This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

Improved EMD-Based Complex Prediction Model for Wind Power Forecasting

Oveis Abedinia, *Member, IEEE*, Mohamed Lotfi, *Member, IEEE*, Mehdi Bagheri, *Member, IEEE*, Behrouz Sobhani, Miadreza Shafie-khah, *Senior Member, IEEE*, and João P. S. Catalão, *Senior Member, IEEE*

***Abstract*-As a response to rapidly increasing penetration of wind power generation in modern electric power grids, accurate prediction models are crucial to deal with the associated uncertainties. Due to the highly volatile and chaotic nature of wind power, employing complex intelligent prediction tools is necessary. Accordingly, this paper proposes a novel improved version of empirical mode decomposition (IEMD) to decompose wind measurements. The decomposed signal is provided as input to a hybrid forecasting model built on a bagging neural network (BaNN) combined with K-means clustering. Moreover, a new intelligent optimization method named ChB-SSO is applied to automatically tune the BaNN parameters. The performance of the proposed forecasting framework is tested using different seasonal subsets of real-world wind farm case studies (Alberta and Sotavento) through a comprehensive comparative analysis against other well-known prediction strategies. Furthermore, to analyze the effectiveness of the proposed framework, different forecast horizons have been considered in different test cases. Several error assessment criteria were used and the obtained results demonstrate the superiority of the proposed method for wind forecasting compared to other methods for all test cases.**

***Index Terms*-Wind Forecasting, Wind Power, Neural Networks, Optimization Methods.**

NOMENCLATURE

Abbreviations

IEMD	Improved empirical mode decomposition
BaNN	Bagging neural networks
PM	Persistence method
ARMA	Auto-regressive and moving average
ARIMA	Auto-regressive integrated moving average
GP	Gaussian process
KF	Kalman filtering
MOWOA	Multi-objective whale optimization algorithm
GARCH	Generalized autoregressive conditional heteroscedasticity
SSO	Shark smell optimization
IMFs	Intrinsic mode functions
HHT	Hilbert-Huang transforms
MFDFA	Multi-fractal de-trended fluctuation analysis
BPNN	Back propagation neural network
PCA	Principal component analysis
EM	Expectation maximization
NNs	Neural networks
SVM	Support vector machine
AUC	Area under curve
ChB-SSO	Chaotic binary shark smell optimization
HIFM	Iterative Forecast Method
RMSE	Root Mean Square Error
NRMSE	Normalized value
ARMAX	Auto-regressive moving average
MAPE	Mean Absolute Percentage Error
RBFNN	Radial Basis Function Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
NNPSO	NN based Particle Swarm Optimization
WT	Wavelet Transform
NMAE	Normalized Mean Absolute Error
PJ-ADMM	Dual decomposition and proximal Jacobian
HSRM	Heteroscedastic spline regression model
RSRM	Robust spline regression model
MTD	Mean trend detector
MMLP	Local predictor
TLGP	Temporally Local Gaussian Process

Variables

$x(t)$	Input signal
$f_{s,t}$	Inversed mean non-stationary frequency $\hat{x}_{s,t}$
$\hat{y}_{s,t}$	Scale-dependent trend Split the trend
$\varepsilon_{s,t}$	Oscillation
$F_q(s)$	Total oscillation function
$h(q)$	Globalized Hurst exponent
$\tau(q)$	Conventional function
α	Singularity exponent
$f(\alpha)$	Multifractal singularity spectrum
WS_{max}	Maximum of wind speed
WS_{min}	Minimum of wind speed
WS_{mean}	Mean value of wind speed
T_{max}	Maximum of temperature
T_{min}	Minimum of temperature
T_{mean}	Mean value of temperature
WP_{max}	Maximum of daily wind generation
WP_{min}	Minimum of daily wind generation
WP_{mean}	Mean value of daily wind generation
N_k	Number of data points in the kth cluster
x_i^k	ith element of the kth cluster
a	Stochasticity in the range of [0,1]
$x_0(t-1)$	Best clustering centroid evaluated up to $t-1$
$K(P(t-1))$	K-means transformation
$Y(t)$	Final desired output
z	Output of hidden layer
u	Output of the result layer
e	Error value
w_i	Hidden layer weights
$C_{bag}(x)$	Optimized arranger
H	Elected based on mean magnitude of K models
S_N	Rated generation capacity of the wind farm,
$X_{ACT(t)}$	Actual values of wind speed/power at hour t
$X_{FOR(t)}$	Forecast values of wind speed/power at hour t
N	Number of hours

I. INTRODUCTION

A. Background and Motivation

Concerns over global warming and the planetary effects of climate change continue to drive the need to restructure energy procurement approaches. Novel energy market regulations and power generation technologies are largely influenced by efforts to mitigate adverse effects of greenhouse gas emissions. Fig. 1 shows that while there is starting to be a noticeable share of clean and renewable sources, there remains a significant global dependence on thermal assets in the energy sector [1-2].

M. Shafie-khah acknowledges the support by FLEXIMAR-project (Novel marketplace for energy flexibility), which has received funding from Business Finland Smart Energy Program, 2017-2021. J.P.S. Catalão acknowledges the support by FEDER funds through COMPETE 2020 and by Portuguese funds through FCT, under POCI-01-0145-FEDER-029803 (02/SAICT/2017). (*Corresponding authors: Miadreza Shafie-khah* and *João P. S. Catalão*).

O. Abedinia and M. Bagheri are with the Electrical and Computer Engineering Department, Nazarbayev University, Nursultan, Kazakhstan and also with the National Laboratory Astana, Nursultan, Kazakhstan (e-mails: oveis.abedinia@un.edu.kz; mehdi.bagheri@nu.edu.kz).

B. Sobhani is with the Electrical Engineering Department, Faculty of Technical Engineering, University of Mohaghegh Ardabili, Ardabil, Iran (e-mail: b.sobhany@gmail.com).

M. Shafie-khah is with the School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland (e-mail: mshafiek@univaasa.fi).

M. Lotfi and J.P.S. Catalão are with INESC TEC and the Faculty of Engineering of the University of Porto (FEUP), Porto 4200-465, Portugal (e-mails: mohd.f.lotfi@gmail.com; catalao@fe.up.pt).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

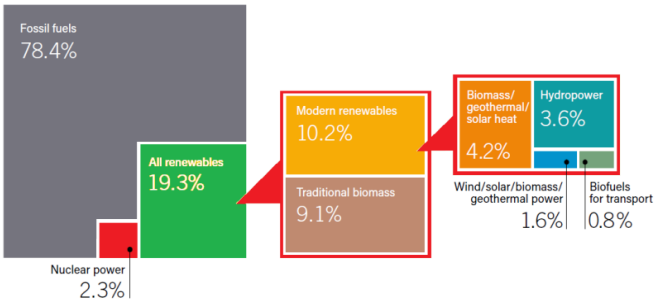

Fig. 1. Detailed breakdown showing the share of different renewable energy sources of the global total final energy consumption as of 2015.

A leading solution thereof is the expanded incorporation of wind generation. Since the supply-demand balance of electrical grids is very important, precise forecasting approaches should be implemented in order to enable optimal operation of the systems. To address this issue, various prediction methods have been addressed in scientific literature over recent years, which are presented subsequently.

B. Literature Review

Several basic prediction algorithms can be employed to wind forecasting. In [3-4], the persistence method (PM) is demonstrated for short-term forecasting of wind energy. Auto-regressive and moving average (ARMA) models are suggested [5]. In [6], authors extend the latter by providing the auto-regressive integrated moving average (ARIMA) model, with the aim of improving the prediction accuracy. Another model, the Gaussian process (GP) is shown in [7]. Another famous approach in this context is Kalman filtering (KF), as presented in [8].

All the previously listed approaches are based on time-domain models which can face difficulties predicting the highly nonlinear, chaotic, wind signals [9]. Therefore, more sophisticated methods have been developed which are better-tailored for this purpose. In [10], a dynamic prediction method for wind speed is introduced. This model consists of two stages: data analysis and forecasting. The process of working the optimal regression model is initially applied to the signal and then experimental results statistically verify the performance of the forecasting model.

In [11], wind speed predication was performed by means of a novel algorithm called the multi-objective whale optimization algorithm (MOWOA). An ensemble wind forecasting model was introduced in [12]. The model aims increase prediction accuracy by means of deep feature extraction to filter the input candidate. Another prediction model based on different generalized autoregressive conditional heteroscedasticity (GARCH) for wind power is proposed in [13], which presents the anticipated instability of the wind signal as a benchmark for latent instability. The improved prediction model derived from multi-step prediction into waves based on the synoptic background is presented in [14]. This model is a hybrid combination of fuzzy sets, inference rules, and optimization methods.

A framework for wind speed forecasting, described in [15], consists of de-trending, subspace identification, and Kalman filtering. Spatio-temporal forecasting of wind speed was developed in [16], consisting of wavelet decomposition and structured-sparse recovery. Another method for predicting wind speed is the Markov-switching model, which is demonstrated in [17]. Bayesian inference is used to predict the Markov-switching model parameters. With the aim of finding the most suitable configuration for wind speed and wind power predict, the Kalman filter is proposed in [18]. In Ref [1], a dual decomposition and proximal Jacobian (PJ-ADMM) model for forecasting of price and power generation in decentralized energy trading systems was presented. A machine learning forecasting model based on numerical weather prediction and meteorological data is presented in [19]. ANN, SVR, and Gaussian process, were the three machine learning approaches applied conjunctively in this model. A new multi structure combination approach is proposed in [20] for probabilistic forecasting of wind speed, compatible for combination with other models for an ensemble forecast. Providing various probability density functions to reduce the prediction error is the main advantage of this model. In order to show the effectiveness, Bayesian learning, kernel density estimation, and beta distribution fitting prediction methods were applied to the test cases.

In Ref [21], a reliable method to determine an accurate uncertainty value for wind power generation using a MILP optimization model is proposed. In Ref [22], a combined technique based on the heteroscedastic spline regression model (HSRM) and robust spline regression model (RSRM) was formulated to achieve the minimum error of wind curve prediction in presence of the inconsistent samples.

A short-term wind speed prediction method based on mean trend detector (MTD) and local predictor (MMLP) is presented in [23]. In this model, wind signals are predicted by combining of two signals: the mean trend based on MTD analysis and the stochastic component based on the MMLP method. In Ref. [24], a Temporally Local Gaussian Process (TLGP) is used for interval prediction of wind power to be used in the unit commitment problem and energy management systems applications.

An SVM improved by Markov model is developed in [25] based on local data to wind speed prediction. This model is used on historical real wind farm data and finite-state Markov model analysis to illustrate the performance and effectiveness of distributional and point prediction results. Apart from linear strategies, nonlinear ones are autonomous in terms of the time series data. From those, NN- and support vector machine (SVM) strategies are more effective in wind forecasting. However, a large quantity of historical data as well as an accurate training model are necessary for those models.

C. Novel Contributions

In this paper, we propose a prediction approach based on the improved empirical mode decomposition (IEMD) in conjunction with a hybrid framework consisting of the bagging neural network (BaNN), K-means clustering method, and a stochastic optimization algorithm. The contributions of the proposed prediction model can be summarized as follows:

- Considering an improved version of empirical mode decomposition, IEMD, adapted for wind power and wind speed signals. Hence, the volatility and fluctuation behavior of wind signal during the forecasting process can be reduced.
- Developing of a filtering and grouping model based on improved k-means for wind signal prediction.
- Developing BaNN-based hybrid forecasting. The proposed forecasting engine is enhanced by using an optimization algorithm to fine-tune BaNN parameters.
- Incorporating a new optimization algorithm based on shark smell optimization (SSO) to tune the free parameters of the forecast engine.

D. Paper Organization

This manuscript is organized as follows: In Section 2, the IEMD model is presented; Section 3 provides the proposed clustering model and forecasting engine; Implementation of the new optimization algorithm (SSO) is addressed in Section 4; Numerical analyses and results are evaluated in Section 5; Finally, the conclusions are presented in Section 6.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

II. Improved Empirical Mode Decomposition

In this section, the proposed forecasting approach is elaborated in detail. The flowchart in Fig. 2 demonstrates the algorithm being proposed. Details of the formulation are subsequently presented in this section.

The implementation of EMD as an algorithm for adaptive data processing in nonlinear and non-stationary time series was first presented in [26]. This method is widely employed in the analysis of nonlinear and non-stationary signals. In practice though, it faces some major challenges [26]. The most prominent one is the usual generation of a number of intrinsic mode functions (IMFs) with complex-valued pseudo components in areas with lower frequencies. Despite the real IMF components that have relatively close resemblance to the original signal, the pseudo components have a distant one.

According to [26], analysis of the interactions among IMFs is implemented through the removal of pseudo components. Additionally, the input signals are often corrupted in practice, and the signal noise influences a number of extracted IMFs. A statistical method named Kolmogorov-Smirnov test (K-S test) is implemented in detection of the difference between the basic distribution and the presumed one.

Also, to evaluate highly accurate data, the spectrums from both the Fourier and Hilbert-Huang transforms (HHT) of the IMFs are calculated in the IEMD. Due to the fact that each IMF is practically mono component, in this utilization, the two-sample K-S test is implemented to examine the statistical affinity between Gaussian noise and each IMF. The IMFs with affinity to the noise are omitted and the rest is accepted.

Combining the two aforementioned methods, the IEMD algorithm using IMF extraction is performed in four steps, for an input time series $x(t)$:

(1) *Decompose* $x(t)$ to obtain n IMFs and a residue.

(2) *Normalize* all n IMFs with respect to $x(t)$, then *compute* the correlation coefficients, cor_t (i = 1,...,n), for each i IMF to signal x(t).

(3) *Compare* $cor_t(i)$ against a threshold X (can be the ratio of the maximum correlation factor λ, such that $X=max(cor_t(i))/\eta$,(i=1,…,n) = λ/η, where η is a larger ratio factor of 1.0). IMFs with a cor_i larger than the threshold are kept and the remaining ones are discarded with the residue.

(4) *Apply* K-S testing on the IMFs. The aggregated density function should be computed, so that the magnitude of identical probability p(D) for both the Gaussian noise and the IMFs can be obtained via K-S testing. If the distributions are found to be identical, the null presumption is applied, else the alternative one will be implemented.

The EMD algorithm in the proposed IEMD-based multi-fractal de-trended fluctuation analysis (MFDFA) method, helps pinpoint the trends that are scale-dependent, i.e. $\hat{x}_{s,t}$. Initially, the time scale is defined for each $c_i(t)$ using the inversed mean non-stationary frequency $f_{s,t}$, and this can be a representation of oscillation mode in the time domain. We can calculate the $f_{s,t}$ using Hilbert transform of $c_s(t)$.

$$s = 1/ < f_{s,t} > . \tag{1}$$

Afterwards, the scale-dependent trend $\hat{x}_{s,t}$ is calculated as follows, where [s] represents the set of scales larger than s.

$$\hat{x}_{s,t} = \sum_{[s]} c_s(t), \tag{2}$$

The MFDFA algorithm is then carried out according to the following steps.

Step 1: Assume $x(t)$ in which $t=1,\ldots,N$ is the data set with the length of N, and the profile $y(t)$ can be constructed as:

$$y(t) = \sum_{k=1}^{t} [x(k) - < x >], t = 1,\ldots,N, \tag{3}$$

in which is the average of $x(t)$ is denoted by <x>.

Step 2. The IEMD method assumes the decomposition of the profile $y(t)$ into two components: $c_i(t)$ and $r_n(t)$, the IMFs and residual, respectively. With the IMFs, $c_i(t)$ $i=1,\ldots,n$, finally the $y(t)$ can be calculated:

$$y(t) = \sum_{i=1}^{n} c_i(t) + r_n(t). \tag{4}$$

Step 3. By using the results of the decomposition, split the trend $\hat{y}_{s,t}$ in every scale and obtain the trend in scale s using the summation of the residue and IMFs with a scale larger than s. Afterwards, deduct the trend in this scale from the profile $y(t)$ to compute the associated oscillation $\varepsilon_{s,t}$.

$$\varepsilon_{s,t} = y(t) - \hat{x}_{s,t}. \tag{5}$$

Step 4. Each oscillation $\varepsilon_{s,t}$ is split into N_s non-overlapping components with equal lengths of s. Due to the fact that N_s is seldom a multiple of the given s, inescapably, a short section of the oscillation will be omitted. To deal with this issue, the same procedure is repeated in the reverse direction. Thus, $2N_s$ segments are obtained totally.

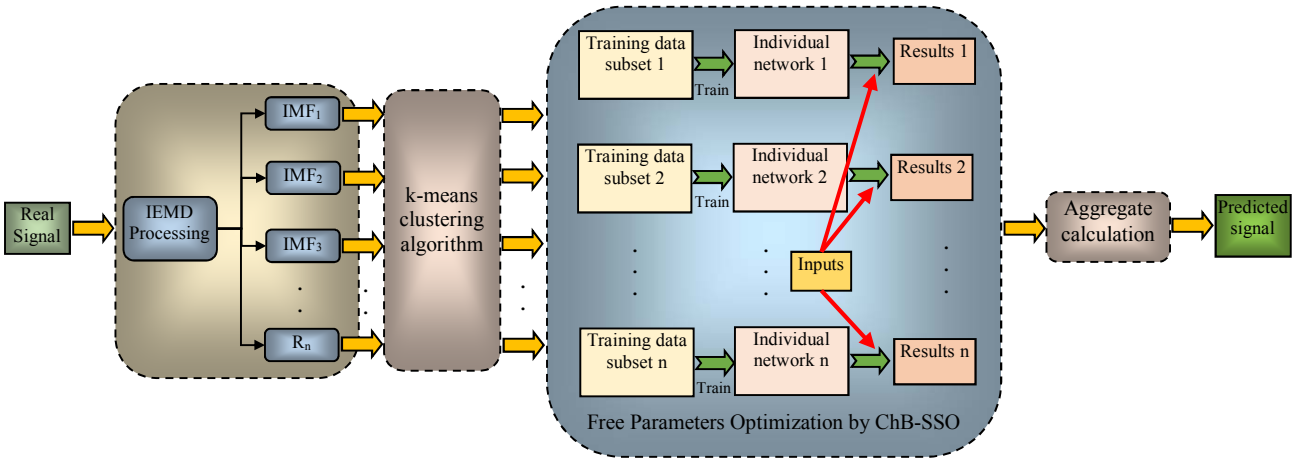

Fig. 2. A flowchart demonstrating the process flow of the proposed hybrid forecasting algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

Step 5. Computation of the de-trended oscillation function *F(v,s)* should be conducted for $2N_s$ total segments. Every segment is represented as ε_v and includes *s* elements. The following formula is used, in which there are $v = 1,\ldots,2N_s$ for the *vth* segment.:

$$F^2(v,s) = \frac{1}{2}\sum_{i=1}^{s}[\varepsilon_v(i)]^2, \quad (6)$$

Step 6. Using the mean of all $2N_s$ segments, the total oscillation function $F_q(s)$ of *qth* order is computed:

$$F_q(s) = \left\{\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(v,s)]^{\frac{q}{2}}\right\}^{\frac{1}{q}}, \quad (7)$$

in which the index *q* can typically have any non-zero value. In case of q=0, the following logarithmic formula which calculates the average should be implemented instead:

$$F_0(s) = \exp\left\{\frac{1}{4N_s}\sum_{v=1}^{2N_s}\ln[F^2(v,s)]\right\}, \quad (8)$$

This should be applied to all scales calculated in step 4 and other values for *q* as well.

Step 7. For each *q*, obtain the scaling pattern by using a logarithmic function for $F_q(s)$ and *s* as follows

$$F_q(s) \sim s^{\square(q)}, \quad (9)$$

in which, the globalized Hurst exponent is presented by *h(q)*. The associated conventional *τ(q)* function for each *q* can then be calculated:

$$\tau(q) = qh(q) - 1 \quad (10)$$

To compute the singularity exponent and multifractal singularity spectrum, *α and f(α)*, respectively, the Legendre transform is used:

$$\alpha = \tau'(q) = h(q) + qh'(q), \quad (11)$$

$$f(\alpha) = q\alpha - \tau(q), \quad (12)$$

in which the time series local roughness is denoted by *α* and the multifractal singularity spectrum *f(α)* can be considered as the subset's fractal dimension with the same singularity strength *α*. This is generally used for measuring the time series local dynamics.

III. Proposed Forecasting Approach

Data extraction methods are extensively employed in categorizing and forecasting procedures. The employed method in this paper uses data extraction and includes K-means clustering and BaNN.

Initially, the data is refined in order to omit inconsistencies in the vector set, after which learning samples are selected and members with maximum relationship are elected to be input into the neural network.

Next, after being refined, the data is grouped using the K-means grouping method for election of the learning group with most affinity to the prediction day. Ultimately, wind generation is predicted using BaNN which has the ability to mitigate problems regarding volatility and over fitting associated with their back propagation neural network (BPNN).

A. Data Refinement

Some parameters regarding the wind generation are gathered to be fed as learning inputs by means of sensor section. Though, they might include some inconsistent data. Moreover, over fitting, i.e. feeding too many samples to the learning section, causes computational complications and may yield unwanted outputs. This is due to the fact that a portion of parameters in the procedure are inappropriate and repetitive. The election of characteristics with maximum relation with wind generation can help enhance the precision of prediction. Data normalization impacts the convergence and precision in the learning process. Refinement of the data is essential for the purpose of achieving precise outcome for the prediction.

- *Elimination of data*: some data in the initial samples might include inconsistent values. For instance, in cases where wind speed or power measurements have negative values. The gap of omitted data should therefore be crowded. The average magnitude procedure is applied in order to determine the average magnitude before and after unwanted data is omitted.
- *Election of candidate inputs*: Conceptually, the more parameters input to the model, the better the differentiation obtained. In practice however, feeding too many parameters might lead to several undesired complications. Hence, choosing a proper group of entry parameters from unprocessed data affects the prediction efficiency significantly. Relief technique is a method used to assign weights to the characteristics, which is done considering the association. Characteristics with weights less than a specified minimum are omitted. The mathematical formulation of the Relief technique is demonstrated in [27]. This method is very efficient due to the fact that there is a linear relationship between the computational time and the sampling time and initial characteristics, i.e. *O(n)*. As compared to other methods such as principal component analysis (PCA), this technique is capable of decreasing the physical problem dimension
- *Normalization of data:* in this step, the unprocessed data is converted to similar orders of scale to obtain faster convergence and more precise prediction. The Normalization process is done considering the simple min-max procedure which is formulated as $\bar{x} = (x - min)/(max - min)$, in which *x* is the initial data and max and min denote the maximal and minimal values of the learning group, respectively. The resulting values in the normalized data set, $\bar{x}$, range from 0 to 1.

The aforementioned steps are essential for obtaining more precise prediction outcomes. Following the data refinement, the proposed method can now be applied and afterwards compared with previously published WPF method.

B. Clustering of Identical Days using Climatic Parameters and Historical Data

In regular operation, wind turbines provide large sets of measured data. The learning inputs can be considered as the time series, and auto regressive procedure is usually applied to model prediction. Nonetheless, the stochastic nature of wind usually results in disparity of learning inputs. Hence, the usage of those learning inputs for achieving the required outcomes may prove to be problematic. To deal with this, a recommended practice is to identify days within the dataset with similar wind behavior and accordingly clustering them in groups of learning inputs. Following this approach enhances the prediction precision significantly.

B.1. Clustering Inputs

The output of wind generation is greatly influenced by meteorological factors such as wind speed and temperature. As previously elaborated, selecting inputs whose data are most analogous in terms of wind speed and temperature from the historical data set is essential. An input including wind speed and temperature is taken as clustering criterion and is defined as follows:

$$S_1=[WS_{max}, WS_{min}, WS_{mean}, T_{max}, T_{min}, T_{mean}] \quad (13)$$

in which WS_{max}, WS_{min}, WS_{mean}, T_{max}, T_{min} and T_{mean} denote maximum, minimum, and mean value of wind speed and temperature, respectively.

It is important to mention that while wind speed and temperature greatly influence wind generation and clustering the training inputs accordingly is greatly beneficial; identical values on two different instances do not guarantee equal generation due to the complexity of the problem with various other dependencies such as terrain, the operation state of generators, and other meteorological factors.

Hence, in addition to meteorological factors, historical data of wind generation associated with current the prediction day is also employed in clustering. As such, the input S2 can be formulated as follows

$$S_2=[WP_{max}, WP_{min}, WP_{mean}] \quad (14)$$

in which $WP_{max}, WP_{min}, WP_{mean}$ denote the maximum, minimum, and mean value of wind generation the daily values, respectively.

Though S_2 for a prediction day is unidentified, we can consider the day before prediction day as a day with a similar S_2. Based on the assumptions in wind generation, in a given season, the generation trends are similar for consecutive days. As such, the samples $S1$ and $S2$ are clustered. Cluster 1 in S_1 and cluster 2 in S_2 are considered to have the most affinity with the prediction day. Ultimately the similar days in clusters 1 and 2 are used as learning inputs for the prediction model.

B.2. Clustering Technique

Clustering is a part of unmonitored training and contains no label for learning inputs. This study implements clustering for extraction of identical samples and classifying them to a group. Conventional clustering technique includes K-means clustering and expectation maximization (EM).

Because of its ability to deal with copious data groups, the K-means technique is implemented in this study. K-means technique is a clustering procedure which uses subdivision and determines the affinity based on gap. The main essence of the technique is random election of k main points and subdivision of the data based on their distance to the main points. Data is groups with their nearest main points of P_k using the Euclidean distance.

This procedure is formulated as follows:

$$P_k = \frac{1}{N_k}\sum_{i=1}^{N_k} x_i^k, \quad (15)$$

in which N_k is the number of data points in the kth cluster and x_i^k is ith element of the kth cluster. The clustering point should be renewed by recalculating the mean value of each cluster and this procedure continues until same values are obtained. Next, the original group is subdivided to k clusters with highly similar data.

B.3. Improved Clustering Method

In Eq. (15), if x is considered as a function of the cluster centroid, it can be written as $E(x)$:

$$E(x(i)) = \min(E(x(1)), E(x(2)), ..., E(x(t))) \quad i=(1,2,...,t) \quad (16)$$

$$x_0(t)=x(i) \quad (17)$$

As such, the following equation can be used to adjust the clustering centroid:

$$x(t)=aK(x(t-1))+(1-a)x_0(t-1) \quad (18)$$

In this equation, a reflects the stochasticity and is set in the range of [0,1]. $x_0(t-1)$ defines the best clustering centroid evaluated up to $t-1$. $K(P(t-1))$ is the K-means transformation through centroid optimization at t-1. In practical application of the proposed model, the best method can be evaluated based on Eq. (18). This formula can implemented by probability r_0 and following equation:

$$x(t)=K(x(t-1)) \quad (19)$$

The clustering technique used in this study can now be described using the following iterative procedure:

- **Step 1:** K points are selected as initial main points for clustering, from the data group with population N.
- **Step 2:** a data point is clustered to the nearest group, based on the computed Euclidean distance to the main points.
- **Step 3:** the average magnitude of the cluster is calculated and used to reselect the clustering main point.
- **Step 4:** the procedures of steps 2 and 3 are repeated and the stops when the same result is obtained.

K-means clustering is implemented on S_1 and S_2 data groups. The data groups are divided into M parts and the learning group is $D=d_1,d_2,...,d_M$, corresponding to the learning days.

The K-means technique is able to choose identical days from D and classify them to a group, which leads to creation of several groups. Based on the Pearson correlation coefficient, the group with the most association is ultimately chosen to feed the neural network.

C. Bagging Neural Networks (BaNN)

Neural networks (NNs) are one of the most efficient methods for data extraction for forecasting. They can easily handle nonlinear problems without the need to formulate a complex mathematical model. Many various variations and structures of NNs exist for different applications. BaNN is widely implemented in prediction techniques and is an enhancement of the simpler backward propagation NN. It is generally formed of three layers: feeding layer, hidden layer, and result layer. The BPNN operation includes two processes: forward propagation of data and backward propagation of the error signal. The conditions of neurons in one layer only influence the next layer in forward propagation. Backward propagation of the error value is performed until the desired result is achieved. The gradient descent procedure is applied to the vector space of the weights. This procedure is essential for dynamic inquiry of a group of weight vector and minimization of error value. Empirically, $2M+1$ nodes should be used in the hidden layer, in which M denotes the nodes of the (previous) feeding layer, T. For optimal prediction, it is recommended that the network is constructed with M between 2 and 10 nodes.

In Fig. 3, the structure of a typical BPNN is shown, in which $[x_1,x_2,...,x_n]$ corresponds to the input vector for the feeding layer. $[w_1,w_2,...,w_n]$ denotes weights among the feeding layer and the hidden layer.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

Weights for the hidden and result layer are denoted by $[v_1, v_2, \ldots, v_n]$. Finally, $\underline{u}$ corresponds to the output of the result layer. In order to calculate the final desired output $Y(t)$, the general Sigmoid function is applied as a relocation function to deal with the problem nonlinearity. This is demonstrated in the following equations (20)-(22):

$$y = f(x) = \frac{1}{1+e^{-x}} \tag{20}$$

The output of hidden layer, z, can be expressed as follows:

$$z = f_1 \sum w_i x_i \tag{21}$$

The output of the result layer, u, can then be expressed as:

$$u = f_2 \sum v_k z_k \tag{22}$$

At this stage, forward propagation is complete. For the backward propagation, the error value e must first be calculated from u and $Y(t)$ as follows:

$$e = \frac{1}{2} \sum (y_i(t) - u_i)^2 \tag{23}$$

IN BPNN, the gradient descent procedure is applied to readjust the hidden layer weights, w_i, in order to minimize the error value. This procedure is applied iteratively as a feedback loop until the error value e converges to zero, resulting in equal magnitudes of u and $Y(t)$.

Computation of partial derivatives is required in this procedure in order to renew the weights of w_i and v_i and is therefore very computationally demanding. When handling a large set of inputs, the BPNN yields better results compared to other techniques such as linear regression and SVM. However, it is susceptible to being trapped in local minima. Moreover, prediction results are highly sensitive to slight changes in the training samples over-fitting issues.

To overcome the aforementioned issues, ensemble training is implemented to enhance the BPNN. Ensemble training is able to incorporate a number of training modules to improve the model's overall stability and prediction precision. Bootstrap aggregating, otherwise known as the bagging algorithm, is a type of ensemble training procedure which is applied in order to enhance BPNNs, resulting in the more efficient BaNNs.

The bagging technique has proved to be useful for volatile training procedures through enhanced feeble training applied on learning inputs. Since feeble training has low precision, the procedure has to be repeated several times on the learning inputs. Afterwards, a series of prediction equations is created and the one with the highest precision is chosen based on voting. The initial learning data set is sampled anew using Bootstrap sampling method and the training technique is used for the subgroups.

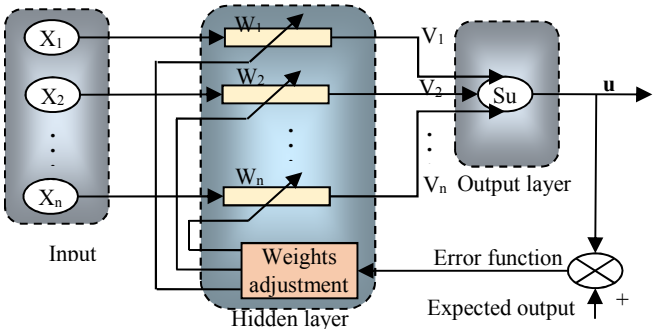

Fig. 3. The structure of a typical three-layered BPNN.

Let $T(x)$ be an arranger, i.e., a tree, which creates a forecasted group label at entry point x. In order to bag T, we need to obtain the samples from the bootstrap technique, $\{(x_i^*, y_i^*)\}^1, \ldots, \{(x_i^*, y_i^*)\}^B$ each of size n with replacement from the training data. Afterwards, the arranger can be formulated as follows:

$$C_{bag}(x) = MajorityVote\{T^{*b}(x)\}_{b=1}^{B} \tag{24}$$

The resulting $C_{bag}(x)$ is an optimized arranger which is chosen based on votes. The most important quality of bagging method which improves the forecasting precision is the stability of its training mechanism.

Initially, the bagging technique is able to arbitrarily select K groups with n samples from the original learning data set.

The overall BaNN is implemented in order to teach subgroups, repeatedly. After that the series $h=\{h_1, h_1, \ldots, h_K\}$ including the prediction model is created. Ultimately the prediction model H is elected based on the computation of mean magnitude of K models. Specifically, all the prediction models are essentially identical. Utilizing the BaNN on the basis of the bagging technique will improve the prediction precision and decrease the computational burden.

IV. Application of Proposed Optimization Algorithm

All free parameters of BaNN are optimized by an optimization method described in this section. The proposed optimization algorithm is named the chaotic binary shark smell optimization (ChB-SSO) algorithm. In [28], a basic version of the algorithm was proposed and applied on some prediction methods in [29]-[30].

The method is a biologically inspired optimization technique, in which the shark's ability to find prey (optimal solution) and move towards it is mathematically simulated. More detailed information of this algorithm can be found in [28]. The enhancement of this algorithm includes two main steps. At first step, the chaotic operator (i.e. logistic map) is incorporated in the SSO algorithm to increase its local and global search capability. A second improvement added to SSO in this study is binary enhancement, leading to the final ChB-SSO model.

By incorporating the developed ChB-SSO algorithm, the training process for the proposed forecast engine is defined as follows:

Step 1: Select the training period (49 days for this study) before the target day, exclusively.

Step 2: Decompose the input signal into different sub-signals using the proposed IEMD.

Step 3: Evaluate candidate inputs, which consisted of past wind power and speed from (200 hours for this study).

Step 4: Elimination of any unwanted data from the initial measurement samples which may include inconsistencies.

Step 5: Election of candidate inputs by the proposed K-means technique.

Step 6: Normalization of data for improved computational performance as detailed in Section 2.

Step 7: Clustering of the identical days from the historical meteorological and power measurements (Section III.B.3).

Step 8: Prediction of the prepared signal using the proposed hybrid BaNN framework.

Step 9: Optimize all the tunable BaNN parameters using the proposed meta-heuristic algorithm.

This iterative procedure is executed until convergence of the error value is achieved.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

The following section demonstrates a case study in which the performance of the proposed method is compared against other methods from previous work. Furthermore, to show the efficiency of proposed model, the training process is presented graphically by comparison with a simple NN forecast engine in Fig. 4.

As shown in this figure, the proposed method has a better ROC chart in comparison with a simple NN and the area under curve (AUC) of the proposed model is higher than the simple NN forecast. As also shown in the figure, the AUC of the proposed model is 0.986 and this value for simple NN based forecast engine is 0.87. Furthermore, the accuracy of the proposed model is 97.4% compared to about 84.7% for the NN. Besides, the sensitivity of the proposed model is 89% compared to 54% for the NN. The specificity value for proposed model and NN are 0.98% and 93%, respectively.

V. Numerical Results and Analysis

Application of proposed prediction approach over different test cases is presented in this section. For this purpose, prediction model is applied on two different real-world standard engineering test cases for comparison with other well-known models. Furthermore, to demonstrate the performance of each individual prediction component, individual analyses are presented.

1. EEMD Analysis

In the first step, the purposed IEMD is applied on the Sotavento test case (on the first November 2005) through the HHT spectrum. As shown in Fig. 5, the first IMF of wind signal (5-a) has high frequency periodic component while and this feature is reduced subsequently. The associated Hilbert time–frequency spectrum is depicted in Fig. (5-b) and the Fourier transform spectrum in Fig. (5-c). In the following, the Sotavento wind farm in Spain and Alberta test case will be presented by related numerical analysis.

A. First Case Study: Sotavento Wind Farm

The Sotavento wind farm was selected as the first case study as it has been already used to validate various models in the past such as the Correlation analysis with Hybrid Iterative Forecast Method (HIFM) [31], two stage MI-MR feature selection with MLP forecast engine [31] and MI-MR feature selection with HIFM [31].

In addition, in order to highlight the impact of IEMD and improved clustering, four variations of the forecasting model are also examined: the proposed model without IEMD and clustering, the proposed model without IEMD, the proposed model without clustering, and finally the proposed prediction model.

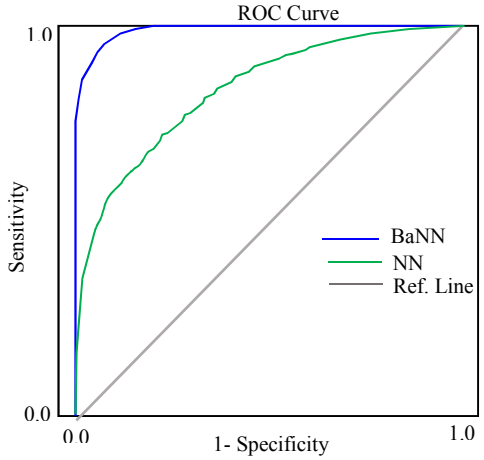

Fig. 4. ROC Curve of BaNN compared to a simple NN forecast engine

TABLE 1
Obtained results for RMSE in day-ahead wind speed forecast of Sotavento wind farm in the four test weeks of year 2005

Test Week	a	b	c	d	e	f	g
Feb.	7.56	7.68	5.71	8.55	4.68	5.31	2.36
May	5.82	5.96	4.26	7.87	4.47	5.23	2.15
Aug.	6.93	7.01	5.92	7.88	4.01	5.02	2.43
Nov.	5.97	6.04	4.55	7.54	4.14	4.77	2.01
Ave.	6.57	6.68	5.11	7.96	4.32	5.08	2.24

a= Correlation + HIFM [31], b= MI-MR feature selection + MLP [31]
c= MI-MR feature selection + HIFM [31], d= Proposed without IEMD and clustering
e= Proposed without IEMD, f= Proposed without clustering model, g=Proposed

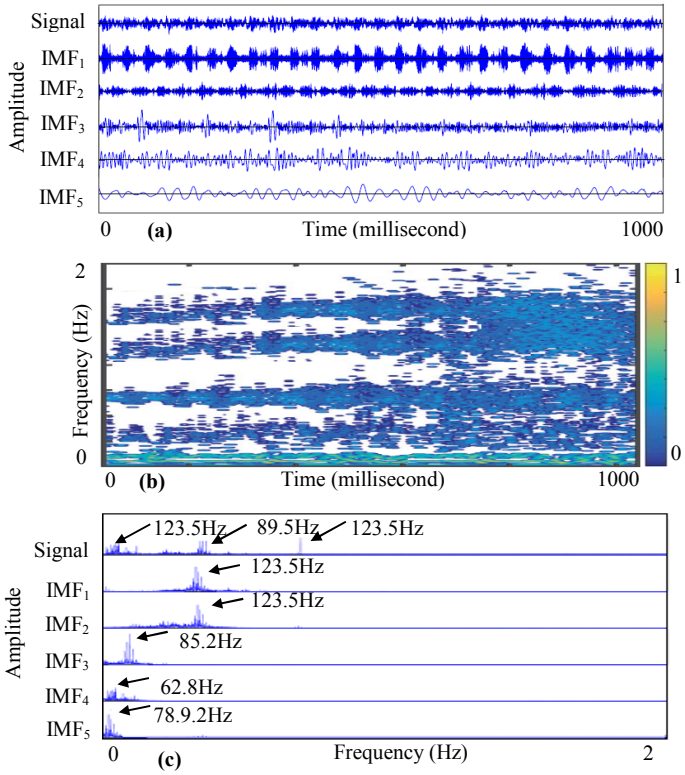

Fig. 5. Decomposition of wind power signal based IEMD in 1th November 2005 for Sotavento test case: (a) original signal and IMFs, (b) Hilbert time–frequency spectrum (c) Fourier transform spectrum in frequency.

In order to guarantee fair comparison, all studies are implemented using the same conditions of 200 lagged hourly values of wind signal, 50 days past to each prediction day are taken as the historical data separated to 49 days as the training set and one day as the validation set. Moreover, the third weeks of February, May, August, and November of 2005 are considered as prediction tests. The obtained numerical results are presented in Table 1. In this table, Root Mean Square Error (RMSE) and its normalized value (NRMSE) are considered as the error evaluation which are calculated as follows:

$$RMSE = \left[\frac{1}{N} \sum_{t=1}^{N} \left(X_{ACT(t)} - X_{FOR(t)} \right)^2 \right]^{1/2} \quad (25)$$

$$NRMSE = \left[\frac{1}{N} \sum_{t=1}^{N} \left(\frac{X_{ACT(t)} - X_{FOR(t)}}{X_N} \right)^2 \right]^{1/2} \quad (26)$$

where S_N is the rated generation capacity of the wind farm, $X_{ACT(t)}$ and $X_{FOR(t)}$ represent the actual and forecast values of the wind speed/wind power signal for hour t, respectively, and N defines the number of hours, which is 168 for test weeks and 24 for test day. The proposed prediction approach can be shown to all compared models as shown in Table I.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

By considering the results in the table, the proposed method improvement can be calculated to be about 56% = (5.11-2.24)/5.11 in comparison with MI-MR feature selection + HIFM. The rate of improvement is about 72% = (7.96-2.24)/7.96 for "d", 48% = (4.32-2.24)/4.32 for "e" and 56% = (5.08-2.24)/5.08 in comparison with "f". This model works on an hourly basis for each day of the prediction horizon. Once each day has elapsed and the wind signal measurements for the 24 hours are available, the historical data can be updated to perform day-ahead wind forecast of the next day. Hence, the forecast horizon consists of 24 forecast steps corresponding to the prediction horizon.

To better demonstrate the comparison of the results, a graphical representation of the results and comparison with other models is shown in the plots of Fig. 6. As depicted in this figure, the proposed model can be seen to better fit the curve of the real values in comparison with other models. In addition, error fluctuation is smaller than the compared models. These results demonstrate the validity of the proposed prediction model. For the second analysis with the Sotavento test case, the proposed model is applied for day-ahead prediction in four test weeks of year 2005 including February, May, August and November.

The obtained results of this comparison are presented in Table 2. In this table, the proposed model is compared with multi-variable auto-regressive moving average (ARMAX) time series, RBF neural network, and MLP neural network. The evaluated error criteria in this comparison are NRMSE and Mean Absolute Percentage Error (MAPE) which is calculated as shown in Eq. (27).

$$MAPE(\%) = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|X_{ACT(t)} - X_{FOR(t)}\right|}{X_{ACT(t)}} \times 100 \qquad (27)$$

To show the effectiveness of the proposed model and compare with other published works, another comparison is made using the four months of April, May, June, and July of 2010 as in [32]. Obtained results are presented in Table 3 through comparison with Persistence method, multivariate ARIMA time series, RBF, multi-layer perceptron (MLP) neural network trained by the efficient Levenberg-Marquardt (LM) learning algorithm, Ridgelet Neural Network (RNN) and hybrid model of mutual information-interaction gain feature selection + neural network + chaotic SSO named (MI-IG + NN + CSSO) [33]. In this table it is shown that the proposed method outperforms all models based on both RMSE and MMAPE error criteria as defined:

$$MMAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|S_{ACT(t)} - S_{FOR(t)}\right|}{S_{AVE-ACT}} \times 100\,,$$
$$S_{AVE-ACT} = \frac{1}{N}\sum_{t=1}^{N} S_{ACT(t)} \qquad (28)$$

The choice of a suitable training model is critical for the proposed forecasting engine. A small training period based on a limited number of samples can result in premature convergence in the learning phase. To prevent this, the forecasting engine is run with different training periods of various lengths and the training period leading to the best forecast results is selected.

To show the efficiency of proposed approach, another comparison is considered based on three training periods of 30 days, 40 days and 50 days. The obtained results are presented in Table 4 based on RMSE and MMAPE error criteria. In this table, the length of 50 days is found to result in the lowest average errors and the proposed model is seen to generate better results in comparison with [33].

TABLE 2
FIRST CASE STUDY: SOTAVENTO (2005) NRMSE AND MAPE (%).

Methods	Error	Feb.	May	Aug.t	Nov.	Ave.
Persistence	NRMSE	16.41	13.93	12.11	23.86	16.57
	MAPE	67.22	56.27	60.27	65.99	62.43
ARMAX	NRMSE	15.52	14.32	11.83	19.33	15.25
	MAPE	61.03	59.24	56.39	60.41	59.26
RBF	NRMSE	14.36	13.04	11.63	17.24	14.06
	MAPE	56.95	53.91	55.75	59.65	56.56
MLP	NRMSE	12.78	12.58	11.42	12.71	12.37
	MAPE	50.12	45.68	54.43	42.34	48.14
Proposed	NRMSE	3.08	2.98	2.44	2.57	2.77
	MAPE	20.34	20.09	19.07	19.63	19.78

TABLE 3
OBTAINED NUMERICAL RESULTS FOR WIND POWER FORECAST OF SOTAVENTO TEST CASE

Test Month	Persistence Method [32]		Multivariate ARIMA [32]		RBF [32]		MLP [32]		RNN [32]		MI-IG + NN + CSSO [33]		Proposed	
	RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE
April 2010	1.124	35.91	0.843	28.74	0.594	25.08	0.514	22.44	0.463	7.75	0.451	8.12	0.342	7.63
May 2010	0.848	30.84	0.742	27.21	0.516	18.11	0.618	19.82	0.435	11.43	0.422	10.12	0.344	9.14
June 2010	0.784	34.33	0.702	28.91	0.593	26.45	0.521	25.24	0.437	16.06	0.434	15.07	0.365	14.05
July 2010	0.826	36.84	0.691	29.54	0.501	27.75	0.467	26.14	0.376	9.33	0.401	8.59	0.347	7.72
Average	0.895	34.48	0.744	28.60	0.551	24.35	0.530	23.41	0.428	11.14	0.427	10.47	**0.35**	**9.63**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

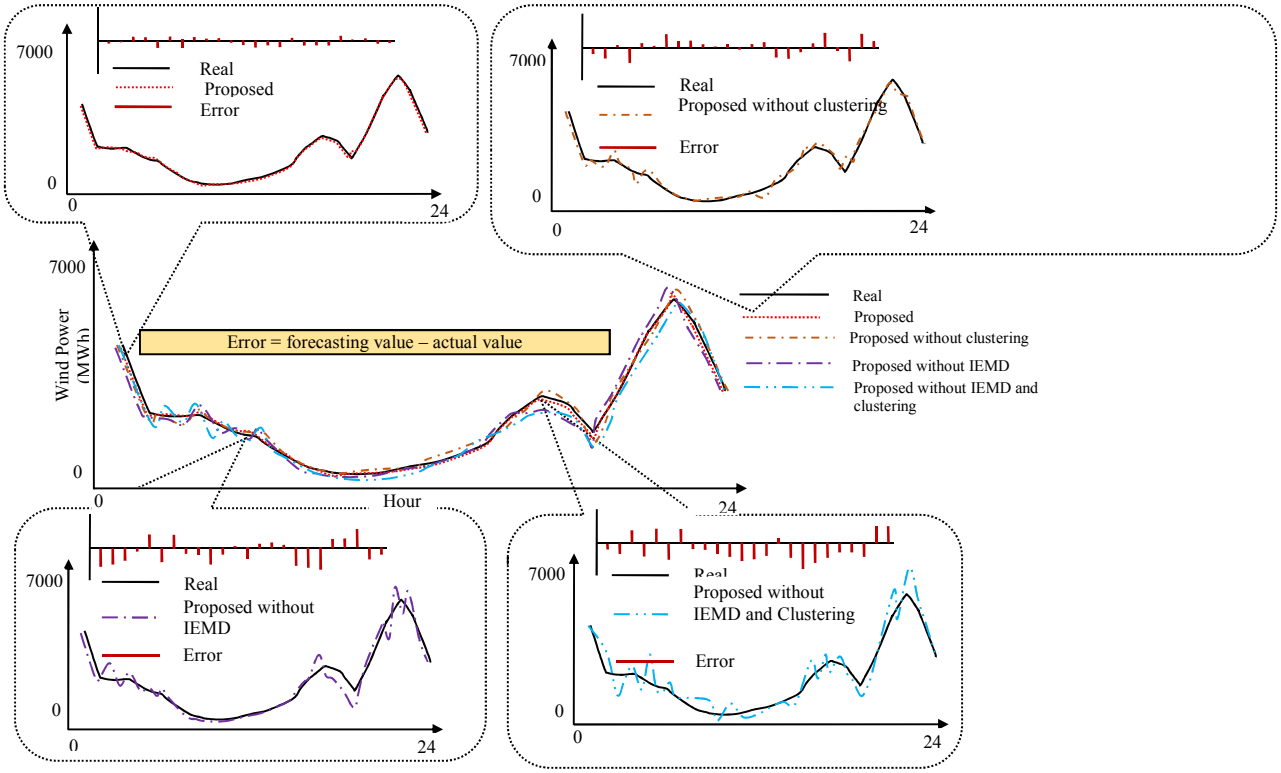

Fig. 6. The forecasting results and actual values for August 18 (Sotavento Wind Farm) for various versions of the proposed model.

TABLE 4
OBTAINED RESULTS FOR WIND POWER FORECAST OF THE SOTAVENTO TEST CASE WITH DIFFERENT TRAINING PERIODS

Test Month	Methods	30 days		40 days		50 days	
		RMSE	MMAPE	RMSE	MMAPE	RMSE	MMAPE
April 2010	MI-IG + NN + CSSO [33]	1.006	22.64	0.486	12.98	0.456	8.12
	Proposed	0.97	20.13	0.39	10.32	0.23	7.12
May 2010	MI-IG + NN + CSSO [33]	1.016	23.14	0.506	12.86	0.406	10.87
	Proposed	0.96	20.12	0.47	10.24	0.38	9.32
June 2010	MI-IG + NN + CSSO [33]	0.836	20.42	0.569	16.48	0.489	15.47
	Proposed	0.72	18.47	0.52	14.72	0.40	14.14
July 2010	MI-IG + NN + CSSO [33]	0.859	17.54	0.468	13.68	0.401	9.14
	Proposed	0.74	16.63	0.44	12.23	0.34	8.70
Average	MI-IG + NN + CSSO [33]	0.929	20.93	0.507	14.00	0.438	10.9
	Proposed	0.85	18.84	0.45	11.88	0.34	9.82

B. Second Case Study: Alberta Wind Farms

The second case study is based on the Alberta (Canada) wind farm datasets which were recorded for the years of 2009 and 2012. For 2009, the dataset from is based on five wind farms located in Pincher Creek, southern Alberta, Canada. Those five farms are called Castler River, Cowley Ridge, Kettles Hill, Summerview, and Summerview-2, with capacities of 44MW, 21MW, 63MW, 70MW, and 66MW, respectively. This leaves the total installed capacity in Pincher Creek at 264MW (P_{max}).

In order to validate and assess the results for Pincher Creek, comparison with [32] was performed, which considered the days of December 3 (Thursday), May 4 (Monday), July 7 (Tuesday), and October 15 (Thursday). The choice of these days was made as to represent the winter, spring, summer, and fall seasons, respectively. The same length of input training data is considered for all models.

The obtained results from the numerical analysis are shown in Table 5. In this table, the suggested model is compared with persistence [34], back propagation NN (BPNN) [34], Radial Basis Function Neural Network (RBFNN) [34], Adaptive Neuro-Fuzzy Inference System (ANFIS) [34], NN based Particle Swarm Optimization (NNPSO) [34], Wavelet Transform (WT) plus BPNN [34], WT plus RBFNN [34], WT plus ANFIS [34], WT plus NNPSO [34], and improved kriging interpolation method plus closed loop NN named IMIK+close-loop NN [30].

The error criteria used for evaluation are MAPE, NRMSE, in addition to the normalized Mean Absolute Error (NMAE), which can be calculated as follows:

$$NMAE = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|X_{ACT(t)} - X_{FOR(t)}\right|}{X_N}\times 100 \qquad (28)$$

The obtained results in Table 5 suggest the superiority of the proposed model as compared to the other models used for comparison. Fig. 7 and 8 provide a graphical representation of the results are presented to display the efficiency of suggested approach.

The comparison of forecasting results MAPE between the proposed method and the other three models which are closest to this method is presented in Fig. 7(a). Fig. 7(b) represents the pie chart comparison of NMAE error criteria for all other models of Table 5.

The results in provide proof for the validity of the proposed method through different error criteria. The improvement of the proposed method in comparison with closest model (i) is; 39% = (8.19-4.99)/8.19 for MAPE, 26%= (4.86-3.61)/4.86 for NMAE and 35% = (6.28-4.09)/6.28 for NRMSE. In Fig. 8, the histogram of NRMSE is presented through comparison with other models.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2020.2976038, IEEE Transactions on Sustainable Energy

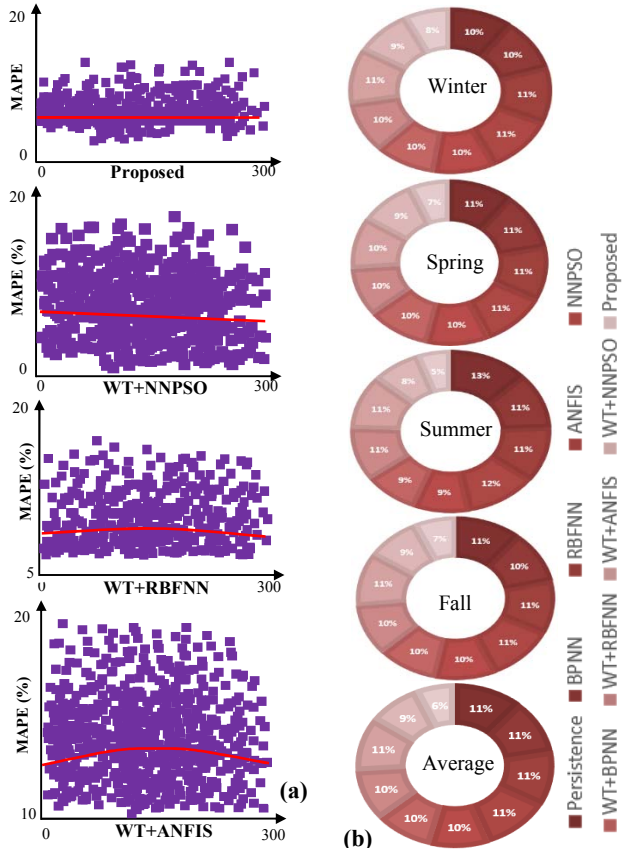

Fig. 7. The inclusive assessment of prediction approach obtained by table 5 (Alberta Case); (a) Comparison of predicted MAPE with three other closest approach, (b) pie diagram comparison based on NMAE for all methods.

TABLE 5
SECOND CASE STUDY: ALBERTA (2009) MAPE, NMAE, AND NRMSE (%).

	Error	a	b	c	d	e	f	g	h	i	j	k	l
Winter	MAPE	10.03	13.62	10.41	14.81	9.54	11.26	8.26	11.08	7.28	7.12	4.18	3.75
	NMAE	4.18	4.32	4.61	4.75	4.09	4.09	4.19	4.55	3.87	4.02	3.23	3.14
	NRMSE	5.41	5.73	5.84	6.13	5.38	5.38	5.43	5.91	5.07	4.86	3.32	2.89
Spring	MAPE	11.31	12.42	11.07	13.51	11.41	11.41	9.22	11.76	8.73	8.50	5.43	4.88
	NMAE	4.58	4.88	4.61	4.71	4.51	4.51	4.18	4.39	4.11	4.12	3.12	2.87
	NRMSE	6.11	6.38	5.89	6.43	6.20	6.20	5.41	6.22	5.83	5.92	4.13	3.78
Summer	MAPE	21.58	17.44	16.72	19.30	12.26	12.26	12.39	16.38	11.27	10.12	7.32	7.63
	NMAE	8.48	7.46	7.32	7.75	5.94	5.94	7.04	7.18	5.29	5.10	4.14	3.67
	NRMSE	11.25	9.23	8.88	10.16	7.33	7.33	8.36	9.63	7.02	6.76	5.44	4.47
Fall	MAPE	14.79	13.93	12.73	12.04	12.82	12.82	14.86	11.08	5.48	5.63	4.07	3.71
	NMAE	7.48	7.26	7.31	7.76	6.85	6.85	7.08	7.32	6.17	6.08	5.21	4.76
	NRMSE	9.19	8.79	8.99	9.38	7.43	7.43	8.60	8.93	7.21	6.43	5.73	5.22
Ave.	MAPE	14.43	14.35	12.73	14.91	11.51	11.94	11.18	12.57	8.19	7.84	**5.25**	4.99
	NMAE	6.18	5.98	5.96	6.24	5.35	5.35	5.62	5.86	4.86	4.83	**3.92**	3.61
	NRMSE	7.99	7.53	7.4	8.02	6.58	6.58	6.95	7.67	6.28	5.99	**4.65**	4.09

a= Persistence [34], b= BPNN [34], c= RBFNN [32] d= ANFIS [34], e= NNPSO [34], f= WT+BPNN [34], g= WT+RBFNN [34], h= WT+ANFIS [34], i= WT+NNPSO [34], j= MI-IG + NN + CSSO [33], k= IKIM+ closed-loop NN [30], l= Proposed

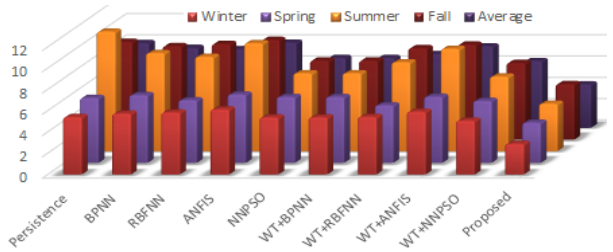

Fig. 8. Comparison for bar diagram of NRMSE with other models.

TABLE 6
SECOND CASE STUDY: ALBERTA (2012) NMAE, AND NRMSE (%).

Methods	Error	March	Jun	Sept.	Dec.	Ave.
Persistence [35]	NRMSE	13.71	15.14	18.44	12.49	14.95
	MAPE	10.08	10.79	13.11	8.84	10.71
RBF [35]	NRMSE	18.32	14.57	18.62	14.11	16.40
	MAPE	13.32	10.45	13.77	10.24	11.95
MLP [35]	NRMSE	15.36	15.62	19.80	12.32	15.78
	MAPE	12.42	11.56	14.54	9.02	11.89
WTNN+MSE [35]	NRMSE	12.38	14.99	17.66	11.65	14.17
	MAPE	9.36	10.64	12.49	8.53	10.26
WTNN+ICSA [35]	NRMSE	12.23	12.48	16.68	11.58	13.24
	MAPE	9.22	9.64	11.73	8.22	9.70
Proposed with ELM	NRMSE	10.65	11.04	15.03	10.12	11.71
	MAPE	7.67	7.98	9.98	7.05	8.17
Proposed	NRMSE	8.73	8.33	12.43	8.24	9.43
	MAPE	6.02	6.34	7.47	6.12	6.49

The results show better performance of the proposed method in terms of forecasting error in comparison with the other models. This is valid for all seasons of the Alberta test case. The third and final analysis was comprised of a one-hour forecasting horizon to validate the suggested approach for very short-term forecasting.

For this, the Alberta dataset was considered for March, June, September and December of 2012, with 861, 941, 941, and 1087MW, respectively.

The obtained results are listed in Table 6 and validated by comparing the forecasting error metrics with against previously published work which used the same data to validate five different algorithms as listed in the table. All the methods the same length of days for the training (59 days) and validation (1 day) sets. The same test weeks were used. [35].

It can clearly be seen that the proposed method provide lower error values (NRMSE and NMAE) for all four test weeks as well as average value of all weeks. This validates the capability of the proposed model also for very short-term forecasting (hour-ahead).

C. Blue Canyon Wind Farm

As the third test case, the Blue Canyon wind farm is considered, which consists of 45 turbines with 74 MW capacity [36]. In this test case, the proposed model is compared with the persistence method [37], Correlation Analysis (CA) + Bayesian Clustering by Dynamics (BCD) + Support Vector Regression (SVR) [37], Modified Hybrid Neural Network (MHNN) + Enhanced Particle Swarm Optimization (EPSO) [38], IKIM+ closed-loop NN [30], and time series model relating the predicted interval [30].

Obtained results are presented in Table 7 with the same test conditions of [37-38 and 30], considering the month of June, 2005. Additionally, various forecasting horizons of 1-h ahead, 24-h ahead, and 48-h ahead have been considered in this table showing better outcomes of the proposed model according to both NMAE and NRMSE error criteria.

TABLE 7
COMPARISON OF PROPOSED METHOD WITH [28-29 AND 21], ON BLUE CANYON WIND FARM THROUGH VARIOUS FORECAST HORIZONS

Test Conditions		Persistence Method [37]	CA+BCD + SVR [37]	MHNN + EPSO [38]	Time series model relating the predicted interval [39]	IKIM+ closed-loop NN [30]	Proposed
Forecast Horizon	Error						
1-h. ahead	NMAE	7.84	6.65	4.12	3.95	3.87	3.12
1-h ahead	NRMSE	11.93	10.54	7.52	5.40	5.43	4.32
24-h ahead	NMAE	21.24	14.38	7.90	6.72	6.64	5.47
24-h ahead	NRMSE	29.84	19.74	12.60	10.14	10.29	9.72
48-h ahead	NMAE	25.42	15.73	10.51	7.98	7.89	7.42
48-h ahead	NRMSE	34.81	21.24	16.58	14.14	14.32	12.14

C. Future Work

Due to critical role of forecasting engines in prediction processes of complex signals associated with renewable power integration (e.g. wind speed/power), the application of improved forecasting engines and more effective feature selection are suggested as future works of this paper. In addition, a probabilistic version of the proposed forecasting model can be considered for future research prospects.

VI. CONCLUSION

This work proposed a new wind power forecasting model based on an improved EMD (IEMD) in a hybrid prediction framework, incorporating K-means clustering, BaNN, and ChB-SSO. The effectiveness of the proposed model was applied on three different test cases (Sotavento, Alberta, and Blue Canyon wind farms) in order to conduct a comprehensive comparative study with numerous previously published models. Furthermore, to prevent the problem of premature convergence during the learning phase, the proposed forecast engine is run with training periods of various lengths and the training period leading to the best forecast results was used. The proposed model was shown to provide accurate results for various forecast horizons through a comparative study with other models with different error evaluation criteria. By evaluating the obtained numerical results, the proposed model was found to outperform all compared studies for all the test cases. In this analysis, the results were compared against 9 published works based on 23 different forecasting approaches. In addition, an individual analysis of each component of the hybrid model was performed. The results of this study can provide useful suggestions for a significant enhancement of wind power prediction.

REFERENCES

[1] Bahrami, S., Amini, M. H., Shafie-Khah, M., & Catalao, J. P. (2018). A decentralized renewable generation management and demand response in power distribution networks. *IEEE Transactions on Sustainable Energy,* 9(4), 1783-1797.

[2] Amini, M. H., Kargarian, A., & Karabasoglu, O. (2016). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. Electric Power Systems Research, 140, 378-390.

[2] H. Bludszuweit, J. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 983–991, Aug. 2008.

[4] T.H.M. El-Fouly, E.F. El-Saadany, M.M.A. Salama, One day ahead prediction of wind speed using annual trends, in: IEEE, Power Engineering Society General Meeting, 2006.

[5] S. Rajagopalan and S. Santoso, "Wind power forecasting and error analysis using the autoregressive moving average modeling," in *Proc. IEEE Power Energy Soc. Gen. Meeting* (*PES'09*), Jul. 2009, pp. 1–6.

[6] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "Arima-based time series model of stochastic wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010.

[7] N. Chen, Z. Qian, I. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 656–665, Mar. 2014.

[8] F. Cassola, M. Burlando, Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, *Applied Energy*, vol. 99, pp. 154-166, 2012.

[9] S. Fan, L. Chen, and W. J. Lee, "Short-term load forecasting using comprehensive combination based on multimeteorological information," *IEEE Trans. Ind. Appl.*, vol. 45, no. 4, pp. 1460–1466, Jul./Aug. 2009.

[10] S. Sun, H. Qiao, Y. Wei, S. Wang, A new dynamic integrated approach for wind speed forecasting, *Applied Energy*, vol.197, pp. 151-162, 2017.

[11] J. Wang, P. Du, T. Niu, W. Yang, A novel hybrid system based on a new proposed algorithm—Multi-Objective Whale Optimization Algorithm for wind speed forecasting, *Applied Energy*, 2017.

[12] C. Feng, M. Cui, B. M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, *Applied Energy*, vol. 190, pp. 1245–1257, 2017.

[13] Z. Shen, M. Ritter, Forecasting volatility of wind power production, *Applied Energy*, vol. 176, pp. 295-308, 2016

[14] J. Zhao, Y. Guo, X. Xiao, J. Wang, D. Chi, Z. Guo, Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method, *Applied Energy*, vol. 197, pp. 183-202, 2017.

[15] H. Akçay, T. Filik, Short-term wind speed forecasting by spectral analysis from long-term observations with missing values, *Applied Energy*, vol. 191, pp. 653-662, 2017.

[16] A. Tascikaraoglu, B. M. Sanandaji, K. Poolla, P. Varaiya, Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform, *Applied Energy*, vol. 165, pp. 735–747, 2016.

[17] Z. Song, Y. Jiang, Z. Zhang, Short-term wind speed forecasting with Markov-switching model, *Applied Energy*, vol. 130, pp. 103-112, 2014.

[18] F. Cassola, M. Burlando, Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, *Applied Energy*, vol. 99, pp. 154-166, 2012.

[19] Du P. Ensemble Machine Learning Based Wind Forecasting to Combine NWP output with Data from Weather Stations. *IEEE Transactions on Sustainable Energy*. 2018 Nov 12.

[20] Lin Y, Yang M, Wan C, Wang J, Song Y. A multi-model combination approach for probabilistic wind power forecasting. *IEEE Transactions on Sustainable Energy*. 2018 May 4;10(1):226-37.

[21] Wang Y, Zhou Z, Botterud A, Zhang K. Optimal wind power uncertainty intervals for electricity market operation. *IEEE Transactions on Sustainable Energy*, 2017 Jul 11;9(1):199-210.

[22] Wang, Y., Hu, Q., Srinivasan, D., & Wang, Z. (2018). Wind power curve modeling and wind power forecasting with inconsistent data. *IEEE Transactions on Sustainable Energy*, 10(1), 16-25.

[23] Wu, J. L., Ji, T. Y., Li, M. S., Wu, P. Z., & Wu, Q. H. (2015). Multistep wind power forecast using mean trend detector and mathematical morphology-based local predictor. *IEEE Transactions on Sustainable Energy*, 6(4), 1216-1223.

[24] Yan J, Li K, Bai E, Zhao X, Xue Y, Foley AM. Analytical Iterative Multistep Interval Forecasts of Wind Generation Based on TLGP. *IEEE Transactions on Sustainable Energy*. 2018 May 29;10(2):625-36.

[25] Yang, L., He, M., Zhang, J., & Vittal, V. (2015). Support-vector-machine-enhanced markov model for short-term wind power forecast. *IEEE Transactions on Sustainable Energy*, 6(3), 791-799.

[26] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *P. Roy. Soc. Lond. A Mat.*, 454(1971) (1998) 903-995.

[27] C. Hyuk-Gyu, H. Park, and H. Kwon, "Similarity measurement among sectors using extended Relief-F algorithm for disk recovery," in *Proc. IEEE ICCIT*, Busan, South Korea, Nov. 2008, pp. 790-795.

[28] O. Abedinia, N. Amjady, A. Ghasemi, A New Meta-heuristic Algorithm Based on Shark Smell Optimization, *Complexity Journal*, DOI: 10.1002/cplx.21634, 2014.

[29] O. Abedinia, D. Raisz, N. Amjady, An Effective Prediction Model for Hungarian Small-Scale Solar Power Output, *IET Renewable Power Generation*, DOI: 10.1049/iet-rpg.2017.0165 , 2017.

[30] N. Amjady, O Abedinia, Short Term Wind Power Prediction Based on Improved Kriging Interpolation, Empirical Mode Decomposition, and Closed-Loop Forecasting Engine, *Sustainability* 9 (11), 2104, 2017.

[31] N. Amjady, F. Keynia, H. Zareipour, "A New Hybrid Iterative Method for Short Term Wind Speed Forecasting," *European Transactions on Electrical Power*, Vol. 21, No. 1, pp. 581-595, January 2011.

[32] N. Amjady, F. Keynia, H. Zareipour, Short-term wind power forecasting using ridgelet neural network, *Electric Power Systems Research,* vol. 81, no. 12, pp. 2099–2107, December 2011.

[33] Abedinia, O.; Amjady, N. Short-Term Wind Power Prediction based on Hybrid Neural Network and Chaotic Shark Smell Optimization. *Int. J. Precis. Eng. Manuf.-Green Technol.* 2015, 2, 245–254.

[34] P. Mandal, H. Zareipour, W. D. Rosehart, Forecasting aggregated wind power production of multiple wind farms using hybrid wavelet-PSO-NNs, *International Journal of Energy Research*, vol. 38, pp. 1654–1666, February 2014.

[35] H. Chitsaz, N. Amjady, H. Zareipour, Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm, *Energy Conversion and Management*, vol. 89, pp. 588–598, 2015.

[36] Oklahoma Wind Farm. Available online: http://bluecanyonwindfarm.com (accessed on 5 July 2015).

[37] S. Fan, James R. Liao, R. Yokoyama, L. Chen, and Wei-Jen Lee, Forecasting the Wind Generation Using a Two-Stage Network Based on Meteorological Information. *IEEE Trans. Energy Convers*. 2009, 24, 474–482.

[38] Amjady, N.; Keynia, F.; Zareipour, H. Wind Power Prediction by a New Forecast Engine Composed of Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization. *IEEE Trans. Sustain. Energy* 2011, 2, 265–276.

[39] T.H.M. El-Fouly, E.F. El-Saadany, M.M.A. Salama, One day ahead prediction of wind speed and direction, in: *IEEE Transactions on Energy Conversion*, 2008, pp. 191-201.

Oveis Abedinia (Member, IEEE) joined Nazarbayev University as a Visiting Professor with different projects in 2019. Furthermore, he is working as a Head of the International Relations of Science and Technology Park (STP), Ardabil, Iran.

Mohamed Lotfi (Member, IEEE) is currently a PhD student at the Faculty of Engineering of the University of Porto (FEUP) and a research assistant at INESC TEC.

Mehdi Bagheri (Member, IEEE) is currently an Associate Professor with the School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan, Kazakhstan. He is also an Associate Editor of IEEE ACCESS.

Behrouz Sobhani is currently an Assistant Professor with the Faculty of Technical Engineering, University of Mohaghegh Ardabili, Ardabil, Iran.

Miadreza Shafie-khah (Senior Member, IEEE) is currently a Tenure-Track Professor with the University of Vaasa, Vaasa, Finland. He is currently an Associate Editor of *IET Renewable Power Generation* and an Editor of the IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY.

João P. S. Catalão (Senior Member, IEEE) is currently a Professor with the Faculty of Engineering, University of Porto (FEUP), Porto, Portugal, and a Research Coordinator with INESC TEC. He is the Promotion and Outreach an Editor of the IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY, an Editor of the IEEE TRANSACTIONS ON SMART GRID and the IEEE TRANSACTIONS ON POWER SYSTEMS, and an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.