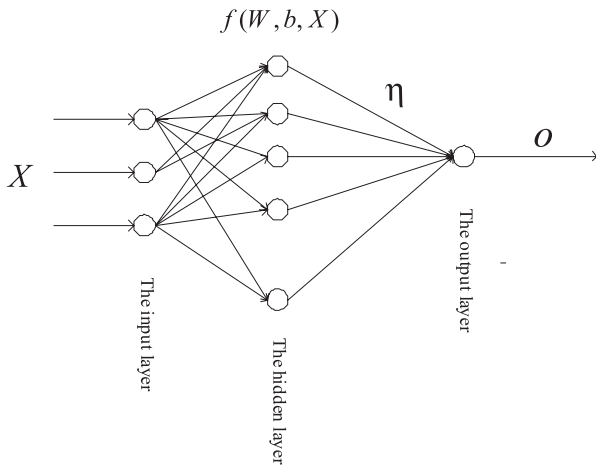The hidden layer

The output layer

Fig. 2. Typical structure of the ELM.

randomly choosing the input weights and analytically determining the output weights [24]. And only the number of hidden layer neurons is set, in which case the unique optimal solution will be obtained.

The typical structure of ELM is shown in Fig. 2, which includes input, hidden, and output layers. Assuming that the network has p input layer neurons and l hidden layer neurons, for N arbitrary distinct samples $(X_i, t_i) \in R^n \times R^m$ the SLFN with hidden neurons can be described as:

$$\sum_{j=1}^{l} f(W_j, b_j, X_i)\eta_j = o_i \quad i = 1, 2, ..., N \tag{7}$$

…where $W_j = [W_{i1}, W_{i2}, ..., W_{in}]^T$ is the input weight vector connecting the input nodes and the j th hidden node, $\eta_j = [\eta_{j1}, \eta_{j2}, ..., \eta_{jn}]^T$ is the weight vector connecting the jth hidden neuron and output neuron, X_i is the ith training example, $o_j = [o_{i1}, o_{i2}, ..., o_{in}]^T$ is the output vector, b_j is the bias of the j th hidden node, and $f(.)$ denotes the activation function of the hidden neuron.

If the above ELM can approximate the N samples with a zero error, then we can obtain:

$$\sum_{i=1}^{N} \|o_i - y_i\| = 0 \tag{8}$$

Hence parameters η_j, b_j, and W_j meet the conditions:

$$\sum_{j=1}^{l} f(W_j, b_j, X_i)\eta_j = t_i \quad i = 1, 2, ..., N \tag{9}$$

The above equation can be described as another form:

$$H\eta = T \tag{10}$$

…where

$$H(W_1, W_2, ..., W_l, b_1, b_2, ..., b_l, X_1, X_2, ..., X_N) =$$

$$\begin{bmatrix} f(W_1, b_1, X_1) & \cdots & f(W_l, b_l, X_1) \\ M & M & M \\ f(W_1, b_1, X_N) & L & f(W_l, b_l, X_N) \end{bmatrix}_{N \times l}$$

$$\eta = \begin{bmatrix} \eta_1^T \\ M \\ \eta_l^T \end{bmatrix}_{l \times m}, \quad T = \begin{bmatrix} t_1^T \\ M \\ t_N^T \end{bmatrix}_{N \times m}.$$

Different from the traditional function approximation theory, the input weights and hidden biases of the ELM are randomly generated. Thus, training the ELM is appropriately equal to find a least-square solution of the linear function $H\eta = T[]$.

$$\|H\hat{\eta} - T\| = \min_{\eta} \|H\eta - T\| \tag{11}$$

Then the solution of the above form is

$$\hat{\eta} = H^+ T \tag{12}$$

…where H^+ represents the Moore-Penrose generalized inverse of the hidden layer's output matrix H.

Evaluation Criterion

To quantitatively measure the performance of the proposed model, mean absolute percentage error (MAPE), root mean square error (RMSE), and the coefficient of determination (R^2) are used to calculate forecasting accuracy. The overall fitting effect of the proposed model can be evaluated by RMSE and R^2. The MAPE is also one important indicator, and the smaller the MAPE values, the closer the predicted power load time series values to those of the actual value. The definitions of the above-mentioned criteria are:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{14}$$

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

…where y_i and $\hat{y}_i$ represent the actual and forecasting power load values at time i, respectively; $\tilde{y}_i$ is mean value of the predicted value; and n is the number of the all load data.

Approaches of ELM-VMD Model

It is worth noting that both the number of hidden layer nodes l and the number of decomposed sub-signal K play an important role in prediction accuracy. Thus, in this paper it makes sense to find their optimal combinations, which this section discusses in detail.

Since the number of sub-signals decomposed by VMD requires it to be given in advance, the range of change varies depending on the number of IMFs obtained by EMD. For instance, if the number of IMFs is n, then parameter K used to determine the number of decomposition to obtain by EMD is $n-3$, $n-2$, $n-1$, n, $n+1$, $n+2$, and $n+3$ [34]. In order to avoid missing the optimal number of hidden layer nodes, in this paper the number of hidden layer nodes is set within a wide range – from 5 to 100. In this paper, predetermined training samples are divided into sub-training samples and sub-testing samples. MAPE, a benchmark for evaluating prediction accuracy, is used to determine the optimal combinations. Thus, there is a three-dimensional relationship among them. The specific steps are presented as follows:

1) Determine sub-training samples and sub-testing samples.
2) Apply VMD to decompose the original power load data into K stationary modes with different frequencies.
3) A rolling forecasting process is studied, and ELM is used to forecast each mode decomposed by VMD; in order to improve the forecasting accuracy of the ELM, all the original power load series and decomposed sub-series must be normalized primarily as follows:

$$\dot{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(16)

…where x is original sample data, $\dot{x}$ represents normalized sample data, and x_{max} and x_{min} represent the maximum and minimum of the data

4) The final power load forecasting result is obtained by aggregating all de-normalized values of the ELM outputs.
5) Calculate MAPE and obtain l and K with the minimum value of MAPE; then l and K are determined.

In fact, parameters l and K are obtained after $7 \times 96 = 672$ calculations. Then, using the parameters already obtained, repeat steps (2), (3), and (4) to forecast short-term power load. The flowchart of the forecasting procedure is shown in Fig. 3.
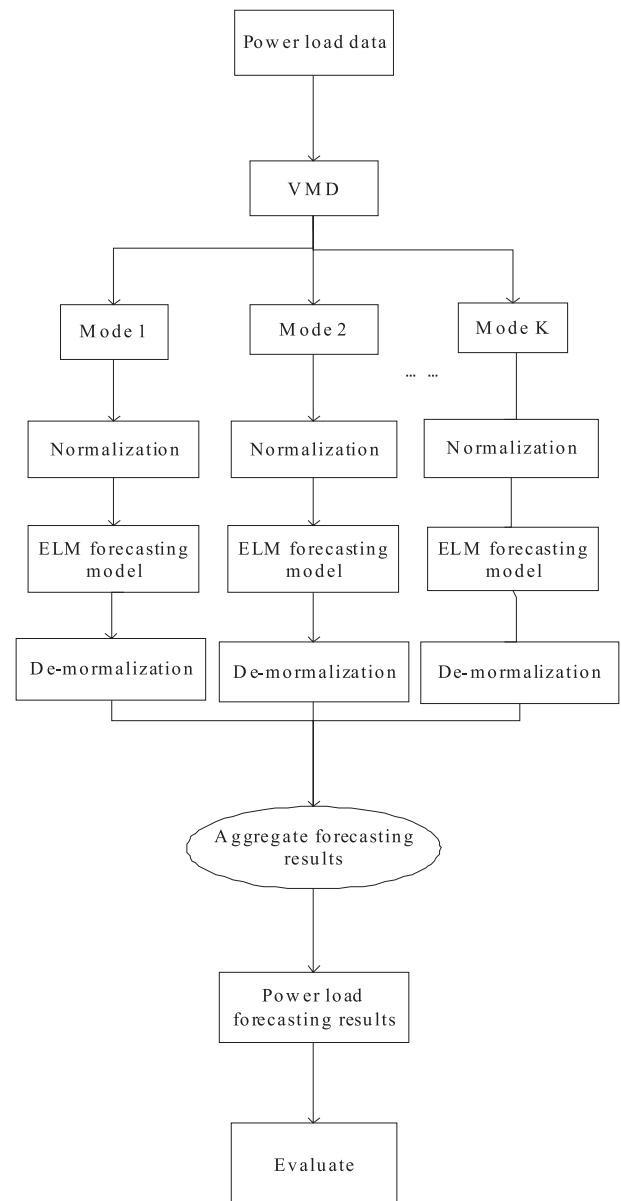

Fig. 3. Forecasting procedures of the ELM-VMD forecasting model.

Results and Discussion

Description of Data

To verify the effectiveness of the proposed ELM-VMD model, five-minute load data from Hebei Province in China was used in this case study. The power load data obtained from one day, with a total of 288 data points, was selected as experimental samples. A rolling forecasting process is studied in which the previous half-hour data is used to forecast the next data. That is, the first six load data points are used to forecast the seventh data, namely the last five load data (x_{n-6}, x_{n-5}, x_{n-4}, x_{n-3}, x_{n-2}, x_{n-1}) are used as the input variables of the ELM, and the output variable is x_n. The details are shown in Fig. 4.
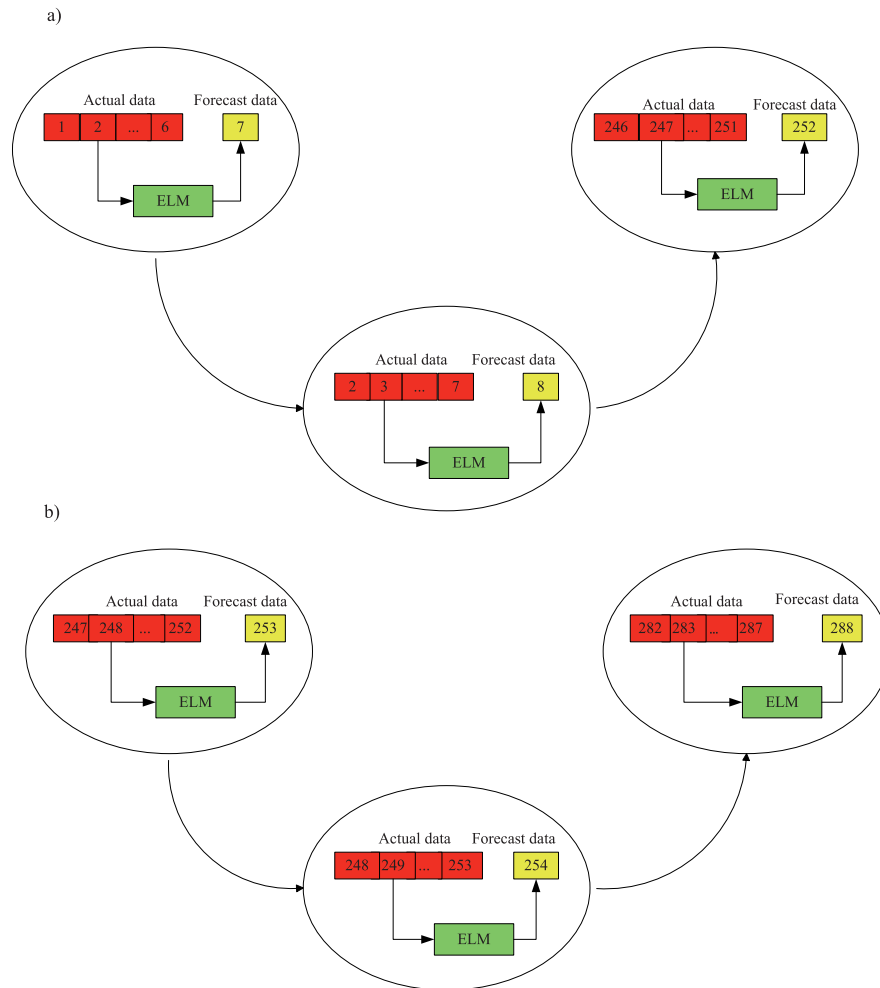
a)

b)

Fig. 4. A rolling-forecasting mechanism for short-term power load forecasting: a) for training set and b) for testing set.

In this paper, the previous 21 hours of data, including 252 data points, were used as the training set, and the remaining 36 data points were used as the testing set. To determine the parameters l and K, the previous 18 hours of the training set data were selected as sub-training set and the remaining data points were used as the sub-testing set. All computations are implemented in a MATLAB (2015a) environment on a computer with Intel core i5, 1.6 to 2.3 GHz CPU, and 4 GB of RAM.

Determine Parameters l and K

The fluctuation of short-term power load is severe, as is shown in Fig. 5. Thus, EMD is exploited to decrease the non-stationary characteristics of the original short-term power load time series. The results are shown in Fig. 6, in which the short-term power load data is decomposed into six different IMFs and one residue. It can be clearly seen that the frequencies of IMF1 and IMF2 are too high, which mainly indicates the random information of power load. The frequencies of IMF3 to IMF6 are moderate, which actually are the periodic components of the original short-term power load series. And the residue represents the long-term variation tendency of the original power load series.

Then, in parameter K the number of decomposed modes varies from 3 to 9, according to the number of IMFs decomposed by EMD. Since the decomposed modes are normalized between 0 and 1, the sigmoid is selected as activation function in this paper. The sub-training power load data is used to train ELM, and the sub-testing sample is used to calculate MAPE. Because the minimum value is not easy to observe in the image, the inverse of MAPE is introduced to have
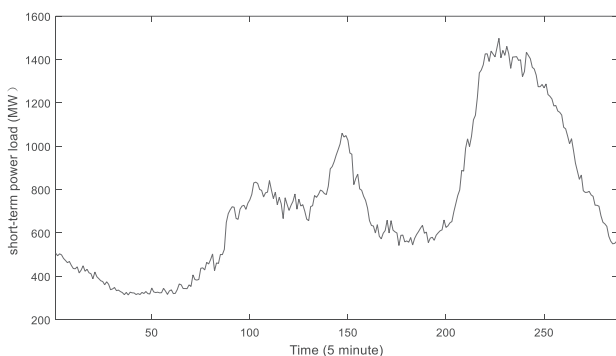

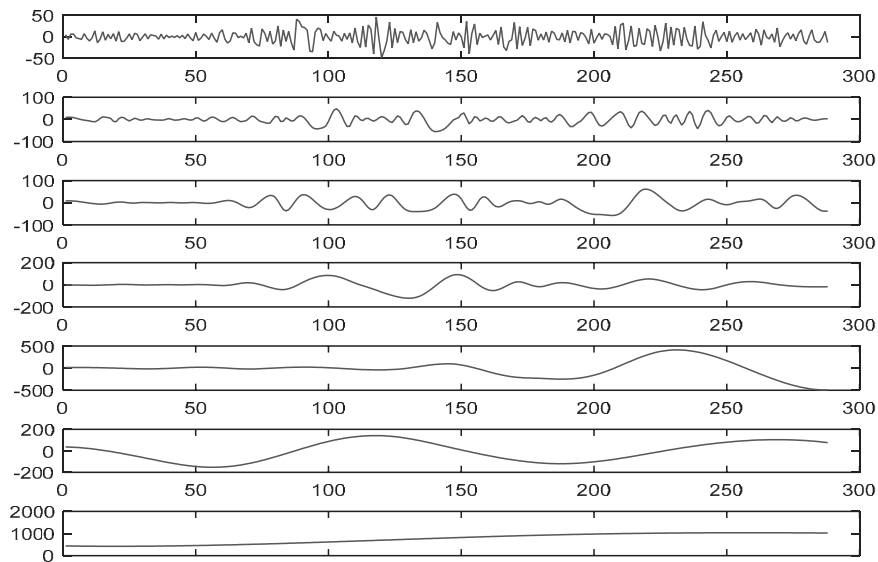Fig. 5. Original short-term power load time series.

Fig. 6. EMD decomposition results for short-term power load time series; from top to bottom: first IMF to residue.

a better representation and understanding of obtained results, which is shown in Fig. 7. It is clearly observed that the inverse of MAPE reaches a maximum with 8 decomposed modes together with 26 hidden layer nodes, namely achieving the best performance in this case. The other conclusions we can draw are that with the increase of decomposition levels, the prediction accuracy is basically increasing. After 8 decomposed levels, the prediction remains almost constant. Therefore, the decomposition of signal into 8 levels is the best choice, which simply increases the computational burden with more decomposition levels. And with the number of hidden layer nodes increasing, the prediction error first increases and then decreases. We can observe that the prediction effect will keep well when the hidden nodes range approximately from 10 to 50.

Then the original short-term power load series was decomposed into 8 modes by VMD (Fig. 8). Figure 8 shows that the first and second modes primarily represent the long-term trend of power load series. While the last two modes, with high frequency, capture short variations of power load series.
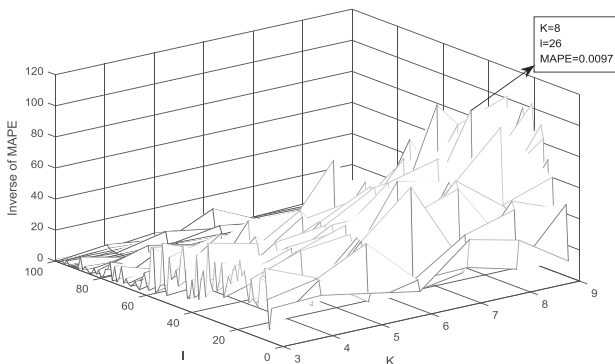

Fig. 7. Results of determining parameters using MAPE as error index.

Short-Term Power Load Forecasting Results

To verify the effectiveness of the ELM as forecasting part in this paper, many forecasting models existing in other literature such as GM (1,1), ARIMA, BP neutral network, and SVM are applied as contrast. Then the forecasting results are obtained by using original power load data that has not been decomposed. The parameters of BP neutral network are set as follows. Obviously, the input layer node and output layer node were set to 6 and 1, respectively. As a general rule of thumb, the hidden layer node was set to 10 in this paper. The maximum number of training steps was set to 100, the learning rate was set to 0.1, and the error precision was set to 1.0×10^5. As for SVM model, the sigmoid kernel function is selected to map the original feature space into a high-dimensional space. The epsilon in loss function of epsilon-SVM is set to 0.01. Owing to the randomness of prediction results for BP neural network, the BP neural network runs 10 times and the average value of the predicted results is used as the final result. Then, the obtained results are shown in Table 1 and the absolute error values are depicted in Fig. 9 for the above-mentioned forecasting tools. The forecasting errors of ARIMA and GM (1,1) are too large, so their absolute error values are not shown in the above Figure 9 for the rendering effects of the whole image. As can be seen from Table 1, the ELM has the best forecasting accuracy among all forecasting tools. The forecasting accuracy of GM (1,1) and ARIMA is far lower than other forecasting tools because they are suitable for stable and smooth time series. Compared with them, intelligence algorithms such as BP neutral network, SVM, and ELM are able to keep on learning, then they can deal with non-linear and non-stationary time series more effectively. The MAPE, RMSE, and R^2 of the ELM are superior to the BP neutral network because the inherent characteristics of the BP neutral network may fall
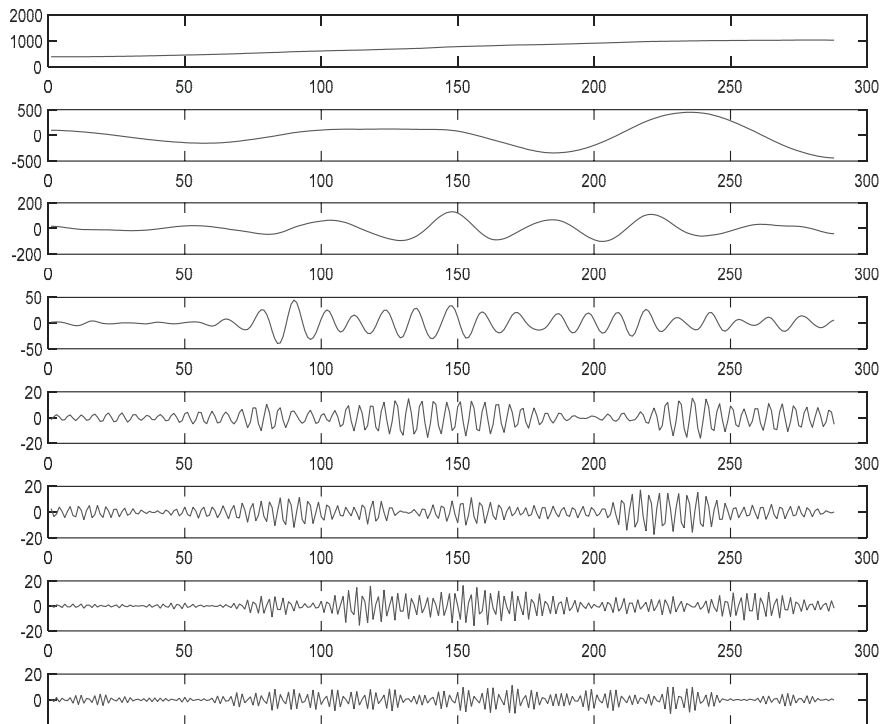
Fig. 8. VMD decomposition results for short-term power load time series; from top to bottom: first mode to eighth.

into local optimum and the hidden nodes of BP neural network highly depend on trial and error procedure. The SVM usually has very good forecasting performance when the input and output dimensions are high, but the SVM does not work well in this rolling prediction. And it even performs worse than the BP neutral network. From the standpoint of computing time using ELM, the computing time of the forecasting process is 0.0605s, which is very short among all applied forecasting tools. Although the computing time of GM (1,1) is rather short (0.0396 s), it has large forecasting errors. What's more, the SVM has the longest forecasting time of 44.0478s. If SVM combined with the VMD is proposed, the forecasting time will be approximately $44 \times 8 = 352$ s. Compared with the ELM-VMD, its forecasting time is much greater than the computing time of the ELM-VMD (approximately $0.6 \times 0.8 = 0.48$ s). Thus, the SVM model combined with other signal decomposed techniques was not taken into consideration because of its too-long calculation time.

Moreover, to demonstrate the superiority of applying VMD to signal analysis, EMD as a signal processing tool is introduced to provide comparison results. When it comes to the ELM-EMD model, to be consistent with the proposed method, using sub-training set and sub-testing set to gain the predetermined optimal hidden nodes. The result is shown as Fig. 10. Thus, in this condition the hidden nodes of the ELM were set to 11. When it comes to the BP-VMD model, the predetermined number of decomposition K ranges from 3 to 9, the BP neural network also runs 10 times and obtains average value.

The obtained forecasting errors of hybrid forecasting methods are given in Table 2. It can be seen that the MAPE and RMSE of the hybrid ELM-VMD model are the smallest and the goodness of fit reaches 0.998, which is the best of all models. As Tables 1 and 2 show, the hybrid models have better prediction accuracy than the corresponding single format. This indicates that the forecasting performance will be improved after

Table 1. Error statistics for single forecasting model.

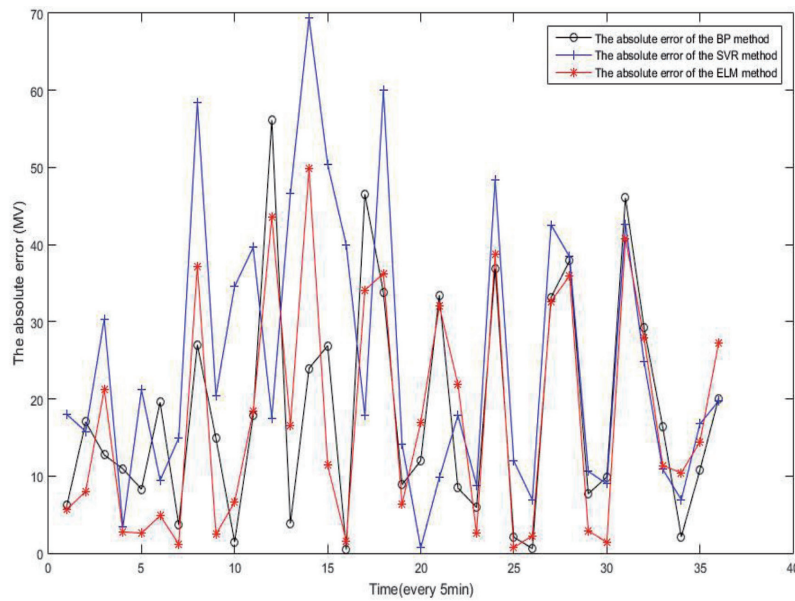| Forecasting model	Error index			Computing time (s)
	MAPE (%)	RMSE (MW)	R^2(MW)	
GM(1,1)	84.59	24.469	0.9774	0.0396
ARIMA	71	22.4301	0.9272	32.2625
BP	2.66	4.6517	0.9869	17.989
SVM	3.07	5.024	0.9904	44.0478
ELM (proposed model)	2.42	4.3732	0.9906	0.0605

Fig. 9. Absolute error of single models in the short-term power load forecasting.

original data are decomposed by signal processing tools. It is worth noting that the BP-EMD model has worse performance than the single BP model because the BP neutral network may fall into the local minimum value. The forecasting finite IMFs and residue are added to obtain the final result, and the forecasting error may increase. As shown in Table 2, the VMD-based approach performs better than the EMD-based approach in terms of MAPE and RMSE, which illustrates the advantage of the VMD-based approach for forecasting short-term power load. One possible reason is that the EMD has the disadvantage of mode mixing, while the EMD has the ability to separate tones of similar frequencies. Table 2 also showed that the ELM-VMD model outperformed the BP-VMD model in terms of MAE, RMSE, and R^2

for all values of K. The BP-VMD model performs better than the ELM-EMD model in terms of MAPE and RMSE, with $K = 6,7,8,9$. This also means that signal preprocessing has great influence on prediction accuracy. The absolute error values are also depicted in Fig. 11. For diverse decomposed modes, it has different forecasting results of the BP-VMD. The case of the best performance with $K = 8$, as the represented case of all cases for the BP-VMD model is shown in Fig. 11.

In summary, the experimental results fully demonstrate the efficiency of ELM in power load forecasting and the advancement of VMD in signal processing. At the same time, it confirms that establishing three-dimensional relationships to determine coefficients l and K is a good idea to obtain
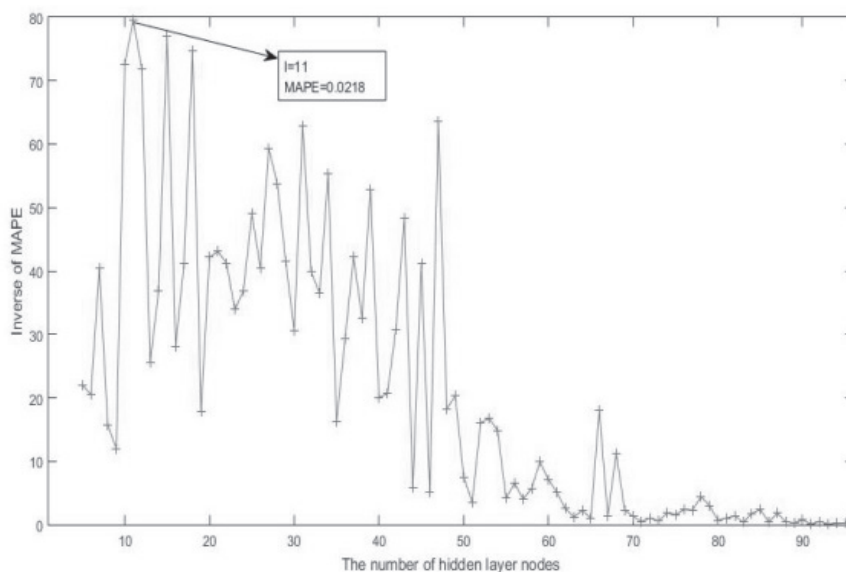

Fig. 10. Result of determining hidden nodes using MAPE as error index.

Table 2. Statistics error for all hybrid forecasting models.

Hybrid model		Error index		
		MAPE (%)	RMSE (MW)	R²(MW)
BP-EMD		3.32	4.6295	0.9938
BP-VMD	K = 3	2.41	4.286	0.9916
	K = 4	2.14	4.0633	0.9938
	K = 5	2.23	4.1313	0.9935
	K = 6	1.5	3.4094	0.9963
	K = 7	1.55	3.457	0.9963
	K = 8	1.18	3.0533	0.9977
	K = 9	1.21	3.0804	0.9977
ELM-EMD		0.0218	3.8119	0.997
ELM-VMD (proposed hybrid model)		0.0107	2.9356	0.998

accuracy forecasting results. And it also indicates that the proposed model is effective and efficient for short-term power load forecasting. In fact, the forecasting process time of ELM-VMD is approximately 0.48s, and its computational time is much smaller than other hybrid models.

Conclusions

In this paper, a new hybrid short-term power load forecasting model based on ELM and VMD has been proposed. First of all, it uses a sub-training set and sub-testing set, establishing three-dimensional relationships, to obtain the number of variational modes and hidden nodes of ELM. Then, in consideration of the volatility of power load, the original load series is decomposed into predetermined modes. Next, rolling prediction is studied step by step and the ELM is applied to forecast individual modes. Finally, the different forecasting series are reconstructed to get the final forecasting results.

Experiments with different statistical criteria (MAPE, RMSE, R^2) clearly testify that the ELM-VMD model achieved the lowest forecasting error.
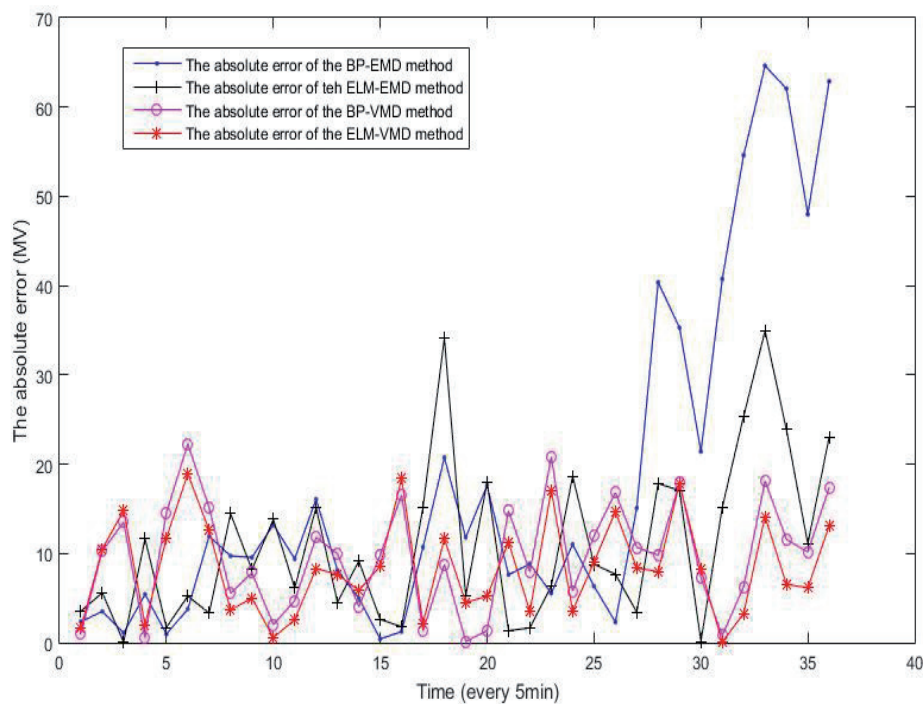

Fig. 11. Absolute error of all hybrid models in short-term power load forecasting.

Compared with other hybrid or single models, it has a very fast forecasting process and saves significant computational time, which means that the ELM-VMD model can be used as a very promising methodology for short-term power load prediction. In this paper, several other conclusions can be acquired as follows:

A) Establishing three-dimensional relationships to obtain the number of variational modes and hidden nodes of ELM, as a highlight of this paper, achieves fine performance.

B) As a new adaptive multiresolution technique, VMD has better performance and it is more robust for analyzing noisy signals than EMD.

C) The intelligent algorithm is more suitable for predicting the non-stationary and nonlinear time series than conventional method.

D) Compared with other forecasting algorithms in the hybrid models, the ELM has the lowest forecasting error and the shortest computation time. As a result, the proposed hybrid model in this paper is simple to implement and it provides fast learning and convergence when the sample size is large.

The whole process – from selecting optimal parameters to obtaining the final forecasting result – is easy to understand. Thus, all of these reasons make it more suitable for forecasting short-term power load and even other time series.

However, in this paper, only the one-step-ahead forecasting model is constructed, with regards to future research directions, to construct the multi-step-ahead forecasting model should also be taken into account. Besides, short-term power load forecasting is equivalent to time series forecasting in this paper, and other factors such as temperature, wind power, and precipitation are excluded from the proposed hybrid model. Thus, a future study should verify the superiority of the proposed model in multi-step-ahead forecasting and incorporate those influencing factors to develop the comprehensive hybrid forecasting model.

Acknowledgements

References

1. SANSTAD A.H., MCMENAMIN S., SUKENIK A., BARBOSE G.L., GOLDMAN C.A. Modeling an aggressive energy-efficiency scenario in long-range load forecasting for electric power transmission planning. Applied Energy , **128** (3), 265, **2014**.

2. LI C., LI, S., LIU Y. A least squares support vector machine model optimized by moth-flame optimization algorithm for annual power load forecasting. Applied Intelligence **45** (4), 1166, **2016**.

3. WANG J., LI L., NIU D., TAN Z. An annual load forecasting model based on supportvector regression with differential evolution algorithm. Applied Energy, **94** (6), 65, **2012**.

4. NIU D., WANG Y., DUAN C., XING M. A New Short-term Power Load Forecasting Model Based on Chaotic Time Series and SVM. Journal of Universal Computer Science, **15** (13), 2726, **2009**.

5. ZHOU X., ZHOU X., ZHANG Z.M., TENTZERIS M.M. Short-term power load forecasting using grey corelation contest modeling. Expert Systems with Applications, **39** (1), 773, **2012**.

6. HONG T., GUI M., BARAN M.E., WILLIS H.L. Modeling and forecasting hourly electric load by multiple linear regression with interactions. Power & Energy Society General, Meeting, 1-8, **2010**.

7. LI W., HAN Z.H. Application of improved grey prediction model for power load forecasting. International Conference on Computer Supported Cooperative Work in Design, 1116, **2008**.

8. NIE H., LIU G., LIU X., WANG Y. Hybrid of ARIMA and SVMs for Short-Term Load Forecasting. Energy Procedia, **16** (5), 1455, **2012**.

9. NAZARKO J., JURCZUK A., ZALEWSKI W. ARIMA models in load modelling with clustering approach. Power Tech, IEEE Russian, 1-6, **2005**.

10. DORDONNAT V., KOOPMAN S.J., OOMS M., DESSERTAINE A., COLLECT J. An hourly periodic state space model for modelling French national electricity load. International Journal of Forecasting, **24** (4), 566, **2008**.

11. VÄHÄKYLA P., HAKONEN E., LÉMAN P. Short-term forecasting of grid load using Box-Jenkins techniques. International Journal of Electrical Power & Energy Systems, **2** (1), 29, **1980**.

12. CHRISTIAANSE W.R. Short-term load forecasting using general exponential smoothing. IEEE Transactions on Power Apparatus & Systems, **90** (2), 900, **1971**.

13. XIA F., FAN L. Application of Artificial Neural Network (ANN) for Prediction of Power Load. Springer Berlin Heidelberg, **115**, 673, **2012**.

14. WU Q. Power load forecasts based on hybrid PSO with Gaussian and adaptive mutation and W v -SVM. Expert Systems with Applications, **37** (1), 194, **2010**.

15. LIU B., YANG R. A novel method based on PCA and LS-SVM for power load forecasting. International Conference on Electric Utility Deregulation & Restructuring & Power Technologies, 759, **2008**.

16. TZAFESTAS S., TZAFESTAS E. Computational Intelligence Techniques for Short-Term Electric Load Forecasting. Journal of Intelligence & Robotic Systems, **31** (1-3), 7, **2001**.

17. HENG J., WANG C., ZHAO X., WANG J. A Hybrid Forecasting Model Based on Empirical Mode Decomposition and the Cuckoo Search Algorithm: A Case Study for Power Load. Mathematical Problems in Engineering, **2016**.

18. LI L.J., HUANG W. A Short-Term Power Load Forecasting Method Based on BP Neural Network. Applied Mechanics & Materials, **494-495**, 1647, **2014**.

19. HONG W.C. Chaotic particle swarm optimization algorithm in a support vector regression electric load

forecasting model. Energy Conversion & Management, **50** (1), 105, **2009**.

20. NIU D., DAI S. A Short-Term Load Forecasting Model with a Modified Particle Swarm Optimization Algorithm and Least Squares Support Vector Machine Based on the Denoising Method of Empirical Mode Decomposition and Grey Relational Analysis. Energies, **10**, **2017**.

21. HUANG G.B., ZHU Q.Y., SIEW C.K. Extreme learning machine: Theory and applications. Neurocomputing, **70** (1), 489, **2006**.

22. ABDOOS A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. Neurocomputing, **203**, 111, **2016**.

23. FIGUEIREDO E.M.N., LUDERMIR T.B. Investigating the use of alternative topologies onperformance of the PSO-ELM. Neurocomputing, **127** (3), 4, **2014**.

24. XIA M., ZHANG Y., WENG L., YE X. Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. Konwledge-Based Systems, **36** (6), 253, **2012**.

25. ZHANG C., ZHOU J., LI C., FU W., PENG T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Conversion & Management, **143**, 360, **2017**.

26. WANG X., ZHANG H., GUO X. Demand Forecasting Models of Tourism Based on ELM. Seventh International Conference on Measuring Technology & Mechatronics Automation, 326, **2015**.

27. HE K., ZHA R., WU J., LAI K. Multivariate EMD-Based Modeling and Forecasting of Crude Oil Price. Sustainability, **8** (4), 367, **2016**.

28. HE K., YU L., TANG L. Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology. Energy, **91**, 601, **2015**.

29. DRAGOMIRETSKIY K., ZOSSO D. Variational Mode Decomposition. IEEE Transactions on Signal Processing, **62** (3), 531, **2014**.

30. LAHMIRI S. Comparing Variational and Empirical Mode Decomposition in Forecasting Day-Ahead Energy Prices. IEEE Systems Journal, **99**, 1, **2015**.

31. SUN G., CHEN T., WEI Z., SUN Y., ZANG H. A Carbon Price Forecasting Model Based on Variational Mode Decomposition and Spiking Neural Networks. Energies, **9** (1), 54, **2016**.

32. YANG W., PENG Z., WEI K., SHI P., TIAN W. Superiorities of variational mode decomposition over empirical mode decomposition particularly in time-frequency feature extraction and wind turbine condition monitoring. Let Renewable Power Generation, **11** (4), 443, **2017**.

33. HUANG G.B., ZHU Q.Y., SIEW C.K. Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE International Joint Conference on Neural Networks, **2**, 985, **2004**.

34. LAHMIRI S. A Variational Mode Decompoisition Approach for Analysis and Forecasting of Economic and Financial Time Series. Expert Systems with Applications, **55**, 268, **2016**.