Neurocomputing 203 (2016) 111–120

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A new intelligent method based on combination of VMD and ELM for short term wind power forecasting

Ali Akbar Abdoos

Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Mazandaran, Iran

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form
14 March 2016
Accepted 21 March 2016
Communicated by Dr. G.-B. Huang
Available online 7 May 2016

Keywords:
Extreme Learning Machine (ELM)
Pattern recognition
Variational Mode Decomposition (VMD)
Wind power

ABSTRACT

Wind power plants are clean energy resources and are increasingly utilizing in power systems. The wind power forecasting is complicated due to the volatile nature of wind speed. Since, accurate wind power forecasting is crucial for wind power plants, a new hybrid pattern recognition method is proposed in this study for short term wind power forecasting. The time series of produced power is estimated through combination of three main steps including: pre-processing, feature selection and regression steps. In the first step, the time series of wind power is decomposed into different modes by using Variational Mode Decomposition (VMD) technique. These modes are then used to construct training patterns and forecasted outputs. In the second step, to eliminate redundant properties, the feature selection method based on Gram-Schmidt Orthogonalization (GSO) is applied on potential candidates. In the last step, Extreme Learning Machines (ELMs) as efficient and fast regression tools are trained using subsets of selected features. Eventually, the power generated by the wind farm is estimated by summing all the predicted modes values. The performance of the proposed wind power forecaster is evaluated using real data collected from two wind farms located in Sotavento Galicia in Spain and Texas in US. The obtained results justify the superiority of the proposed method in accurate forecasting and saving computational time comparing to some previously reported methods.

1. Introduction

As a result of high pollution level of fossil-fired power generation units and its negative impacts on environment, renewable energy resources such as wind, solar and wave energy are now seriously considered as viable alternatives of electricity energy. Among the aforementioned resources, the wind power has the highest growth rate in the power systems. The produced electric power of wind farms can affect the power system operation such as planning, scheduling and dispatching. Forecasting the produced wind power is a big challenge because of random nature of wind speed. Thus, an accurate wind power forecaster is of great importance for system operators [1,2].

Many efforts have been devoted to short term wind power forecasting in a time span of minutes to few days. The proposed methods can be classified into two main groups: physical and statistical [1,2]. In physical methods, wind speed and climate variables are forecasted using Numerical Weather Prediction (NWP) models, and the results are plugged into the manufacturer power curves, in order to estimate the future wind power generation. In [3] the Wind Atlas Analysis and Application Program (WAsP) as well as PARK are used for corrections of wind speed predictions but in [4] the NWP model is directly used to forecast the wind speed. It has been shown in [5] that eta model can effectively predict the wind energy up to 36-hour ahead. However, these methods need huge and complicated computational processes. Moreover, the forecasting errors of NWP models are common and can lead to remarkable errors in wind power forecasters.

From statistical point of view, a model is constructed to reveal the relationship between wind power/speed and other variables using historical data [6–11]. Autoregressive (AR) model [6], Autoregressive Moving Average (ARMA) model [7], Autoregressive Integrated Moving Average (ARIMA) model [8], Persistence (PER) [9], New-Reference (NR) [10] and Gaussian regression model [11] are some of the statistical methods which have been used for short term wind power/speed forecasting.

Nowadays, statistical methods based on artificial intelligence have drawn more attention due to their ability to recognize nonlinear and complicated relations between training patterns. Most of these pattern recognition based forecasters are independent of NWP data. Artificial Neural Network (ANN) with different structures such as Feed Forward Neural Network (FFNN) [12–19], Radial Basis Function Neural Network (RBFNN) [20], Elman neural network [21] and Extreme Learning Machines (ELM) [22] have been used to learn exemplar patterns of wind speed/power. Besides, Support Vector

E-mail address: abdoos_a@yahoo.com

http://dx.doi.org/10.1016/j.neucom.2016.03.054
0925-2312/

Machines (SVMs) [23] and Adaptive Neuro Fuzzy Interface Systems (ANFIS) [24,25] have also been used as regression cores for some forecasting engines. Recently, some methods have been presented based on the combination of NWP and learning machines [26,27]. In order to obtain the optimum structure of ANN, many different adjustable parameters should be examined which is a time consuming and boring task. Unlike ANN, the SVM has less tuning parameters but its training time increases exponentially by increasing the dimension of input vector. In some researches, the heuristic search algorithms such as Particle Swarm Optimization (PSO) and Fruit Fly Optimization (FFO) have been have been used to determine the adjustable parameters of regression tools [28–30].

To decrease the forecasting errors, the preprocessing step has been utilized in some forecasting algorithms to analyze the time series of wind power/speed signals. Wavelet Transform (WT) as a powerful signal analysis tool has been used for decomposition of signals into approximation and details levels [21, 23, 25 and 31]. WT analysis suffers from two main drawbacks: firstly, different mother wavelets should be examined to obtain the best performance of forecaster engine and secondly, the required number of decomposition levels should be determined by trial and error. Differential Empirical Mode Decomposition (EMD) has been used in [18,19] to analyze the wind power. Using the EMD method, any complicated data set can be decomposed into several components as Intrinsic Mode Functions (IMFs). Since the decomposition is based on the local characteristic time scale of data, it can be applied to non-stationary time series of produced wind power. The main disadvantage of EMD is that the number of IMFs can be changed according to harmonic content of signals.

In this paper, a new hybrid intelligent method is presented for forecasting of wind farms output power. In the first step, Variational Mode Decomposition (VMD) [32] is employed to decompose the wind power signal into several modes which have specific sparsity properties while producing main signal. The time series of each decomposed mode is utilized to create the training patterns. In order to eliminate redundant data from exemplar patterns, the well-known Gram-Schmidt Orthogonalization (GSO) based feature selection method [33,34] is applied on the potential candidates of input matrices. Then, the selected features are used for training of the efficient regression tool namely Extreme Learning Machines (ELMs) [35–37]. Since ELM requires only a single-pass training stage without any iteration for weights adjustment, it has very fast learning process. The number of ELM hidden neurons can be easily determined based on dimensions of input and output vectors [38]. The proposed forecasting engine is evaluated using historical data of two different wind farms: Sotavento Galicia in Spain [39] and Texas in US [40]. The proposed intelligent method has advantages from the following aspects:

- Small variations in time series of produced wind power are detected more precisely by using VMD analysis in preprocessing step.
- The generalization capability of learning machines increases by elimination of redundant features using GSO based feature selection method.
- The GSO ranks features according to their orthogonality, therefore no adjustable parameter is needed in the ranking process.
- The execution time of the proposed method is very low because ELM as a powerful regression tool has extremely fast learning speed without local minima issues.
- The applicability of the proposed method is tested in real power systems using historical data collected from two different wind farms.
- The proposed method is independent of climate conditions due to the fact that future generated power is estimated based on historical data of wind farm output power.
- Compared to other forecasting engines, the proposed method has few adjustable parameters to execute forecasting process.

2. Preliminaries

2.1. Variational Mode Decomposition

The VMD is a signal processing technique that decomposes a real valued signal into different modes called u_k, which have specific sparsity properties while producing main signal. It is assumed that each mode k can be concentrated around a center pulsation ω_k which is determined during the decomposition process. So, the sparsity of each mode is its bandwidth in spectral domain. In order to obtain the mode bandwidth, the following steps should be fulfilled: 1) Applying Hilbert transform to each mode u_k in order to obtain unilateral frequency spectrum, 2) Shifting the mode's frequency spectrum to “baseband”, by using an exponential tuned to the respective estimated center frequency and 3) Estimation of the bandwidth through the H^1 Gaussian smoothness of the demodulated signal, i.e. the squared L^2–norm of the gradient. Thus, the decomposition process is realized by solving the following optimization problem [32]:

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

where $f(t)$ is the main signal to be decomposed, $\{u_k\} := \{u_1, \ldots, u_K\}$ and $\{\omega_k\} := \{\omega_1, \ldots, \omega_K\}$ implicate the set of all modes and their center frequencies, respectively. $\delta(t)$ is the Dirac distribution and * denotes convolution. In order to address the constraint, both penalty term and Lagrangian multipliers λ are considered. 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. So, the above optimization problem is changed to unconstraint one as below [32]:

$$\mathcal{L}(\{\boldsymbol{u_k}\}, \{\boldsymbol{\omega_k}\}, \lambda) := \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-i\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{2}$$

Then the Alternate Direction Method of Multipliers (ADMM) is used for solving the original minimization problem (2) by finding the saddle point of the augmented Lagrangian L in a sequence of iterative sub-optimizations [32]. Plugging the solutions of the sub-optimizations into the ADMM, and directly optimizing in Fourier domain, the complete algorithm for variational mode decomposition can be found in [32]. According to ADMM optimization technique, u_k and ω_k should be updated to realize the VMD analysis process. To update the modes, the optimization problem of relation (2) is solved with respect to u_k. This sub optimal problem is represented as follows [32].

$$u_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-i\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\} \tag{3}$$

The solution of this quadratic optimization problem is readily found by letting the first variation vanish for the positive frequencies [32]:

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

which is clearly identified as a Wiener filtering of the current residual, with signal prior $1/(\omega-\omega_k)^2$. The full spectrum of the real mode is then simply obtained by Hermitian symmetric completion. Conversely, the real part of the inverse Fourier transform of this filtered analytic signal yield the mode in time domain.

In the second sub-problem, the optimization problem of relation (2) is solved with respect to ω_k. The center frequencies do not appear in the reconstruction fidelity term, but only in the bandwidth prior. The relevant problem thus reads [32]:

$$\omega_k^{n+1}=\underset{\omega_k}{\operatorname{argmin}}\left\{\alpha\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-i\omega_k t}\right\|_2^2\right\} \tag{5}$$

The solution of above suboptimization problem in frequency domain is represented as follows [32]:

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{6}$$

which puts the new ω_k at the center of gravity of the corresponding mode's power spectrum. This mean carrier frequency is the frequency of a least squares linear regression to the instantaneous phase observed in the mode.

2.2. Gram-Schmidt orthogonalization based feature selection method

GSO process is a simple forward selection method which can rank features based on their orthogonality with respect to target. Suppose that the kth feature of M patterns is denoted by vector $X_k=[x_{k1},x_{k2},\ldots,x_{kM}]^T$ and $Y=[y_1,y_2,\ldots,y_M]^T$ represents the vector of output target. In order to select the best correlated feature with output, the cosine of angle between each input feature X_k and target Y is calculated as an evaluation criterion [33,34]:

$$\cos(\varphi_k)=\frac{\langle X_k.Y\rangle}{\|X_k\|\|Y\|} \tag{7}$$

where φ_k is the angle between input kth feature of vector X_k and output target Y, N is the number of all features and $\langle X_k.Y\rangle$ denotes the inner product between X_k and Y. If the output is fully proportional to input, the φ_k is zero and inversely if the output is fully uncorrelated to input, the φ_k is $\pi/2$ [33]. So, in an iterative procedure, the feature that maximizes the above mentioned evaluation criterion, is selected as the most correlated feature to target. As depicted in Fig. 1, for selection of the next feature, the output vector and all other candidate features are mapped to null space of the selected feature and then input features and output vectors are updated with new data. The ranking procedure is repeated until all candidate features are sorted, or when a predetermined stopping condition is met [33,34].

2.3. Extreme Learning Machine

The ELM was firstly introduced by Huang et al. [35] for both classification and regression purposes. It has founded based on the Single Layer Feed-forward Network (SLFN) as shown in Fig. 2 [35]. It includes input, hidden and output layers. For N arbitrary distinct samples $(\mathbf{p}_j,\mathbf{o}_j)\in R^n\times R^m$, SLFNs with L hidden nodes and activation function $G(\boldsymbol{a}_i,b_i,\boldsymbol{p}_j)$ are mathematically modeled as [35,36]:

$$\sum_{i=1}^{N}\beta_i G\left(\boldsymbol{a}_i,b_i,\boldsymbol{p}_j\right)=o_j \quad j=1,2,\ldots,N \tag{8}$$

where $\boldsymbol{a}_i$ denotes the input weight vector connecting the ith hidden node and the input nodes, b_i is the threshold or impact factor

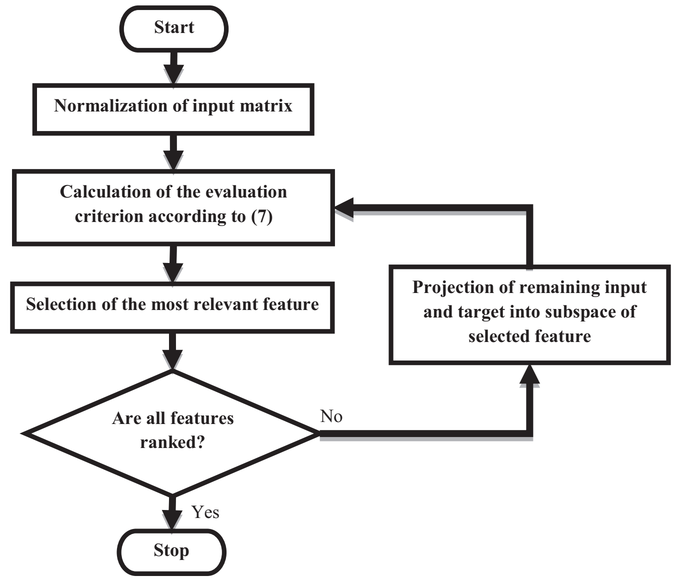


Fig. 1. Implementation steps of GSO based feature selection method.

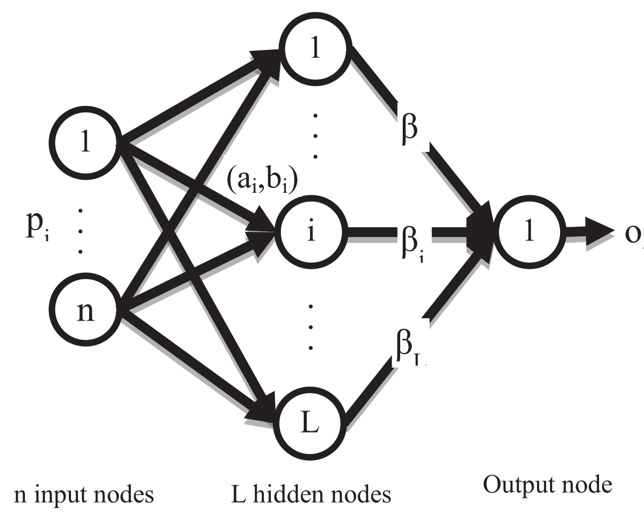


Fig. 2. ELM structure.

of the ith hidden node and β_i is the weight vector connecting the ith hidden node and the output node. The formula (8) can be written in matrix form as [35–37]:

$$H\beta=\mathbf{O}$$

$$\text{where } \boldsymbol{H}=\begin{bmatrix}\boldsymbol{h}(\boldsymbol{p}_1)\\ \vdots\\ \boldsymbol{h}(\boldsymbol{p}_N)\end{bmatrix}=\begin{bmatrix}G(\boldsymbol{a}_1,b_1,\boldsymbol{p}_1) & \ldots & G(\boldsymbol{a}_L,b_L,\boldsymbol{p}_1)\\ \vdots & \ldots & \vdots\\ G(\boldsymbol{a}_1,b_1,\boldsymbol{p}_N) & \ldots & G(\boldsymbol{a}_L,b_L,\boldsymbol{p}_N)\end{bmatrix}_{N\times L},$$

$$\beta=\begin{bmatrix}\beta_1^T\\ \vdots\\ \beta_L^T\end{bmatrix}_{L\times m} \text{ and } \boldsymbol{O}=\begin{bmatrix}o_1^T\\ \vdots\\ o_L^T\end{bmatrix}_{N\times m} \tag{9}$$

$\boldsymbol{H}$ is called the hidden layer output matrix that its ith column is the output of the ith hidden node with respect to inputs p_1, p_2, …, p_N.

It has been proven that for any non-constant piecewise continuous function G, the continuous target function $f(\boldsymbol{p})$ can be approximated by the SLFN without adjustment of hidden node parameters [35]. So, all hidden node parameters can be randomly generated without the knowledge of the training data. That is for any continuous target function G and any randomly generated sequence $\left\{(\boldsymbol{a_i}, b_i)_{i=1}^{L}\right\}$, $\lim_{L\to\infty}\|f(\mathbf{p})-f_L(\mathbf{p})\| = \lim_{L\to\infty}\|f(\mathbf{p}) - \sum_{i=1}^{L}\beta_i G(\boldsymbol{a_i}, b_i, \mathbf{p})\| = 0$ holds with probability one if β_i is chosen to minimize $\|f(\mathbf{p})-f_L(\mathbf{p})\|, i=1,\ldots,L$ [36]. So, three stages should be proceeded for realization of ELM:

Step 1: Assigning hidden node parameters randomly,

Step 2: Calculation of the hidden layer output matrix $\boldsymbol{H}$,

Step 3: Calculation of the output weight $\beta = \boldsymbol{H}^{\dagger}\boldsymbol{T}$.where $H^{\dagger}$ is the Moore–Penrose generalized inverse of hidden layer output matrix H [37].

3. Methodology

The proposed forecasting algorithm is based on the recognition of wind power patterns. On the other hand, the future generated wind power is estimated using previous historical data. The structure of the proposed short term wind power forecaster is depicted in Fig. 3. The forecasting procedure is realized through three main steps: decomposition of the time series of produced wind power, feature selection and regression.

In the first step, the historical data of produced wind power is decomposed into different modes using VMD analysis. The time series of produced power of Sotavento Galicia wind farm on January 2014 together with three decomposed modes are shown in Fig. 4. The stochastic and complex nature of produced wind power can be clearly seen in the main signal. Therefore, the analysis of time series of wind power can be helpful for precise modeling of its intermittent characteristics. Mode 1 containing low frequency components is a good replica of smooth changes of produced wind power and higher order modes i.e. modes 2 and 3 include higher frequency components. Thus, the small changes of the produced wind power can be more precisely recognized using high order decomposed modes. In order to remove the inappropriate effect of different variation range of decomposed modes on the training process of learning machines, decomposed modes are normalized between 0 and 1. In this work, historical data of the previous week is considered for creation of exemplar patterns. For prediction of the next sample of produced wind power, the historical data of previous day is used as potential features. In this way, each column of the input matrix pertains to time series of previous 24-hour data and the elements of output matrix imply the next produced wind power. Since the regression accuracy of learning machine is affected by the dimension of input matrix, it is necessary to eliminate the redundant data by using an appropriate feature selection method. There are many feature selection methods which can be generally classified into three main groups [33]: wrapper, filter and embedded methods. The wrapper method scores feature subset based on the performance of the forecaster model that needs a huge computational burden. Filters use arithmetic analysis of features without evaluation of forecaster model performance. The embedded method attempts to take advantages of the filter and wrapper approaches which permits the latter to utilize the information delivered by the filter algorithm to accelerate the convergence of the wrapper algorithm. In the second step of the proposed method, the GSO based feature selection method is used for selection of the most informative features subset with minimum dimension. Since the GSO based feature selection method provides a feature ranking rather than an explicit best feature subset such as auto correlation and mutual information methods [31], desirable numbers of selected features can be set for evaluation of the proposed method without

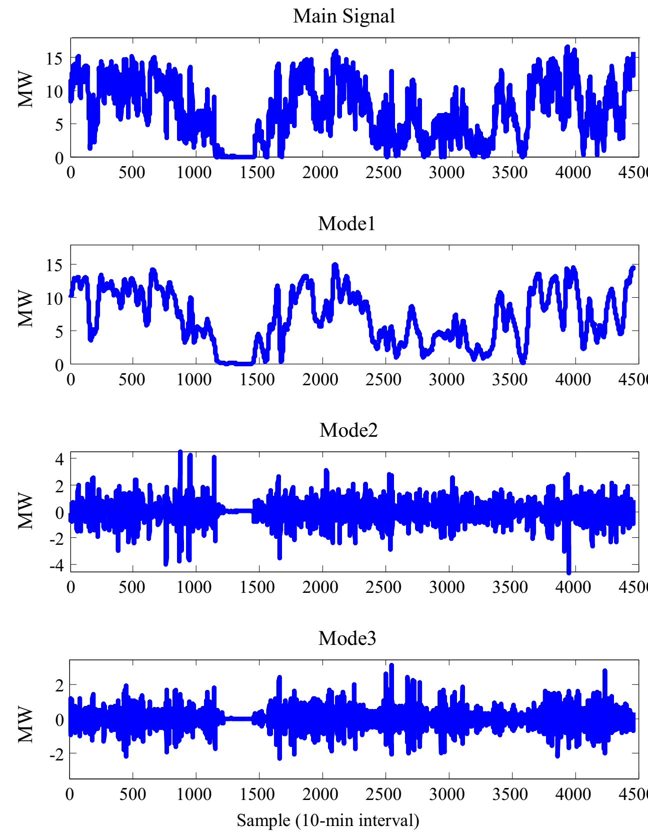


Fig. 4. Produced wind power signal of January 2014 and decomposed modes.

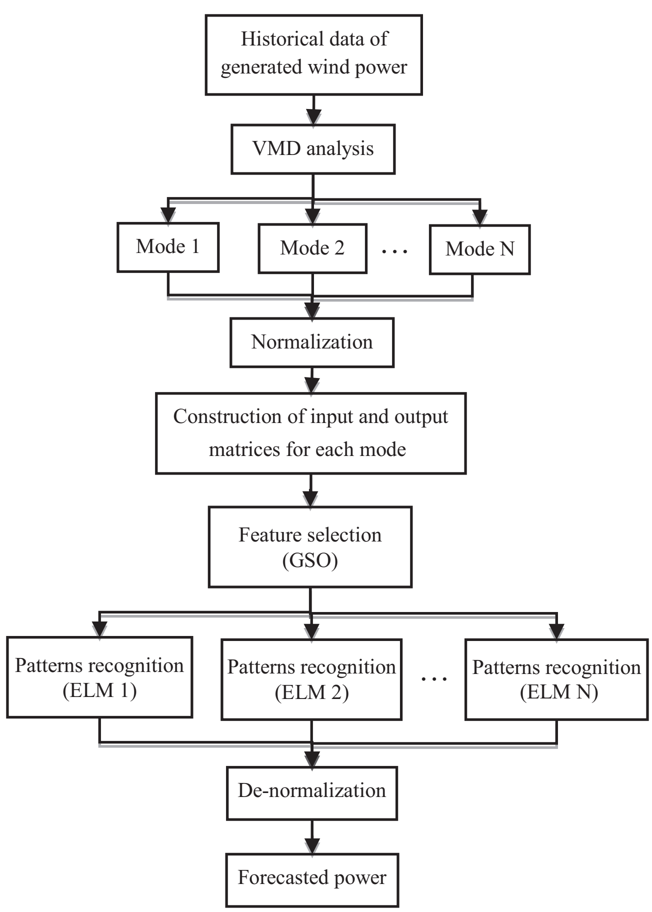


Fig. 3. Flowchart of the proposed method.

adjustment of any parameter. Removal of redundant features from training patterns brings following advantages:

- The required memory space decreases,
- Both computational burden and required time of training process decreases,
- The generalization capability of learning machine increases.

In the third step, a learning machine is needed for training of exemplar patterns of each decomposed mode. The ELM is utilized as the regression core of the proposed forecasting engine to extract the complicated input–output mapping between historical and forecasting data. ELM has a simple structure with only two adjustable parameters i.e. number of hidden neurons and activation function. Since the training phase is implemented through single stage without any iterative process, its training time is extremely short. Due to the random nature of ELM and to avoid undesirable effect of bad trained regression tools on forecasting accuracy, the training process is repeated 10 times. After the elimination of 4 outlier forecasted values, the mean of 6 remaining values are considered as the final forecasted value of ELM. The outlier values are defined as those forecasted values which have the largest difference from the mean. For each mode, the exemplar patterns are applied to the ELM in training phase. Thus, the number of required ELM is equal to the number of decomposed modes. The output of each ELM denotes the forecasted wind power of the same mode order. The final forecasting result of the produced power is obtained by summing all de-normalized values of the ELMs outputs.

4. Numerical results

Before proceeding of the forecasting process, two main parameters should be determined: forecast interval and forecast horizon. The time interval between available samples denotes the forecast interval and the forecast horizon is the length of time into the future for which forecasts are to be prepared. Both parameters are heavily dependent on the application of forecast results. For example, hourly wind power forecasts can be used for unit commitment while the interval forecast for economic dispatch can be as short as 10 min. In this study, we select 10 min and 4 h for forecast interval and forecast horizon, respectively. Thus, the time interval between two consecutive samples of wind power is 10 min and the forecast engine predicts the next 24 samples $(24\times 10\ \text{min}=240\ \text{min}=4\ \text{h})$ of wind farm produced power. In each step, the new forecasted wind power is used for prediction of next sample and this repetitive process is continued till the 24th sample is predicted. To obtain the forecast results of a certain day, the input of forecast engine should be updated with real data of wind power after forecasting of 24 samples (4 h).

In this study, the 10-minutly lagged values of wind farm produced power in the past 24 h are considered as candidate input. Thus, each candidate vector of input matrix has $24\times 6=144$ elements. At first sight, the dimension of input vector may look very large but the elimination of redundant features by using GSO method removes the undesirable effect of large dimension on the regression tool. Therefore, the meaningful features which have the greatest impact on the output matrix can be involved in the forecasting process. Moreover, the exemplar patterns of input matrix is constructed using historical data of the previous week. The large exemplar patterns can mislead the forecaster engine. On the other hand, few exemplar patterns increase the risk of over fitting to the forecaster model and decreases the generalization capability of regression core. Since 10-minuetly stored data of wind power is considered in this study, there are $7\times 24\times 6=1008$ exemplar patterns using historical data of the previous week.

In the forecasting process, the exemplar patterns are created for each mode. So, the number of ELMs and decomposed modes are identical. The ELM has very simple structure with only two adjustable parameters: activation function and number of hidden nodes. Since, the decomposed modes are normalized between 0 and 1, the log-sigmoid is selected as activation function. The ELM performance depends partly on the number of hidden nodes. As the ELM is a type of SLFN, as a general role of thump, the number of hidden nodes can be determined as follows [38]:

$$L=\text{round}\left(\sqrt{n+m}+\text{rand}(1\sim 10)\right) \tag{10}$$

where L is the number of hidden nodes, n and m denote dimensions of input and output vectors, respectively. In this study, the dimension of input vector depends on the number of selected features and the dimension of output vector is 1. In order to obtain a forecaster engine with optimum structure, two adjustable parameters should be determined: number of selected features and number of decomposed modes. Using exhaustive search method for investigation of all combination of possible solution is not practical because of extensive computational burden. Therefore, the cross-validation technique is used for tuning of above mentioned parameters. Although the globally optimum values of adjustable parameters are not determined by using cross-validation technique but it yields acceptable results with low execution time. The number of selected features is chosen from sequential numbers varying from 10 to 70 with step size of 10 and the number of decomposed modes is sequentially selected between 1 and 10. To evaluate the performance of the proposed method, four days of different seasons are considered: 1 May, 12 April, 18 August and 27 November, all of year 2014. So, for 70 different states, two forecasting error indices i.e. Normalized Root Mean Square Error (NRMSE) and Normalized Mean Absolute Error (NMAE) are calculated as follows:

$$\text{NRMSE}=\sqrt{\frac{1}{N_s}\sum_{s=1}^{N_s}\left(\frac{\text{WP}_{act(s)}-\text{WP}_{for(s)}}{\text{WP}_{cap}}\right)^2}\times 100 \tag{11}$$

$$\text{NMAE}=\frac{1}{N_s}\sum_{s=1}^{N_s}\left|\frac{\text{WP}_{act(s)}-\text{WP}_{for(s)}}{\text{WP}_{cap}}\right|\times 100 \tag{12}$$

where $\text{WP}_{act(s)}$ and $\text{WP}_{for(s)}$ indicate the actual and forecasted values of wind power for sth sample, respectively. Also, N_s indicates number of samples, which is 144 for a test day, and WP_{cap} is the total installed capacity of wind farm.

4.1. Sotavento Galicia wind farm

To demonstrate the effectiveness of the proposed method, the actual produced power of Sotavento Galicia wind farm in Spain is used in this research work. The wind farm consists of 24 wind turbines with nominal output power ranging from 600 to 1320 kW that provides total installed power capacity of 17,560 kW. The wind speed, wind direction and produced power of wind farm are available in 10-minute interval time [39]. Since short term wind power forecasting is considered in this study, there is no need to predict the meteorological variables using NWP models.

To have a better representation and understanding of obtained results, the inverse of error indices i.e. NRMSE and NMAE are shown in Figs. 5 and 6, respectively. It should be noted that error indices are calculated for four days of interest and then the mean value is considered for evaluation of different test cases. It is clearly seen that the case with 20 selected features together with 10 decomposed modes leads to the best forecasting accuracy of 5.51% and 3.58% for NRMSE and NMAE, respectively. An important point which can be found from the cross-validation test is that both NRMSE and NMAE criteria are overly improved when the number of decomposed modes increases. In the most cases, after nine decomposition levels, the forecasting accuracy does not improve significantly and remains almost constant. So, the decomposition of signal into more levels increases only the computational burden without considerable improvement on forecasting accuracy.

To show the effectiveness of the ELM as the regression core of the proposed method, forecasting errors have been calculated for other regression tools such as ANN, SVM and Wavelet Neural Network (WNN). Obtained results are presented in Table 1 for above mentioned regression tools. The forecast results are also depicted in Fig. 7 for four days of interest. The obtained results for January 22, 2014 (Fig. 7(a)) show that the proposed method can more precisely trace the produced wind power as compared to other methods. The actual produced power of January 22, 2014 (solid blue line) has two sharp changes which are estimated by our proposed method (dashed red line) with low forecasting error while these changes are not accurately forecasted by other regression tools. For April 12, 2014 (Fig. 7(b)), the forecasted values by using ANN become zeros for some instants in which the actual values of produced power are more than 2 MW. For this day, the WNN model has also the local minimum at sample 78 while the real produced power reaches the peak value. Besides, for August 18, 2014 (Fig. 7(c)), the SVM and ANN have two undesirable peaks while the output power of wind farm is about zero. Moreover, for November 27, 2014, (Fig. 7(d)), the forecasted power of applying other regression tools makes distance from actual value and sometimes, the forecasted values of WNN method reach to zero, incorrectly. As it can be seen from Table 1, the ELM has the best forecasting accuracy among all regression tools for four days of interest. The average of NRMSE and NMAE of each model as well as required time of forecasting process are also provided in Table 1. All computations related to forecasting process are implemented in MATLAB environment under computer with Intel core i7, 2.5 to 3.5 GHz CPU, and 12 GB of RAM. Results show that the ELM has the best forecasting accuracy with minimum training time. The average NRMSE and NMAE are 5.51% and 3.58%, respectively. Using ELM, the execution time of forecasting process is 39 s, which is the shortest time among all applied models. The SVM has the longest training time of 1435 s (about 24 min) while its forecasting errors are less than ANN and WNN. Although the training time of ANN is rather short (571 s) but it has the largest forecasting errors.

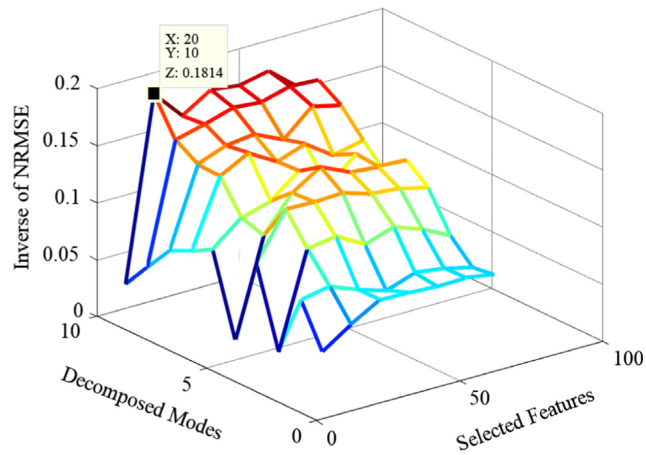


Fig. 5. Results of cross-validation test using NRMSE as error index.

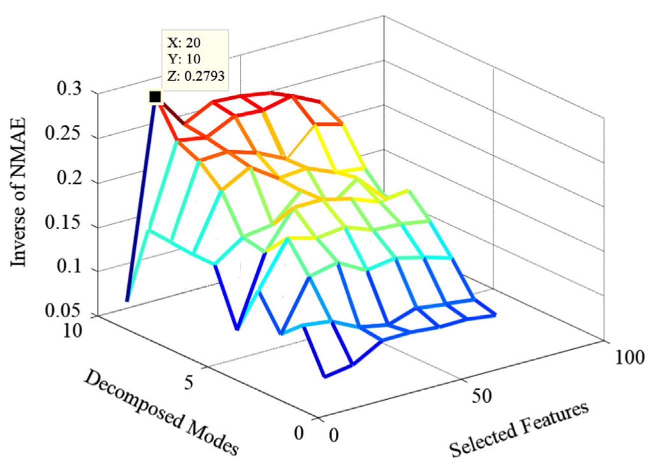


Fig. 6. Results of cross-validation test using NMAE as error index.

Moreover, the preference of applying VMD in the preprocessing step is justified by providing comparison results with two conditions: without signal analysis and using WT and EMD as signal processing tools. The forecasting errors of above mentioned conditions are given in Table 2. Without considering the signal processing stage, the average forecasting errors i.e. NRMSE and NMAE are 13.1% and 10%, respectively. Using WT in pre-processing stage and decomposition of wind power signal into 9 detail levels and one approximation, the average of NRMSE and NMAE are 8.50% and 6.73%, respectively. EMD is also utilized as another signal processing technique for decomposition of wind power signal into 10 modes. EMD yields average forecasting errors i.e. 10.55% and 7.51% in terms of NRMSE and NMAE, respectively. So, the decomposition of time series of wind power signal in preprocessing step can improve the forecasting accuracy of method and among utilized signal processing tools, the VMD has the best performance.

Besides, we have provided comparison results with some short term wind power forecasting methods in Table 3. It should be

Table 1
Comparison of the proposed method with different regression tools.

Model	Error	Test day				Average	Execution time (s)
		January 22	April 12	August 18	November 27		
ANN	NRMSE	20.1545	11.5013	11.8095	15.7773	14.81	571
	NMAE	14.0790	7.1851	6.9743	12.1262	10.09	
SVM	NRMSE	15.3555	10.5963	7.9613	18.4664	13.09	1435
	NMAE	11.0796	6.7620	5.6588	13.8007	9.325	
WNN	NRMSE	15.5720	8.5654	7.3094	17.6251	12.26	954
	NMAE	10.7389	6.3426	4.6120	10.7303	8.10	
ELM (Proposed method)	**NRMSE**	**5.7053**	**5.8785**	**3.2788**	**7.1851**	**5.51**	**39**
	NMAE	**3.9153**	**4.0780**	**1.8947**	**4.4352**	**3.58**	

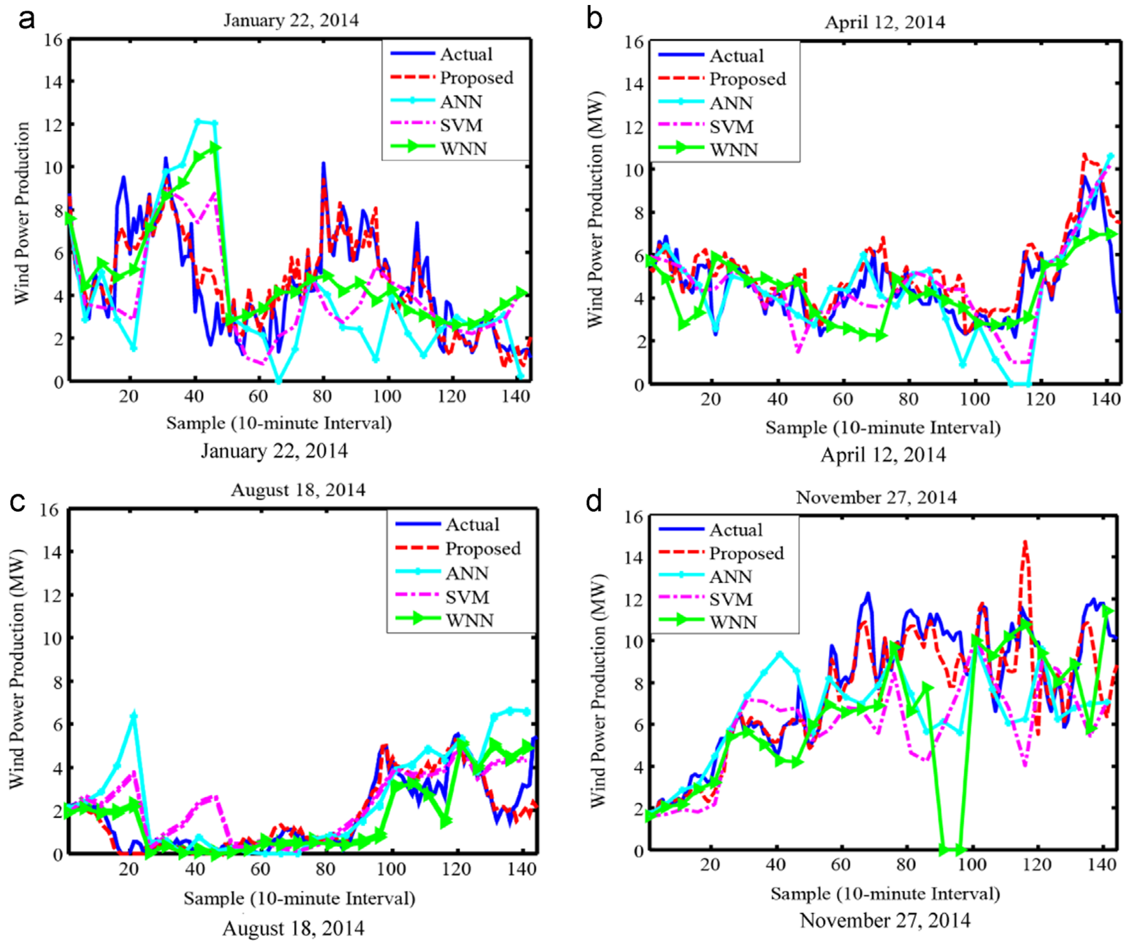


Fig. 7. Forecast results of different models for four different days. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 2
Forecasting accuracy with and without considering signal processing.

Model	Error	Test day				Average
		January 22	April 12	August 18	November 27	
Without signal processing	NRMSE	12.0975	16.8910	10.6015	12.8060	13.1
	NMAE	9.2427	13.5804	7.9080	9.2950	10
WT	NRMSE	7.7578	11.7088	5.0430	9.4966	8.50
	NMAE	5.9462	10.4819	3.1715	7.3472	6.73
EMD	NRMSE	13.5032	8.9183	8.0095	11.7762	10.55
	NMAE	9.6623	6.5184	5.2979	9.5749	7.51

Table 3
Comparison results with some short term forecasting models.

Forecasting model	NRMSE	NMAE
WT+ANN [12]	15.47	11.23
WT+AWNN [13]	8.39	7.04
WT+MI+ANFIS [25]	5.35	4.83
WT+Chaotic time series [31]	6.62	5.97
VMD+GSO+ELM (Proposed method)	**5.11**	**3.58**

noted that forecasting methods are comparable if they have exactly the same forecasting interval and horizon. Since papers with the same forecasting interval and horizon are rarely found, we have applied the forecaster models while produced power of Sotavento Galicia wind farm is used in the prediction process. For all models, the forecasting interval and forecasting horizon are set to 10 min and 4 h, respectively. Obtained results in Table 3 show that our proposed method can yield the best forecasting accuracy among all applied models. The evaluation criteria of our proposed method i.e. NRMSE and NMAE are 5.51% and 3.58%, respectively. Other presented models in Table 3, apply WT in the preprocessing step as signal analysis tool. The combination of WT and ANN in Ref. [12] has the worst forecasting results with NRMSE and NMAE of 15.47% and 11.23%. In Ref. [13], Adaptive Wavelet Neural Network (AWNN) is used as the regression core that yields NRMSE and NMAE of 8.39% and 7.04%, respectively. In Ref. [25], Mutual Information (MI) is added as an extra step for elimination of meaningless data while Adaptive Network-based Fuzzy Inference System (ANFIS) is utilized for training of historical patterns. By addition of feature selection step, the forecasting accuracy is considerably improved owing to removal of the redundant data. The combined WT and chaotic time series in Ref. [31] has NRMSE and NMAE of 6.62% and 5.97%, which are better than ANN and AWNN.

In order to demonstrate the capability of the proposed method for different forecast intervals, the hourly wind power data are also used in the training and testing processes. The historical data of previous month is used for creation of exemplar patterns while the data of previous week is considered as potential candidates of input vectors of exemplar patterns for prediction of next hour wind power. So, there are $30 \times 7 \times 24 = 5040$ patterns so that each pattern includes $24 \times 7 = 168$ hourly lagged values of wind power as potential candidates. The forecasting horizon is determined 4 h. Therefore, to predict the produced wind power of next day, the forecasting process should be repeated six times. It means, after prediction of wind power of the next 4 h, the forecaster engine will be updated by adding new real data of the recent 4-hour window. The same 20 features together with 10 decomposed modes are considered for evaluation of the forecaster engine performance. The real and forecasted wind power of July 25, 2015, together with forecasted error are given in Table 4. The NRMSE and NMAE errors of the forecasted wind power of the day of interest are 3.96% and 3.06%, respectively. Besides, the preference of the proposed method is shown by providing comparison results with hybrid WT+ANN method for which the NRMSE and NMAE are 12.96% and 10.11%, respectively. The difference between forecasted and real values is also given in the last column of Table 4. The mean absolute errors of WT+ANN and VMD+ELM are 536.78 kW and 1775.7 kW, respectively. Results show that the proposed method can also be effectively utilized for hourly wind power forecasting.

Table 4
Hourly wind power forecasting of July 25, 2015 using VMD+ELM (proposed method) and WT+ANN.

hour	Actual power (kW)	Forecasted power (kW)		Forecasted error (kW)	
		WT+NN	VMD+ELM	WT+ANN	VMD+ELM
1	12897.78	11568.93	12356.41	1328.9	541.4
2	10965.84	10591.6	10132.18	374.2	833.7
3	7650.24	10655.33	8326.24	−3005.1	−676.0
4	6946.14	92443.7	6256.79	−2298.2	689.4
5	7183.14	86673.8	7076.46	−1484.2	106.7
6	6415.44	91696	6476.59	−2754.2	−061.2
7	7187.94	80342.2	6605.28	−846.3	582.7
8	7090.92	80854.1	6158.12	−994.5	932.8
9	7125.84	78706.7	7206.31	−744.8	−080.5
10	7328.16	89059.4	8269.65	−1577.8	−941.5
11	12214.14	77139.6	11660.74	4500.2	553.4
12	8831.52	64081	9192.76	2423.4	−361.2
13	6874.26	76891.9	6838.80	−814.9	35.5
14	6809.28	66805.1	6811.06	128.8	−1.8
15	10665.24	56534.5	9019.48	5011.8	1645.8
16	10104.96	58539.3	8870.10	4251.0	1234.9
17	8321.16	72479	8262.22	1073.2	58.9
18	11511.6	73231.7	11086.29	4188.4	425.3
19	8232.18	79962.6	8506.91	235.9	−274.7
20	8079.48	91015.4	8293.04	−1022.1	−213.6
21	10086.54	89833.1	10004.38	1103.2	82.2
22	9176.04	95100.5	8441.00	−334.0	735.0
23	9295.14	84861.10	8828.49	809.0	466.6
24	8326.14	96399	6977.81	−1313.8	1348.3

4.2. National Renewable Energy Laboratory

To show the applicability of the proposed forecast engine for different wind farms, the US National Renewable Energy Laboratory (NREL) historical data are also used for further evaluation [40]. More than 32,000 wind turbines have been installed in different sites in the US. Wind speed and output power of each wind turbine are available for years 2004, 2005 and 2006 in 10-minute interval. In this study, we considered the sum of output power of ten wind turbines with ID 1–10 as the generated power of a local wind farm. These turbines have been installed in the similar locations at latitude of 31.2 and longitude of −102.2 in Texas State. The capacities of these turbines vary between 25 and 35 kW and their installation elevations are between 850 and 950 m.

In this study, the same days of interest are considered for evaluation of the proposed method using historical data of year 2006. Table 5 gives the comparative results with some other pattern recognition based methods. Both forecasting errors i.e. NRMSE and NMAE have been calculated for days of interest and the last column gives the average of calculated errors. The main conclusion which can be drawn from tabulated results in Table 5 is that the type of classifier, feature selection method and preprocessing tool have direct effect on the forecasting accuracy. For example the VMD analysis in the preprocessing step yields the best detection accuracy as compared to WT and EMD. On the other hand, the combined method ELM+CRO uses the feature selection method based on Coral Reefs Optimization (CRO) technique but its forecasting errors (NRMSE=8.37% and NMAE=5.3%) are more than our proposed method (NRMSE=7.25% and NMAE=4.35%) which uses the GSO based feature selection method. As results show, our proposed method yields the best forecasting accuracy among combined intelligent methods i.e. WT+ANN, EMD+ANN and ELM+CRO.

For further analysis and in order to show the applicability of the proposed method for different forecasting interval, the hourly historical data is also used for day-ahead wind power prediction. Table 6 gives the comparative results of different forecasting models. Obtained results show that our proposed method has the best forecasting accuracy (NRMSE=4.49% and NMAE=3.16%) as compared to other intelligent methods. The low forecasting errors justify that our proposed method can be used effectively for prediction of wind power with different forecasting interval. Besides,

Table 5
Comparison of the proposed method with different forecasting models using 10-minutly historical data of NREL.

Forecasting model	Error	Test day				Average
		January 22	April 12	August 18	November 27	
EMD+ANN [18]	NRMSE	14.5	10.77	4.33	16.43	11.51
	NMAE	8.77	6.9	2.28	9.11	6.76
WT+ANN [12]	NRMSE	7.33	9.31	5.56	17.27	9.87
	NMAE	5.15	6.8	3.44	10.28	6.42
ELM+CRO [22]	NRMSE	8.68	8.45	5.41	10.94	8.37
	NMAE	6.21	6.37	3.15	5.47	5.3
VMD+GSO+ELM (Proposed method)	**NRMSE**	**9.38**	**6.95**	**4.43**	**8.22**	**7.25**
	NMAE	**5.23**	**5.06**	**2.45**	**4.65**	**4.35**

Table 6
Comparison of the proposed method with different forecasting models using hourly historical data of NREL.

Forecasting model	Error	Test day				Average
		January 22	April 12	August 18	November 27	
EMD+ANN [18]	NRMSE	9.5	11.83	3.2	19.02	10.88
	NMAE	7.27	8.89	2.02	11.12	7.33
WT+ANN [12]	NRMSE	5.15	9.02	3.96	17.33	8.87
	NMAE	4.13	7.11	2.62	10.33	6.05
ELM+CRO [22]	NRMSE	6.9	12.11	3.29	5.93	7.06
	NMAE	4.83	8.21	2.2	3.86	4.77
VMD+GSO+ELM (Proposed method)	**NRMSE**	**5.02**	**3.84**	**3.88**	**5.23**	**4.49**
	NMAE	**3.71**	**2.70**	**2.41**	**3.82**	**3.16**

the proposed method can be applied for different wind farm with different conditions such as: wind farm capacity, location and weather conditions.

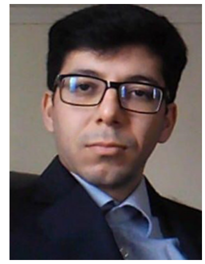
5. Conclusion

In this paper, a new wind power forecaster is proposed based on an intelligent hybrid pattern recognition methodology. VMD as a powerful signal processing technique is used for decomposition of time series of produced wind power. To improve the generalization capability of the forecaster engine and reduce the required memory, non-informative data are removed by using GSO based feature selection method. In order to find the relation between exemplar patterns and desired output, ELM is used as a powerful regression core. Then, the cross-validation technique is applied to obtain the optimum structure of the forecaster engine through performance evaluation of the proposed method for different numbers of selected features and decomposition modes. Simulation results show that the forecaster engine with 10 decomposition modes and 20 selected features has the lowest forecasting errors. Moreover, the performance of the proposed method have been evaluated for both 10-minute and 1-hour forecasting intervals. The obtained results from historical data of two different wind farms justify that the proposed method can yield better performance in aspects of prediction accuracy and processing speed in comparison to some previous reported algorithms. For further studies, the influence of some climate parameters such as wind direction and wind speed can be considered as potential input features.

References

[1] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, Renew. Sustain. Energy Rev. 34 (2014) 243–254.
[2] I. Colak, S. Sagiroglu, M. Yesilbudak, Data mining and wind power prediction: a literature review, Renew. Energy 46 (2012) 241–247.
[3] L. Landberg, Short-term prediction of the power production from wind farms, J. Wind Eng. Ind. Aerodyn. 80 (1999) 207–220.
[4] M.S. Roulston, D.T. Kaplan, J. Hardenberg, L.A. Smith, Using medium-range weather forecasts to improve the value of wind energy production, Renew. Energy 28 (2003) 585–602.
[5] L. Lazic, G. Pejanovic, M. Zivkovic, Wind forecasts for wind power generation using the Eta model, Renew. Energy 35 (2010) 1236–1243.
[6] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, E. Feitosa, A review of the young history of wind power short-term prediction, Renew. Sustain. Energy Rev. 12 (2008) 1725–1744.
[7] J.L. Torres, A. Garcia, M. De Blas, A. De Francisco, Forecast of hourly average wind speed with ARMA models in Navarre (Spain), Sol. Energy 79 (2005) 65–77.
[8] R. Kavassery, K. Seetharaman, Day-ahead wind speed forecasting using f-ARIMA models, Renew. Energy 34 (2009) 1388–1393.
[9] H. Madsen, G. Kariniotakis, H.A. Nielsen, T.S. Nielsen, P. Pinson, A protocol for standardizing the performance evaluation of short-term wind power prediction models. in: Proceedings of the global wind power conference and exhibition, Chicago, US, 2004, pp. 28-31.
[10] H. Madsen, P. Pinson, G. Kariniotakis, H.A. Nielsen, T.S. Nielsen, Standardizing the performance evaluation of short-term wind power prediction models, Wind Energy 29 (2005) 475–489.
[11] J. Yan, K. Li, E. Bai, Z. Yang, A. Foley, Time series wind power forecasting based on variant Gaussian process and TLBO, Neurocomputing (2016), in press.
[12] J.P.S. Catalao, H.M.I. Pousinho, V.M.F. Mendes, Short-term wind power forecasting in Portugal by neural networks and wavelet transform, Renew. Energy 36 (2011) 1245–1251.
[13] K. Bhaskar, S.N. Singh, AWNN-assisted wind power forecasting using feed-forward neural network, IEEE Trans. Sustain. Energy 3 (2012) 306–315.
[14] L. Han, C.E. Romero, Z. Yao, Wind power forecasting based on principle component phase space reconstruction, Renew. Energy 81 (2015) 737–744.
[15] C. Gonzalez-Mingueza, F. Munoz-Gutierrez, Wind prediction using weather research forecasting model (wrf): a case study in Peru, Energy Convers. Manag. 81 (2014) 363–373.
[16] S. Fan, J. Liao, R. Yokoyama, L. Chen, W.J. Lee, Forecasting the wind generation using a two-stage network based on meteorological information, IEEE Trans. Energy Convers. 24 (2009) 474–482.
[17] S. Salcedo-Sanz, A.M. Perez-Bellido, E.G. Ortiz-Garcia, A. Portilla-Figueras, L. Prieto, F. Correoso, Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks, Neurocomputing 72 (2009) 1336–1341.
[18] H. Liu, C. Chen, H. Tian, Y. Li, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, Renew. Energy 48 (2012) 545–556.
[19] H. Liu, H. Tian, X. Liang, Y. Li, New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks, Renew. Energy 83 (2015) 1066–1075.
[20] G. Li, J. Shi, On comparing three artificial neural networks for wind speed forecasting, Appl. Energy 87 (2010) 2313–2320.
[21] H. Liu, H. Tian, X. Liang, Y. Li, Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks, Appl. Energy 157 (2015) 183–194.
[22] S. Salcedo-Sanz, A. Pastor-Sanchez, A. Blanco-Aguilera, L. Prieto, R. Garcia-Herrera, Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – extreme learning machine approach, Energy Convers. Manag. 87 (2014) 10–18.
[23] Y. Liu, J. Shi, Y. Yang, W.J. Lee, Short-term wind-power prediction based on wavelet transform – support vector machine and statistic-characteristics analysis, in: Proceedings of the IEEE Industrial and Commercial Power Systems Technical Conference (I&CPS), 1–5 May 2011, Baltimore, USA.
[24] B. Zhu, M.Y. Chen, N. Wade, L. Ran, A prediction model for wind farm power generation based on fuzzy modelling, Procedia. Environ. Sci. 12 (2012) 122–129.
[25] G.J. Osorio, J.C.O. Matias, J.P.S. Catalao, Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information, Renew. Energy 75 (2015) 301–307.
[26] S. Salcedo-Sanz, A.M. Perez-Bellido, E.G. Ortiz-Garcia, A. Portilla-Figueras, L. Prieto, F. Correoso, Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks, Neurocomputing 72 (2009) 1336–1341.
[27] M. Khashei, M. Bijari, G. Raissi-Ardali, Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs), Neurocomputing 72 (2009) 956–967.
[28] C. Ren, N. An, J. Wang, L. Li, B. Hu, D. Shang, Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting, Knowledge-Based Syst. 56 (2014) 226–239.
[29] H.-Z. Li, S. Guo, C.-J. Li, J.-Q. Sun, A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm, Knowledge-Based Syst. 37 (2013) 378–387.
[30] X. Kong, X. Liu, R. Shi, K.Y. Lee, Wind speed prediction using reduced support vector machines with feature selection, Neurocomputing 169 (2015) 449–456.
[31] X. An, D. Jiang, C. Liu, M. Zhao, Wind farm power prediction based on wavelet decomposition and chaotic time series, Expert Syst. Appl. 38 (2011) 11280–11285.

[32] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (2014) 531–544.
[33] I. Guyon, Practical feature selection from correlation to causality, Mining Massive Data Sets for Security, IOS Press, Amsterdam, 2008, pp. 27–44.
[34] H. Stoppiglia, G. Dreyfus, R. Dubois, Y. Oussar, Ranking a random feature for variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1399–1414.
[35] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[36] G.-B. Huang, L. Chen, C.-K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Trans. Neural Netw. 17 (2006) 879–892.
[37] G.-B. Huang, L. Chen, Convex incremental extreme learning machine, Neurocomputing 70 (2007) 3056–3062.
[38] Y.-W. Huang, D.-H. Lai, Hidden node optimization for extreme learning machine, in: Proceedings of the AASRI Procedia Conference on Modelling Identification and Control, 3, 2012, pp. 375–380.
[39] ⟨http://sotaventogalicia.com/en/real-time-data/historical⟩.
[40] National Renewable Energy Laboratory [Online], Available at ⟨http://wind.nrel.gov/Web_nrel/⟩.



Ali Akbar Abdoos received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from Semnan University, Semnan, Iran, in 2006, 2008 and 2013 respectively. He is now with Babol Noshrvani University of technology as assistant professor. He is interested in power system protection and application of artificial intelligence in power systems.