BENTHAM OPEN

CrossMark

The Open Electrical & Electronic Engineering Journal
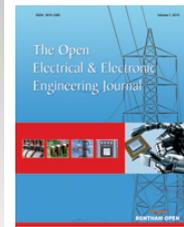
RESEARCH ARTICLE

Short-term Wind Power Prediction Using GA-ELM

Xinyou Wang[1,*], Chenhua Wang[2] and Qing Li[3]

[1] *Institute of Technology, Gansu Radio & TV University, Lanzhou 730030, P.R. China*

[2] *Northwest Engineering Corporation Limited, PowerChina, Xi'an 710065, P.R. China*

[3] *State Grid Xinjiang Electric Power Company, Electric Power Research Institute, Grid technology Center, Urumqi 830000, P.R. China*

Abstract: Focusing on short-term wind power forecast, a method based on the combination of Genetic Algorithm (GA) and Extreme Learning Machine (ELM) has been proposed. Firstly, the GA was used to prepossess the data and effectively extract the input of model in feature space. Basis on this, the ELM was used to establish the forecast model for short-term wind power. Then, the GA was used to optimize the activation function of hidden layer nodes, the offset, the input weights, and the regularization coefficient of extreme learning, thus obtaining the GA-ELM algorithm. Finally, the GA-ELM was applied to the short-term wind power forecast for a certain area. Compared with single ELM, Elman algorithms, the experimental results show that the GA-ELM algorithm has higher prediction accuracy and better ability for generalization.

INTRODUCTION

Wind power, as a green renewable energy resource, has gained more and more significance in the recent years around the world. With the rising wind power capacity in wind farm, the penetration of wind resources in power system has been increasing in the recent years. However, wind power is characterized as intermittent with stochastic fluctuations, which can pose significant challenges to peak load regulation [1 - 3].Thus, highly efficient, accurate wind power forecasting is crucially important for maintaining the power balance and economic operation of the power system.

Generally speaking, the prediction methods for wind power can be summarized as physical approaches, statistical approaches and hybrid approaches based on computational intelligence [4 - 6], *etc*. Physical method is based on numerical weather prediction (NWP) using weather forecast data like wind speed, wind direction, pressure and temperature. Physics information usually obtained from the local meteorological service and transformed to the wind turbines at the wind farm is converted to wind power by motor power curves [7].Statistical method is based on mapping relations between wind speed, wind direction and output data. Typically, time series analysis approaches and some artificial intelligence approaches are involved [8, 9]. Computational intelligence method is based on algorithm as wavelet analysis, artificial neural network (ANN) and support vector machines (SVM). Normally, the nonlinear relationship between the input and the output is described from historical time-series in wind power, thus the models for wind power prediction are obtained [10 - 13].

Extreme learning machine [14, 15], proposed by Huang et al. (2006), is a sort of single-hidden layer feedforward neural networks (SLFNs). In ELM, the SLFN weights and biases are randomly initialized, and the output weight is determined then. Essentially, its hidden layer does not need to be tuned. Hence, compared to some classical methods,

---

* Address correspondence to this author at the Institute of Technology, Gansu Radio & TV University, Lanzhou, P.R. China; Tel: 18293131209; E-mail: wangxiny@gsrtvu.cn

ELM learns much faster and higher generalization performance. Furthermore, its implementation is easy, which avoids many difficulties faced by gradient-based learning methods such as learning epochs, learning rate, *etc*.

Wind power time series are characteristics of nonstationarity and intermittency, due to stochastic nature effect. Thus, preprocess data is aid to improve the performance of prediction model. Since in ELM, the input weight, bias, regularization coefficient and other initial parameters are randomly determined, there is significant influence on the fitting performance, the convergence rate and the prediction accuracy. This paper has proposed a new hybrid approach based on computational intelligence which combines genetic algorithm with ELM. Data preprocessing is conducted with GA algorithm, for the extraction of the model dimension. And the activation function of hidden layer, as well as bias, input weights and regularization coefficient are optimized with GA algorithm. The proposed model has been validated by using data obtained from the National Renewable Energy Laboratory (NERL) and compared with the Elman model and ELM model to show its superiority.

1. GA METHOD

Genetic Algorithm is simulated Darwin evolution of Natural selection and Genetic mechanism of biological evolution process calculation model which was proposed by J. Holland professor in 1975 [16]. It is a kind of thorough simulation of Natural evolution search optimal solution method for some complex problems. Parameter coding, identification of initial group, fitness function, genetic operation and control parameters are the core context of genetic algorithm [17 - 19]. Genetic operation mainly includes three operators: selection operation, cross operation and variation operation. Control parameters mainly include the size of group, the probability of the genetic operation, *etc*. The particular algorithm process is as shown in Fig. (**1**).
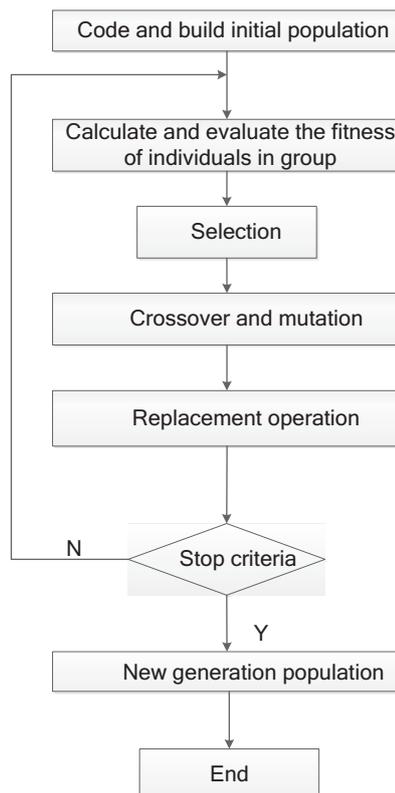
Fig. (1). Genetic algorithm basic operation flowchart.

Firstly, points in feasible region were encoded. Then, a random group code (chromosomes or individuals) was chosen and set as the initial group, and the individual fitness of each code was calculated. The fitness represents the optimization information of target function here. Based on the fitness of individuals in the group on the basis of evaluation, some individuals were chosen as the samples assemble before reproduction process, under a selection mechanism. The individuals with higher fitness were pledged to maintain more samples, while the individuals with

slower fitness had smaller samples in the selection mechanism. In the reproduction process, chosen sample was changed under certain crossover rate and mutation rate using crossover and mutation operators, thus generating a new individual. Finally, the next generation group was generated by the replacement of the new individuals from the olds. The algorithm keeps repeating the evolution, selection, reproduction and replacement operation until the termination condition judgment is satisfied.

The computational steps of the ELM optimized by using GA are explained as follows:

Step 1: Select coding strategy and convert the parameters to strings, and construct an initial group with individuals randomly selected t, as follows $s = \{X_1, X_2, \dots X_N\}$, $X_i = (x_{i,1}, x_{i,2}, \dots x_{i,n})$, while n represents the dimension of the solution space. The value of each individual's fitness is calculated based on the fitness of function.

Step 2:
   a. Selection operation: According to the reproduction rate $p_r(f)$, select an individual X_i for reproducing N times. The more the fitness, the higher the reproduction rate.
   b. Cross operation: The population forms N^2 pair chromosomes after reproduction. Under certain crossover rate p_c, use crossover operator to cross genetic code in each chromosome for the generation of new chromosomes.
   c. Variation operation: according to a certain probability p_m, certain places in the chromosome go breaking.

Step 3: Repeat the above a,b,c until getting enough individuals to generate new generation.

Step 4: Go back to step 2 if termination condition judgment is not satisfied. The termination of a computation is obtained if satisfied.

2. ELM METHOD

As a simple and efficient learning algorithm, ELM is a single hidden-layer feed forward neural network extended to the generalized SLFNs. In ELM, only the number of the hidden layer node needs tuning, while the input weights and hidden layer biases do not need to be adjusted. Moreover, by the use of the Moore–Penrose generalized inverse for solution of network weights, smaller weights are the norms, avoiding several issues like local minima, improper learning rate and overfitting, *etc.* in gradient descent-based learning methods. Therefore, it provides not only extremely fast learning speed but also good generalization performance.

For SLFNs with $\tilde{N}$ hidden nodes, given N learning sample matrix $\{x_i, y_i\} \in R^n \times R^m$, where $x_i = [x_{i1}, x_{i2}, \dots x_{in}]^T$, $y_i = [y_{i1}, y_{i2}, \dots y_{im}]^T$, $i = 1, \dots N$ so the node output takes the form as

$$g(x) = \sum_{i=1}^{\tilde{N}} \beta_i h_i(x) = \sum_{i=1}^{\tilde{N}} \beta_i G(x; w_i, b_i), x \in R^n, \beta_i \in R^m \tag{1}$$

Where w_i, b_i are the learning parameters of the network hidden layer node, $\beta_i \in R^m$ is the weight vector connecting the ith hidden node and the output nodes, $h_i(x)$ is the output function $G(x; w_i, b_i)$ of the ith hidden node.

Suppose SLFNs can approximate these N samples with zero error means that $\sum_{j=1}^{N} \|o_j - y_j\| = 0$, there β_i, w_i and b_i such that:

$$g(x) = \sum_{i=1}^{\tilde{N}} \beta_i G(x_j; w_i, b_i) = y_j, j = 1, \dots, N \tag{2}$$

Equations (2) can be written in matrix form as:

$$H\beta = Y \tag{3}$$

Where
$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} K(w_1 \cdot x_1 + b_1) & \cdots & K(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ K(w_1 \cdot x_N + b_1) & \cdots & K(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m}$$

Where, H is called the hidden layer output matrix of the neural network; the ith column of H is the ith hidden node output with respect to inputs $x_1, x_2, ... x_N$; the ith row of H is the feature mapping of hidden layer with respect to inputs x_i, namely $x_i : h(x_i)$, and $h(x_i) = \left[G(x_i; w_1, b_1), ..., G(x_i; w_{\tilde{N}}, b_{\tilde{N}}) \right].$

For any infinitely differential activation function $G(x; w_i, b_i)$, if the hidden layer node and node parameters can be generated randomly, the maximum number of the hidden layer node $\tilde{N}$ should be less than the sample number N, with the interpolation views. In fact, when $\tilde{N} = N$, the training error is equal to zero. When $\tilde{N} < N$, SLFNs can still approximate these training samples with tiny error, while the matrix H is not square matrix, thus $\hat{\beta}_i$, $\hat{w}_i$ and $\hat{b}_i$ exist, enabling

$$\left\| H(\hat{w}_1, ..., \hat{w}_{\tilde{N}}, \hat{b}_1, ..., \hat{b}_{\tilde{N}}) \hat{\beta} - T \right\|$$
$$= \min_{w_i, \beta_i, b_i} \left\| H(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}) \hat{\beta} - T \right\| \tag{4}$$

Since $G(x; w_i, b_i)$ is infinitely differentiable in any interval, equations (4) can be linear systems, thus the training of the ELM is equivalent to solve the equation (3) for its least square solution, we have

$$\left\| H(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}) \hat{\beta} - T \right\|$$
$$= \min_{\beta_i} \left\| H(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}) \hat{\beta} - T \right\| \tag{5}$$

So, the smallest norm least squares solution of the output weight matrix is

$$\hat{\beta} = H^{\dagger} Y \tag{6}$$

To improve the stability and generalization performance, regularization coefficient is introduced for its regularization least square solution with the idea of ridge regression.

Therefore, the regularization least square solution of output weight matrix β for Equation (3) takes the form as

$$\hat{\beta} = H^T \left(HH^T + \frac{I}{C} \right)^{-1} Y \tag{7}$$

Resulting from Equation (7), the network output of the ELM can be expressed as

$$g(x) = h(x)\hat{\beta} \tag{8}$$

3. GA-ELM METHOD

Firstly, the data pretreatment is carried out using GA for efficient extraction of input dimension for model in the feature space. On the basis of this, a model based on ELM for short-time wind power prediction is constructed. Then, GA is used to optimize the type, bias, input weight and regularization coefficient of the activation function for ELM, thus the hybrid prediction model for wind power based on GA-ELM is obtained. The GA-ELM algorithm adopts the following steps:

Step 1: Normalize the sample data, and map the training sample set $\{x_i, y_i\} \in R^n \times R^m$ to the interval [0,1].
Step 2: Optimize the input dimension of the sample set $x_i = [x_{i1}, x_{i2}, ... x_{in}]^T$.
Step 3: Construct the wind power prediction model based on ELM. In this paper, sigmoid function and linear function are chosen as activation functions of the hidden layer node.
Step 4: Optimize the type of activation function $s_j^\lambda \in \{0, 1, 2\}, j = 1, \cdots, h$, regularization coefficient C, input weight $w_i \in [0,1]$, bias $b_i \in [0,1]$, and construct the wind power prediction model based on GA-ELM.
Step 5: Apply the GA-ELM to a short-time wind power prediction in a special region. The model is constructed by the time-series method, that is

$$\hat{y}_{t+h} = f(x_t), \forall t = \Delta...l \tag{9}$$

Where $f(x_t)$, is constructed by ELM method, Δ represents the embedding dimension of prediction model, x_t is the multiple-dimension input vector constructed by the historical wind power values $(y_{t-1}, y_{t-2},......, y_{t-\Delta})$.

4. SHORT-TIME WIND POWER PREDICTION EXPERIMENTS

The proposed GA-ELM model is applied for Western dataset supplied by American National Renewable Energy Laboratory (NREL). As in [20], the mean absolute percentage error (MAPE), max error (ME), and root mean square error (RMSE) are used to measure the prediction performance. The definitions are expressed as

$$MAPE = \frac{1}{k}\sum_{i=1}^{K}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{10}$$

$$MAXERROR = \max\left|y_i - \hat{y}_i\right| \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{K}\left(y_i - \hat{y}_i\right)^2}{k}} \tag{12}$$

Where, y_i is the real power value at the time of prediction, $\hat{y}_i$ is the prediction value of the model, k is the sample number of the testing set.

4.1. Validation experiment Result I

All of the data in this experiment come from NREL [21]. The modeled data is sampled every 10 min temporally. 1200 data are randomly selected for experiments, and the fore 80 percent data are used for training, the remnant are used for testing. A 2h-head multi-step prediction model is constructed, according to the Equation (13) as follows:

$$\hat{y}_{t+2h} = f(x_t), \forall t = \Delta..l \tag{13}$$

In present experiment, the node number of the ELM hidden layer was set as 120, the maximum iteration was selected as 60 and the maximum group number was set as 100. The proposed GA-ELM model was compared with the ELM and Elman models to further evaluate its performance. The results are shown in Table **1** and Fig. (**2**).

Table 1. Comparison among the Elman, ELM, and GA-ELM models.

Prediction method	MAPE	RMSE	MAXERROR
Elman	11.0096	2.1819	13.2566
ELM	15.5132	2.9079	18.0702
GA-ELM	6.0790	1.4799	11.1989
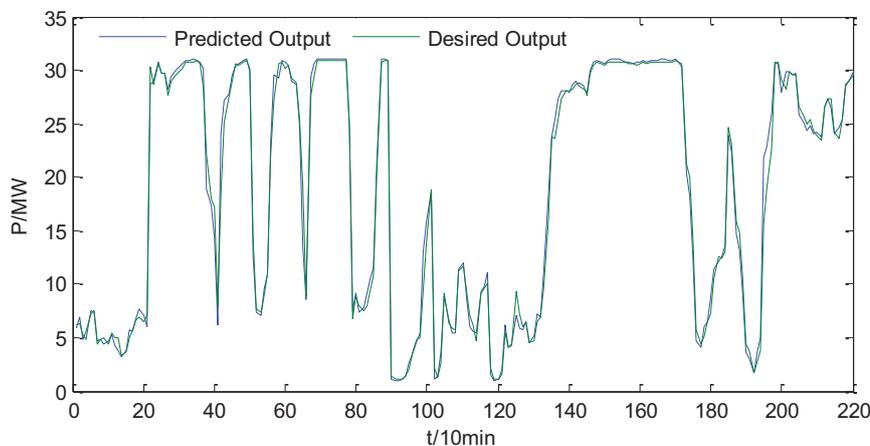
Fig. (2). Comparison of the real value and the 2-h-ahead prediction value based on GA-ELM.

From Table **1**, it is obvious that the MAPE, RMSE and MAXERROR values of the GA-ELM are all smaller than those of Elam and ELM. In addition, as shown in Fig. (**2**), the proposed GA-ELM has better fitting performance.
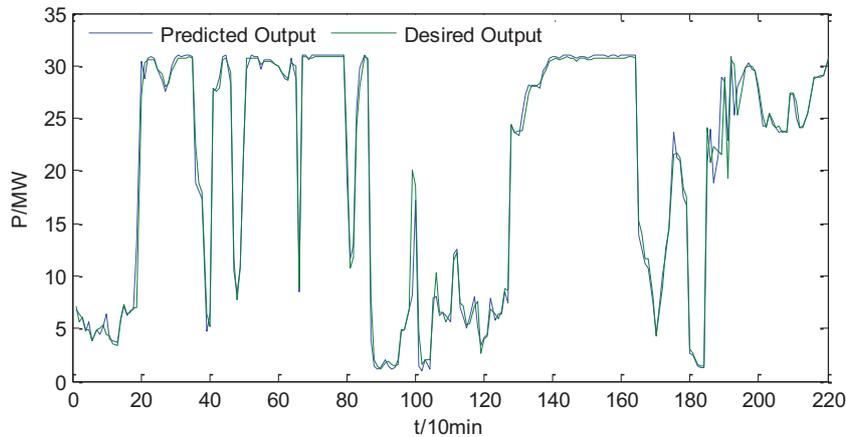
Fig. (3). Comparison of the real value and the 1-h-ahead prediction value based on GA-ELM.

4.2. Validation Experiment Result II

In this experiment, all the data and the parameters are the same as those in the experiment I. Apart from that, 1h-head multi-step prediction model and1/2 h-head multi-step prediction model are constructed according to the Equation (13). The results are shown in Tables **2** and **3**, Figs. (**3** and **4**).
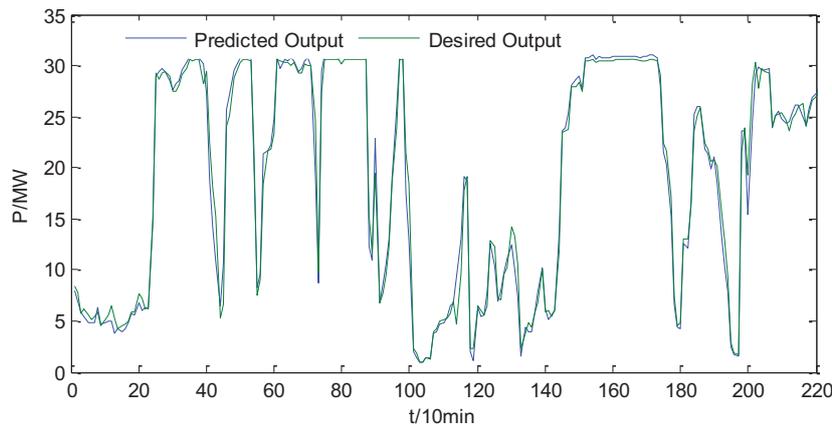
Fig. (4). Comparison of the real value and the 1/2-h-ahead prediction value based on GA-ELM.

From Tables **2** and **3**, it can be seen that: (a) all the models forecast the wind speed effectively; (b) among all involved models, the hybrid GA-ELM model has the best performance.

Table 2. 1h-ahead prediction comparison among the Elman, ELM, and GA-ELM models.

Prediction method	MAPE	RMSE	MAXERROR
Elman	9.2430	2.2872	12.1871
ELM	15.7708	2.9498	18.6572
GA-ELM	6.9519	1.1603	6.9477

From Figs. (**3** and **4**), it can be analyzed that: (a) When comparing the hybrid GA-ELM model with the Elman and the single ELM mode, the hybrid GA-ELM model has improved the performance of the latter obviously in both predictions; (b)In both predictions, the GA-ELM model shows better generalization ability of the others significantly.

Table 3. 1/2h-ahead prediction comparison among the Elman, ELM, and GA-ELM models.

Prediction method	MAPE	RMSE	MAXERROR
Elman	9.1176	1.4901	9.802
ELM	14.9209	3.0004	17.4408
GA-ELM	6.2001	1.3808	11.9234

4.3. Validation experiment Result III

In this case study, the GA-ELM method was applied to the historical wind power data of the Xinjiang region's Wind farm. The evaluation indexes were the same as in the above experiments. From September 23 to 30, the data was randomly sampled every 15 min in that wind farm. The fore 80 percent data were used for training, and the remnant was used for testing. Thus, 2h-head multi-step prediction model was constructed by means of the Equation (13).

In the present experiment, the node number of the ELM hidden layer was set as 120, the maximum iteration was selected as 60 and the maximum group number was set as 100. The proposed GA-ELM model was compared with the ELM to further evaluate its performance. The results are shown in Table **4** and Fig. (**5**).

Table 4. Comparison among the Elman, ELM, and GA-ELM models.

Prediction method	MAPE	RMSE	MAXERROR
Elman	14.9997	4.4708	19.4312
ELM	13.8229	1.7845	7.2912
GA-ELM	9.1183	0.9875	4.3561
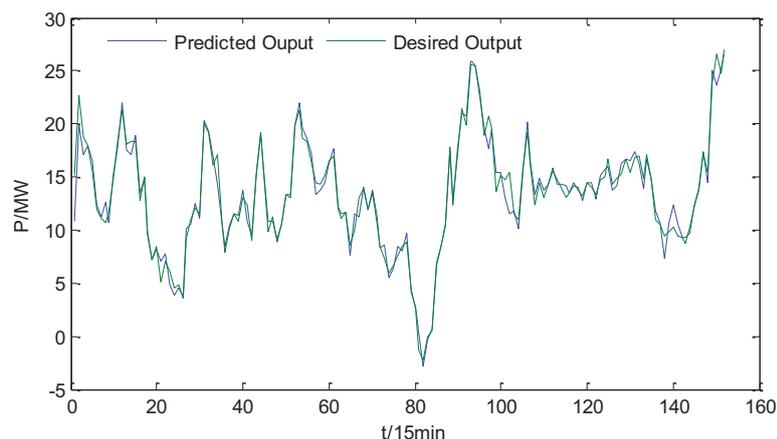
Fig. (5). Comparison of the real value and the 2-h-ahead prediction value based on GA-ELM.

As shown in Fig. (**5**), compared with Elman and common ELM approach, experimental results show smoothness and effectiveness of the proposed method. With reference to Table **4**, the prediction accuracy of the GA-ELM is superior to ELM. Moreover, from Tables **1** to **4**, in case study,all of the evaluation indexes verify the better results of the GA-ELM method proposed in this paper.

CONCLUSION

In this paper, a new hybrid method was proposed for the wind power high-precision predictions by combining the GA algorithm and the ELM algorithm.

The history data was preprocessed with GA method for optimizing, and on this basis, a wind power prediction model was constructed using ELM algorithm. Moreover, GA was used to optimize the type, bias, input weight and regularization coefficient of the activation function for ELM, thus obtaining the hybrid prediction model for wind power based on GA-ELM.

The experiments have been carried out for a dataset obtained from the NREL and an historical wind power data of Xinjiang wind farm. Results have shown that the GA-ELM model is effective for short-term wind power prediction,

significantly outperforms the Elman model, and is better than the ELM in terms of prediction accuracy.

The proposed hybrid forecasting method has low complexity, operates in real-time and is easy to implement. Therefore, it is suitable for the short term wind power high precision predictions for the safety of the wind power conversion.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCE

[1]   X. Wang, P. Guo, and X.B. Huang, "A review of wind power forecasting models", *Energy Procedia,* vol. 12, pp. 770-778, 2011.
[http://dx.doi.org/10.1016/j.egypro.2011.10.103]

[2]   X.K. Gu, G.F. Fan, and H.Z. Dai, "Summarization of wind power prediction technology", *Power System Technology.,* vol. 31, no. 2, pp. 335-338, 2007.

[3]   X. Zhao, S.X. Wang, and T. Li, "Review of evaluation criteria and main methods of wind power forcasting", *Energy Procedia,* vol. 12, pp. 761-769, 2011.
[http://dx.doi.org/10.1016/j.egypro.2011.10.102]

[4]   B. Ernst, B. Oakleaf, and M.L. Ahlstrom, "Predicting the wind", *IEEE Power and Energy Magazine,* vol. 5, no. 6, pp. 79-89, 2007.
[http://dx.doi.org/10.1109/MPE.2007.906306]

[5]   A.K. Choudhary, K.G. Upadhyay, and M.M. Tripathi, "Estimation of wind power using different soft computing methods", *International Journal of Electrical System,* vol. 1, no. 1, pp. 1-7, 2011.

[6]   K. Mohammadi, S. Shamshirband, and P.L. Yee, "Predicting the wind power density based upon extreme learning machine", *Energy,* vol. 86, pp. 232-239, 2015.
[http://dx.doi.org/10.1016/j.energy.2015.03.111]

[7]   M. Lange, and U. Focken, "New developments in wind energy forecasting", *IEEE Power and Energy Society General Meeting 2008 - Conversion and Delivery of Electrical Energy in the 21st Century,* pp. 1-8, 2008.
[http://dx.doi.org/10.1109/PES.2008.4596135]

[8]   G. Giebel, G. Kariniotakis, and R. Brownsword, "The state-of-the-art in short-term prediction of wind power – a literature review", Available: http: //www.anemos-project.eu/download//ANEMOS_D1.1_StateOfTheArt_v1.1.pdf

[9]   A.R. Garcia, and E. De-La-Torre-Vega, "A Statistical wind power forecasting system – A Mexican wind-farm case study", In: *European Wind Energy Conference and Exhibition – EWEC Parc Chanot*, Marseille, France, 2009.

[10]   A.M. Foley, P.G. Leahy, and A. Marvuglia, "Current methods and advances in forecasting of wind power generation", *Renew. Energy,* vol. 37, no. 1, pp. 1-8, 2012.
[http://dx.doi.org/10.1016/j.renene.2011.05.033]

[11]   Y. Xiuyuan, X. Yang, and C. Shuyong, "Wind speed and generated power forecasting in wind farm", *Proceedings of the CSEE,* vol. 25, no. 11, pp. 1-5, 2005.

[12]   L.J. Wang, and L. Dong, "Combined prediction of wind power generation in multi-dimension embedding phase space", *Control and Decision.,* vol. 25, no. 4, pp. 576-581, 2010.

[13]   J.W. Zeng, and W. Qiao, "Short-term wind power prediction using a wavelet support vector machine", *IEEE Transaction on Sustainable Energy.,* vol. 2, no. 3, pp. 255-264, 2012.
[http://dx.doi.org/10.1109/TSTE.2011.2180029]

[14]   G.B. Huang, Q.Y. Zhu, and C.K. Siew, "Extreme learning machine: Theory and applications", *Neuro Computing,* vol. 1/2/3, pp. 489-501, 2006.
[http://dx.doi.org/10.1016/j.neucom.2005.12.126]

[15]   N.Y. Liang, G.B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks", *IEEE Transactions on Neural Networks,* vol. 17, no. 6, pp. 1411-1423, 2006.
[http://dx.doi.org/10.1109/TNN.2006.880583] [PMID: 17131657]

[16]   J.H. Holland, *Adapation in Natural and Arifical Systems.,* University of Michigan Press: Ann Arbor, MI, USA, 1975.

[17]   P.J. Angeline, G.M. Saunders, and J.B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks", *IEEE Transactions on Neural Networks,* vol. 5, no. 1, pp. 54-65, 1994.
[http://dx.doi.org/10.1109/72.265960] [PMID: 18267779]

[18]   V. Maniezzo, "Genetic evolution of the topology and weight distribution of neural networks", *IEEE Transactions on Neural Networks,* vol. 5, no. 1, pp. 39-53, 1994.
[http://dx.doi.org/10.1109/72.265959] [PMID: 18267778]

[19]   R. Gunter, "Continuous analysis of canonical genetic algorithms", *IEEE Transactions on Neural Networks,* vol. 5, no. 1, pp. 39-53, 1994.
[PMID: 18267778]

[20]   M. Xu, Y. Qiao, and Z.X. Lu, "A comprehensive error evaluation method for short-term wind power prediction", *Dianli Xitong Zidonghua,* vol. 35, no. 12, pp. 20-26, 2011.

[21]   C.W. Potter, D. Lew, J. McCaa, S. Cheng, and S. Eichelberger, "Creating the dataset for the western wind and solar integration study(U.S.A)", In: *7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms,* Madrid: Spain, 2008, pp. 325-338.
[http://dx.doi.org/10.1260/0309-524X.32.4.325]