Energy Conversion and Management 89 (2015) 588–598

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm

Hamed Chitsaz [a,*], Nima Amjady [b], Hamidreza Zareipour [a]

[a] Schulich School of Engineering, University of Calgary, Alberta, Canada
[b] Electrical Engineering Department, Semnan University, Semnan, Iran

ARTICLE INFO

Article history:
Received 13 June 2014
Accepted 1 October 2014
Available online 29 October 2014

Keywords:
Wind power forecasting
Wavelet neural network
Clonal optimization

ABSTRACT

With the integration of wind farms into electric power grids, an accurate wind power prediction is becoming increasingly important for the operation of these power plants. In this paper, a new forecasting engine for wind power prediction is proposed. The proposed engine has the structure of Wavelet Neural Network (WNN) with the activation functions of the hidden neurons constructed based on multi-dimensional Morlet wavelets. This forecast engine is trained by a new improved Clonal selection algorithm, which optimizes the free parameters of the WNN for wind power prediction. Furthermore, Maximum Correntropy Criterion (MCC) has been utilized instead of Mean Squared Error as the error measure in training phase of the forecasting model. The proposed wind power forecaster is tested with real-world hourly data of system level wind power generation in Alberta, Canada. In order to demonstrate the efficiency of the proposed method, it is compared with several other wind power forecast techniques. The obtained results confirm the validity of the developed approach.

1. Introduction

In recent years, wind power has been the fastest growing renewable electricity generation technology in the world [1,2]. The worldwide wind capacity reached approximately 300 GW by the end of June 2013, out of which 13.9 GW were added in the first six months of 2013 [3]. In particular, Canada installed 377 MW during the first half of 2013, which is 50% more than in the previous period of 2012 [3]. In the Province of Alberta, Canada, in particular, the installed capacity reached 1087 MW in late 2012, and is expected to grow to 2388 MW by 2016 [4]. Despite the environmental benefits of wind power [5], it has an intermittent nature [6], which could affect power systems security [7] and reliability [8].

One approach to deal with wind power intermittency in operation time scale is to forecast it over an extended period of time. Accurate wind power forecasting can improve the economical and technical integration of large capacities of wind energy into the existing electricity grid [9]. Wind forecasts are important for system operators to control balancing, operation and safety of the grid [10]. On the other hand, wind power forecast errors might sometimes require system operators to re-dispatch the system in real time. The costs of re-dispatch affect electricity prices and system performance [11]. Moreover, reserve requirements are connected to wind forecast uncertainty [12]. Hence, reducing the costs of re-dispatch and contribution of spinning reserves by more accurate wind power prediction can effectively increase system operation efficiency. For instance, the economic benefits of accurate wind forecasting were assessed by GE Energy for the New York State Energy Research and Development Authority (NYSERDA) and the NYISO – all terms are defined in Appendix A. In that study, it was estimated that $125 Million, or 36%, of the cost reduction is associated with state-of-the-art wind power forecasting. It was about 80% of the estimated cost reduction that could be achieved with a perfect wind power production forecast [13].

Hence, various approaches have been proposed to improve wind power forecasting accuracy in the literature. In one group of models, wind speed and other climate variables are predicted using Numerical Weather Prediction (NWP) models, and those forecasts are used to predict the wind power output of a wind turbine or a wind farm [14,15] using turbine or farm production curves. In another group of models, the NWP forecasts, or self-generated climate variables forecasts, are fed into secondary time series models to predict wind power output for a turbine, a farm or system level wind power production. The time series models may be built based on ensemble forecasting [16], statistical approaches [17,18], or artificial intelligence techniques [19,20]. In a third group of models,

* Corresponding author. Tel.: +1 587 896 8388; fax: +1 403 282 6855.
E-mail address: hchitsaz@ucalgary.ca (H. Chitsaz).

http://dx.doi.org/10.1016/j.enconman.2014.10.001
0196-8904/

only past power production values are used in univariate models to predict future wind power values [21,22]. Despite improvements in wind power forecasting methods, wind power forecasts still suffer from relatively high errors, ranging from 8% to 22% (in terms of normalized Mean Squared Error) depending on several factors, such as, forecasting horizon, type of forecasting model, size of wind farm and geographic location [23].

The contribution of the present paper is to propose a new forecasting technique for short-term wind power forecasting. In particular, we develop a wind power forecasting engine based on Wavelet Neural Network (WNN) with multi-dimensional Morlet wavelets as the activation functions of the hidden neurons and Maximum Correntropy Criterion as the error measure of the training phase. We propose a stochastic search technique, which is an improved version of Clonal search algorithm, for training the forecasting engine. The significance of the proposed forecasting technique is that the combination of the WNN and the proposed training strategy is capable of capturing highly non-linear patterns in the data and result in improved forecast accuracy. Particularly, high exploitation capability of the proposed training strategy enables it to find more optimal solutions for the optimization problem of WNN training.

The remaining sections of the paper are organized as follows. A brief literature review on wind power forecasting models is provided in Section 2. The architecture of the wind forecasting model is introduced in Section 3.1. The proposed training strategy is then presented in Section 3.2. The results of the proposed wind forecasting method, obtained for the real-world test cases, are compared with the results of several other prediction approaches in Section 4. Section 5 concludes the paper.

2. Literature review

In this section, a literature review of the existing wind power forecasting models is provided. As mentioned in Section 1, the forecasting methods based on time series, either statistical models or artificial intelligence models, use historical wind power data recorded at the wind farms along with the historical data of the exogenous meteorological variables such as wind speed, temperature and humidity, providing that the data is available. Auto-Regressive Moving Average (ARMA) models [24], Auto-Regressive Integrated Moving Average (ARIMA) model [25], and Fractional ARIMA (FARIMA) model [18], have already been applied to wind speed and wind power prediction. Although time series models are simple forecasting methods and can be easily implemented, most of them are linear predictors, while wind power is generally a non-linear function of its input features.

Artificial intelligence techniques, especially Artificial Neural Networks, have been used in several papers to predict wind power generation [26]. Recurrent neural network [27], Radial Basis Function (RBF) neural network [28], and Multi-Layer Perceptron (MLP) neural networks [22], have been proposed for wind power forecasting. Although neural networks can model nonlinear input/output mapping functions, a single neural network with traditional training mechanisms has limited learning capability and may not be able to correctly learn the complex behavior of wind signal. To remedy this problem, combinations of neural networks with each other and with fuzzy inference systems such as Adaptive Neuro-Fuzzy Inference System (ANFIS) [29,30], and Hybrid Iterative Forecast Method (combining MLP neural networks) [31], have also been suggested for wind speed and power prediction. However, such models and especially fuzzy logic models, involve high complexity and a long processing time in the case of many rules [32].

Another approach to tackle the complex behavior of wind power time series is using wavelet transform. In [33], it has been discussed that wavelets can effectively be used for both stationary and non-stationary time series analysis, and that is one of the reasons for the wide and diverse applications of wavelets. Wind speed and power prediction approaches based on wavelet transform, as a preprocessor to decompose wind speed/power time series, and ANFIS [34], Auto Regressive Moving Average (ARMA) [35], Artificial Neural Network (ANN) [36], and Support Vector Regression (SVR) [37], as forecast engines, have been presented. As for SVM-based models, they highly depend on appropriately tuning of parameters and involve complex optimization process [32]. Wavelet can also be applied in a more efficient structure called wavelet neural network, in which wavelet functions are used as the activation functions of the neurons in neural networks. In [38], wavelet has been used in the form of WNN for wind speed prediction and it is trained by extended Kalman filter. Since such a model consists of many scaled and shifted wavelets of the utilized mother wavelet, it requires a powerful training algorithm to efficiently train the model and not to be trapped in local optima while finding the best input/output mapping function of the model.

In [39], a wind power prediction strategy including a Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization (EPSO) has been proposed. In this paper, a developed evolutionary algorithm, i.e., EPSO, is presented to empower the training phase of the utilized neural network, which is generally a combination of three simple MLPs. Taking into consideration the advantages of wavelet transform in the form of WNN and evolutionary algorithms as the training algorithms, we propose a wind power prediction model, which is elaborated in the next section.

3. Methodology

In this section, we provide the details of the proposed forecasting engine and its training strategy. Briefly, the proposed forecasting technique is composed of a WNN structure with Morlet wavelet functions as activation functions in the hidden layer and a new training strategy. These components are described next.

3.1. The developed wavelet neural network

Wavelet transform has been used in some recent research works for wind forecasting, as a preprocessor to decompose wind speed/power time series to a set of sub-series [34,37,35,36]. The future values of the sub-series are predicted by ANFIS [34], SVR [37], ARMA [35] and ANN [36] and then combined by the inverse WT to form the forecast value of wind power/speed. Another approach to utilize wavelet in a forecast process is through constructing wavelet neural network in which a wavelet function is used as the activation function of the hidden neurons of an ANN. For instance, WNNs with Mexican hat and Morlet wavelets, shown in Fig. 1, as the activation function of the hidden neurons have been applied for another application, i.e. price forecast of electricity markets, in [40,36], respectively. Due to the local properties of wavelets and the ability of adapting the wavelet shape according to the training data set instead of adapting the parameters of the fixed shape activation function, WNNs offer higher generalization capability compared to the classical feed forward ANNs [40]. Recently, a WNN using Mexican hat mother wavelet function is proposed for wind speed forecast [38]. However, Morlet wavelet has vanishing mean oscillatory behavior with more diverse oscillations with respect to Mexican hat wavelet, which can be seen from Fig. 1, and so it can better localize high frequency components in frequency domain and various changes in time domain of severely non-smooth time series, e.g. wind power. In [41], it is mentioned

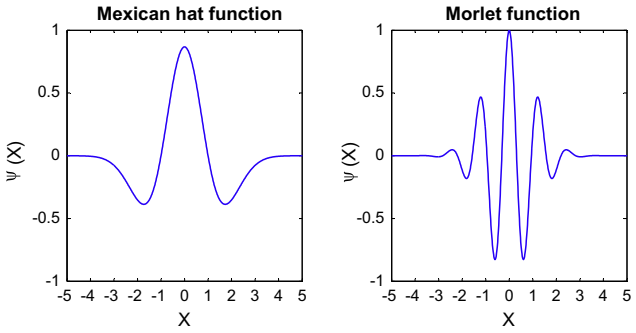

Fig. 1. Mexican hat and Morlet wavelet functions.

that Morlet leads to a better electricity price forecast compared with Mexican hat. In this paper, we propose a WNN with multi-dimensional Morlet wavelet as the activation function of the hidden neurons for wind power forecasting. In the proposed method, we implement a new training algorithm which can efficiently search for the global optimum solution, while a simple gradient method is used as the training algorithm of the WNN in [41]. In addition, since the performance of these two activation functions has not been illustrated in [41], we compare these functions in numerical results in order to demonstrate the effectiveness of Morlet wavelet over Mexican hat in wind power forecast.

Architecture of the WNN is shown in Fig. 2, which is a three-layer feed-forward structure. In Fig. 2, $X = [x_1, x_2, \ldots, x_m]$ is the input vector of the forecast process and y is the target variable. The inputs $x_1, x_2, \ldots, x_m$ of the forecasting engine can be from the past values of the target variable and past and forecast values of the related exogenous variables. For instance, past values of wind power along with the past and forecast values of wind speed, wind direction, temperature and humidity can be considered for wind power prediction, provided that their data is available [20].

A feature selection technique can be used to refine these candidate features and select the most effective inputs for the forecast process. In this research work, we use the feature selection method of [39]. This method is based on the information theoretic criterion of mutual information and selects the most informative inputs for the forecast process by filtering out the irrelevant and redundant candidate features through two stages. In the first stage, so-called irrelevancy filter, mutual information between each candidate input, i.e. $x_i(t)$, and the target variable is computed. The candidate input with higher value of mutual information has more common

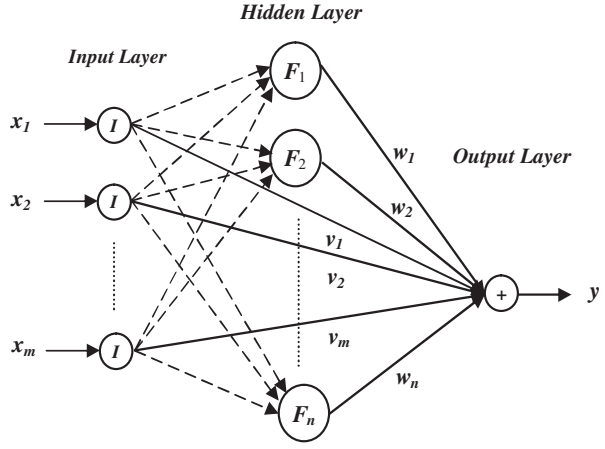

Fig. 2. Architecture of the proposed wind forecasting engine (WNN with Morlet wavelet).

information content with the target variable. The candidate inputs with calculated mutual information value greater than a relevancy threshold TH_1 are considered as the relevant features of the forecast process, which are retained for the next stage, while other candidate inputs with mutual information value lower than TH_1 are considered as irrelevant features, which are filtered out. In the second stage, so-called redundancy filter, redundant features among the selected candidate inputs from the first stage are found and filtered out. Higher value of mutual information between two selected candidates, e.g., $x_k(t)$ and $x_l(t)$, means more common information between these two candidates and thus, they have a higher level of redundancy. Therefore, the redundancy of each selected feature $x_k(t)$ with the other candidate inputs is measured. Afterwards, if the measured redundancy becomes greater than a redundancy threshold TH_2, $x_k(t)$ is considered as a redundant candidate input. Hence, between this candidate and its rival, which has the maximum redundancy with $x_k(t)$, one with lower relevancy should be filtered out [39]. The selected candidate features in the second stage are considered as the inputs of the wind power forecast engine. Moreover, tuning the values of the thresholds TH_1 and TH_2 is performed by cross validation technique. Since this method is not the focus of this paper, it is not further discussed here. The interested reader can refer to [39] for details of this feature selection method. In addition, here, the target variable is wind power of the next time interval that the forecasting engine presents a prediction for it. Multi-period forecast, e.g. prediction of wind power for the next forecast steps, is reached via recursion, i.e. by feeding input variables with the forecaster's outputs. For instance, forecasted wind power for the first hour is used as $y(t-1)$ for wind power prediction of the second hour provided that $y(t-1)$ is among the selected candidate inputs of the feature selection technique.

The forecasting engine should construct the input/output mapping function of $X \Rightarrow y$. The activation function of the input layer nodes is identity function, i.e. $I(x) = x$. In other words, the input layer only propagates the inputs of the WNN to the next layers. The activation function of the hidden layer nodes of the WNN, i.e. multi-dimensional Morlet wavelet, is constructed as follows:

$$F_i(x_1, x_2, \ldots, x_m) = \prod_{j=1}^{m} \psi_{a_i,b_i}(x_j), \quad \forall i = 1, 2, \ldots, n \tag{1}$$

$$\psi_{a_i,b_i}(x_j) = \psi\left(\frac{x_j - b_i}{a_i}\right) \tag{2}$$

where n indicates the number of hidden neurons of WNN, and one-dimensional Morlet wavelet $\psi(\cdot)$ is defined as follow:

$$\psi(x) = e^{-0.5x^2}\cos(5x) \tag{3}$$

In (1) and (2), ψ_{a_i,b_i} is the scaled and shifted version of $\psi(\cdot)$ with a_i and b_i as the scale and shift parameters, respectively. Each activation function $F_i(\cdot)$ has its own a_i and b_i. Based on (1), $F_i(\cdot)$ is m-dimensional wavelet function of $x_1, x_2, \ldots, x_m$ constructed by the tensor product of one-dimensional Morlet wavelets ψ_{a_i,b_i}. Finally, the output of the WNN, denoted by y in Fig. 2, is computed as follows:

$$y = \sum_{i=1}^{n} w_i F_i(x_1, x_2, \ldots, x_m) + \sum_{j=1}^{m} v_j x_j \tag{4}$$

where, w_i is the weight between ith hidden neuron and the output node, and v_j is the direct input weight between jth input and the output node. In other words, the output of the WNN is obtained by a combination of multi-dimensional wavelet functions, i.e. $F_i(x_1, x_2, \ldots, x_m)$, as well as a combination of inputs, i.e. x_j. Thus, the proposed WNN not only can benefit from the capabilities of wavelet functions, such as their ability to capture cyclical behaviors,

but also can capture trends of the signal. Based on the above formulation, the vector of the free parameters of the WNN, denoted by Z, is as follows:

$$Z = [v_1, \ldots, v_m, w_1, \ldots, w_n, a_1, \ldots, a_n, b_1, \ldots, b_n] \quad (5)$$

Therefore, the WNN has $NP = m + 3n$ free parameters, which should be determined by the proposed training strategy.

3.2. The proposed training strategy

We propose new training strategy to train the developed WNN. This strategy is based on improved Clonal selection algorithm. In the following, at first, the Clonal Selection Algorithm (CSA) is briefly introduced. Then, the proposed Improved CSA (ICSA) is presented and adapted as the training strategy to optimize the free parameters of WNN.

As an antigen (e.g., a virus) invades the human body, the biological immunity system selects the antibodies that can effectively recognize and destroy the antigen. The selection mechanism of the immunity system operates based on the affinity of the antibodies with relation to the invading antigen. CSA is an efficient optimization method, inspired by the biological immunity system selection mechanism, proposed by De Castro and Van Zuben [42]. This method has successfully been applied to optimization and pattern recognition domains [42,43] and also unit commitment of power systems [44] in recent years. The performance of CSA can be summarized as the following step by step procedure [42]:

Step 1 : Randomly produce the initial population of CSA within the allowable ranges. Each individual of the population, so called antibody in CSA, is a candidate solution for the optimization problem including its decision variables, called genes in CSA [42]. Here, each antibody of the population of CSA includes NP free parameters of WNN, shown in (5). The number of antibodies in the population is denoted by N. The generation number g is set to zero ($g = 0$). In general, antibody k in generation g is as follows:

$$Z_k^g = [Z_{k,1}^g, Z_{k,2}^g, \ldots, Z_{k,NP}^g], \quad \forall k = 1, 2, \ldots, N \quad (6)$$

where the lth gene or decision variable $Z_{k,l}^g (l = 1, \ldots, NP)$ can be from $v_1, \ldots, v_m$ or $w_1, \ldots, w_n$ or $a_1, \ldots, a_n$ or $b_1, \ldots, b_n$ as shown in (5). The four sets of v_j, w_i, a_i and b_i of each individual are randomly initialized with uniform distribution in the intervals $[-1,1]$, $[-1,1]$, $[0.5,2]$ and $[-3,3]$, respectively. It is noted that the interval of $[-1\ 1]$ is the most common range for weight and bias initialization of neural networks [45]. According to Fig. 1, the output of Morlet wavelet function becomes very close to zero for the input values bigger than 3 and smaller than -3 and therefore, the initial values for b_i is set in the interval of $[-3\ 3]$. Finally, with regard to the initialization of a_i, the interval of $[0.5\ 2]$ is chosen as it is not an extreme interval to make the function too spread/dense.

Step 2 : Determine the affinity of the antibodies with respect to the antigen. In the optimization problems, usually there is no explicit antigen population to be recognized, but an objective function to be optimized (maximized or minimized). Thus, in optimization tasks, an antibody affinity corresponds to the evaluation of the objective function for the given antibody [44]. Here, the proposed WNN is trained, i.e. its free parameters are optimized, by CSA. Thus, training error of the WNN is considered as the objective function of CSA, which should be minimized. Since training error of ANN-based forecasting engines is widely measured in terms of Mean Squared Error (MSE), it is considered as the measure of WNN training error for the present.

Step 3 : Sort antibodies based on their training error values in terms of MSE (i.e., the objective function values) such that the best antibody with the lowest MSE ranks the first.

Step 4 : Copy the selected antibodies based on their position in the sorted population:

$$nc_k = Round\left(\frac{\beta N}{k}\right), \quad \forall k = 1, 2, \ldots, N \quad (7)$$

where nc_k is the number of antibodies copied from kth antibody; $Round(\cdot)$ function rounds up/down its real argument to the nearest integer value; β is a constant coefficient which indicates rate of copy. Thus, an antibody with lower MSE and higher rank (lower k) will be copied more than those with higher MSE. At the end of this step, the number of copied antibodies will be NC as follows:

$$NC = \sum_{k=1}^{N} Round\left(\frac{\beta N}{k}\right) \quad (8)$$

Performance of this step is graphically shown in Fig. 3. As seen, the first antibody with the highest rank is copied nc_1 times, the second one nc_2 times and so on.

Step 5 : Mutate the NC antibodies, produced in Step 4, using the hypermutation operator [43].

Step 6 : Determine MSE value for each mutated antibody. Among the NC mutated and N original antibodies, select NS antibodies ($NS < N$) with the lowest MSE values. These NS antibodies enter directly to the next generation.

Step 7 : Randomly generate $N - NS$ new antibodies for the next generation. These randomly generated antibodies enhance search diversity of CSA, and consequently, the algorithm takes the chance to escape from the local optima.

Step 8 : Increment the generation number ($g = g + 1$). If the termination criterion, such as maximum number of generations, is satisfied, the algorithm is terminated and the best antibody of the last generation, owning the lowest MSE, is determined as the final solution of CSA; otherwise, go back to Step 2 and repeat this cycle. The termination criterion used for the training of the WNN will be described later.

In the above algorithm, N, NS, and β are user-defined settings of CSA. In this algorithm, antibodies are evolved through the mutation, which is the key operator of CSA. The proposed Improved CSA (ICSA) includes two enhancements for this operator as follows. De Castro and Van Zuben [42] discuss that the mutation rate should be inversely proportional to the antigenic affinity: the higher the affinity, the smaller the mutation rate. Thus, in an optimization problem, more optimal candidate solutions should be less mutated. This general idea is modeled in the proposed ICSA based on the following relation:

$$N_k^{Mut} = Round[e^{(-\rho \frac{MSE_{min}}{MSE_k})} \times NP], \quad \forall k = 1, \ldots, N \quad (9)$$

where MSE_k is MSE of kth antibody and MSE_{min} is the lowest MSE among the antibodies of the current population; N_k^{Mut} represents number of genes of the antibodies copied from kth antibody, produced in Step 4, that should be mutated ($N_k^{Mut} \leqslant NP$); the coefficient ρ controls the mutation rate such that higher ρ leads to lower values of N_k^{Mut}. Thus, fewer decision variables are mutated in antibodies with lower MSE (more optimal candidate solutions) and more decision variables are mutated in antibodies with higher MSE (less optimal candidate solutions), and consequently, the individuals of ICSA population are mutated in a coordinated manner. As a result, not only ICSA does allow antibodies with lower MSE to be copied more

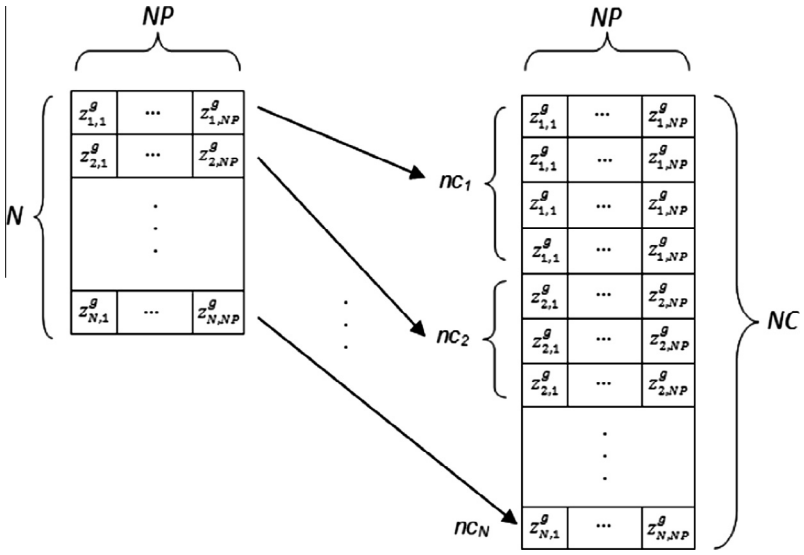

Fig. 3. Representation of Step 4 (copy operator) of CSA.

than those with higher MSE, but also controls the mutation process of copied antibodies in accordance with their MSE using (9). Note that although the mutation operation is applied to the *NC* copied antibodies as shown in Step 5, only *N* values of N_k^{Mut} should be computed, since the antibodies copied from one individual have the same N_k^{Mut}.

After determining N_k^{Mut} for the antibodies of ICSA, the set of genes that should be mutated are randomly selected among the decision variables of each antibody. The hypermutation operator of Step 5 adds a normal random variable with zero mean and constant variance to each decision variable of the mutating antibody [39]. Here, a more effective mutation operation inspired from Differential Evolution (DE) algorithm [46] is proposed for ICSA as follows:

$$Z_{k,l}^{g+1} = \left[1 - e^{-\rho\frac{MSE_{min}}{MSE_k}}\right] Z_{k,l}^{g} + e^{-\rho\frac{MSE_{min}}{MSE_k}}\left(Z_{k1,l}^{g} - Z_{k2,l}^{g}\right), \quad 1 \leqslant k \neq k_1 \neq k_2 \leqslant NC, 1 \leqslant l \leqslant NP \tag{10}$$

where $Z_{k,l}^{g}$ and $Z_{k,l}^{g+1}$ represent gene l of antibody k in two successive generations g and $g+1$. As seen, to mutate kth antibody, this mutation operator uses the values of decision variables in two other antibodies (i.e. $Z_{k_1,l}^{g}$ and $Z_{k_2,l}^{g}$) as it is applied in DE mutation operator.

In [46], it has been discussed that DE mutation operator by computation of difference between two randomly chosen individuals from the population (here, antibodies k_1 and k_2), determines a function gradient in a given area (not in a single point), and therefore, prevents trapping the solution in a local optimum of the objective function. Moreover, to enhance the search diversity of the proposed mutation operation, it is separately applied to each gene of the mutating antibodies. Additionally, the $exp(\cdot)$ term of (9) and its complement $1 - exp(\cdot)$ are used in (10). If kth antibody $Z_{k,l}^{g}$ is a good candidate solution, the $exp(\cdot)$ term and its complement become close to 0 and 1, respectively. Thus, the next generation decision variables mainly take their values from the current generation decision variables and small gradient terms are added to them. Consequently, ICSA can search promising areas of the solution space with small steps or high resolution, which is known as exploitation capability. Based on this capability, a stochastic search technique can extract optimal solutions of the solution space from its promising areas. High exploitation capability of the proposed ICSA allows it to find more optimal solutions for the optimization problem of WNN training. In other words, the WNN using ICSA can better learn the complex input/output mapping function of the wind power forecast process and so predict its future values with higher accuracy. However, if kth antibody is a poor candidate solution, the $exp(\cdot)$ term and its complement become close to 1 and 0, respectively, and so the next generation decision variables mainly obtain their values from large gradient terms. Hence, ICSA can move out from the non-promising areas of the solution space by large steps. Finally, note that due to the effect of the mutation operation of (10), even the copied antibodies of the same antibody lead to different individuals after the mutation operation. Thus, at the end of Step 5 of the proposed ICSA, *NC* diverse candidate solutions are added to *N* original antibodies (Fig. 3), which further enhance the search ability of ICSA.

The termination criterion is an important aspect of the proposed training strategy, as it can affect the performance of the proposed forecasting engine. A low number of ICSA generations may lead to insufficient training of the forecasting engine and cause that the WNN cannot correctly learn the input/output mapping function of the forecast process, i.e. $X \Rightarrow y$. On the other hand, a large number of ICSA generations increases the computation burden and more importantly may lead to over-fitting problem of the WNN. When over-fitting occurs in a neural network based forecast engine, it memorizes the training samples instead of learning them, and thus, while the neural network obtains very low training error, its generalization capability (i.e. its ability to reply to unseen forecast samples) degrades. To avoid these problems, a termination criterion based on cross-validation technique is used for the training phase of the proposed forecasting engine. In this technique, the whole gathered historical data is divided to training and validation samples. The WNN is trained by the proposed ICSA trying to minimize MSE of training samples in each generation. However, at the end of each generation (Step 8), the WNN is tested on the unseen validation samples. For instance, validation samples can be some samples at the end of the historical data interval, i.e. the closest historical data to the forecasting horizon. When MSE of the unseen validation samples, as a measure of the WNN's forecast error, begins to rise, the generalization capability of the WNN begins to degrade, and therefore, the training phase should be terminated. Then, the best individual of the WNN in the generation leading to the minimum validation error, which is expected to yield the maximum generalization capability of the WNN, is selected as the final solution of the training phase indicating the free parameters of the WNN.

Although MSE has widely been used in forecasting models as a training error measure, its applicability to train a neural network is optimal only if the probability distribution of the prediction errors is Gaussian. However, wind power forecast error presents a non-Gaussian shape [47]. Minimizing the squared error is equivalent to minimizing the variance of the error distribution. Accordingly, the higher moments (e.g., skewness, kurtosis, etc.) are not captured, but they contain information that should be passed to the free parameters of the neural network instead of remaining in the error distribution. Ricardo Bessa et al. in [47] proposed some new training error criteria based on minimizing the information content of the error distribution (instead of minimizing the variance in MSE). Maximum Correntropy Criterion is defined as follow:

$$\mathrm{MCC} = max\left\{\frac{1}{N_e}\sum_{i=1}^{N_e} G(\epsilon_i, \sigma^2)\right\} \tag{11}$$

where G is the Gaussian kernel, N_e is the number of training samples, ϵ_i is the error for ith training sample, and σ^2 is the variance. Therefore, MCC approximates a non-Gaussian shape by summation of Gaussian functions corresponding the errors of training samples. For more detailed information relating to MCC, refer to [47]. It should be noted that MCC is a maximization problem, while the proposed training algorithm in Section 3.2, i.e. ICSA, was described as a minimization problem since it was based on MSE. Hence, to adapt this criterion to the proposed training algorithm, minimization of $\frac{1}{\mathrm{MCC}}$ or $-$MCC can be considered instead of maximization of MCC. It is noted that the value of correntropy is always positive [48]. Moreover, the values of MSE_k and MSE_{min} are simply replaced by $(\frac{1}{MCC})_k$ and $(\frac{1}{MCC})_{min}$ (or $(-MCC)_k$ and $(-MCC)_{min}$), respectively, in Eqs. (9) and (10). The main reason for including MSE in the structure of the proposed algorithm is because of the popularity and common use of MSE criterion in training phase of forecasting models. Moreover, the minimization of errors for training samples is often the objective function of forecasting problems, and thus, it is more tangible to deal with a minimization problem in this area.

4. Numerical results

In order to generate forecasts, two parameters must be decided first, namely, forecast interval and forecast horizon. Forecast interval determines the length of each time step into the future (e.g., 10 min or one hour), and forecast horizon determines how many time steps into the future are of interests. Both factors depend on the application of the forecasts. For instance, if the forecasts are used for very short-term adjustments in operation schedules, the forecast horizon could be as short as a few hours. Furthermore, the forecast interval may also vary depending on the application. For example, while a unit commitment algorithm may consider hourly intervals, an economic dispatch algorithm may look into shorter intervals (e.g., 5 min). Note that in generating the forecasts, selecting the forecast interval is sometimes limited by the availability of the data. For example, in Alberta, the meteorological towers that measure weather factors at wind farm sites are set to collect the data for every 10-min interval. Thus, the forecast interval cannot be less than 10 min if the meteorological data are of interests in the models.

In this paper, we have selected to generate hourly wind power production forecasts for Alberta's power system for up to 6 h into the future as our test case. Alberta's electricity market is a real-time market with an hourly settlement interval. Although the system marginal price is determined every minute, the supply and demand offers must be submitted to the system operator for hourly intervals. The bids and offers may be changed up to two hours before the operation hour, which is refereed to as the T-2 window. The majority of slower generators do not strategically adjust their prices and bid at \$0/MWh. Faster units, on the other hand, actively watch the supply–demand balance in the market and adjust their strategies. These units normally act based on the developments in the market in the short-term, usually the next few hours. In particular, for any given operation hour, forecasts of wind power generated before the start of the T-2 Window, when the market participants can still change their bids and offers, is of important value. In addition, given the real-time nature of the market, and considering the fact that only short lead-time units behave strategically, the system operator is mainly concerned with the supply–demand balance for the next hour or two. Thus, while forecasts into longer horizons are important, the ones for the short forecast horizons are particularly valuable for the system operator and market participants in Alberta.

The forecasts of meteorological variables, e.g., wind speed, through NWP models are known to be useful in wind power forecasts for longer forecast horizons [28]. Thus, we choose not to include such data into our model since we are focusing on short-term prediction, and generate forecasts solely based on past power production values. More specifically, 100 hourly lagged values of wind power are considered as the candidate inputs, which are processed by the feature selection technique to select a minimum subset of the most informative features for the proposed forecasting engine. Furthermore, 60 days prior to each forecast day are considered as the historical data divided to 59 days as the training set and one day before the forecast day as the validation set. The second week of March, June, September, and December of year 2012 are considered as test weeks. For each test week, the forecasts are updated in two different ways in this paper. In the first part of the numerical results, we update the forecasts every 6 h, and thus, 28 sets of forecasts are generated for each of the representative weeks. In this part, the error measures are evaluated based on the average of errors for the individual forecasts over the entire week. In the second part of the results, we test the model by updating the forecasts every hour, i.e., for every test hour, there will be six versions of forecasts produced at previous hours. For these forecasts, we evaluate the error measures for each forecast horizon, as further discussed later in this section.

To show an example for selection of inputs using the feature selection technique, the selected features for the third forecasting window of September test week, i.e. the third 6-h of September 8, 2012, are as follows:

$$\begin{aligned} X = [x_1, x_2, \ldots, x_m] = [&WP(t-1), WP(t-2), WP(t-3), WP(t-4),\\ &WP(t-5), WP(t-6), WP(t-7), WP(t-8), WP(t-10),\\ &WP(t-11), WP(t-18), WP(t-19), WP(t-20),\\ &WP(t-21), WP(t-22), WP(t-23), WP(t-24),\\ &WP(t-25), WP(t-26), WP(t-28)] \end{aligned}$$

These 20 features are selected from 100 candidate inputs $WP(t-1), WP(t-2), \ldots, WP(t-100)$.

Two error criteria are used in this paper to evaluate forecast errors: normalized Root Mean Square Error (nRMSE) and Normalized Mean Absolute Error (nMAE) defined as follows:

$$\mathrm{nRMSE} = \sqrt{\frac{1}{\mathrm{NH}}\sum_{t=1}^{\mathrm{NH}}\left(\frac{\mathrm{WP_{ACT(t)}} - \mathrm{WP_{FOR(t)}}}{\mathrm{WP_{Cap}}}\right)^2} \times 100 \tag{12}$$

$$\mathrm{nMAE} = \frac{1}{\mathrm{NH}}\sum_{t=1}^{\mathrm{NH}}\left|\frac{\mathrm{WP_{ACT(t)}} - \mathrm{WP_{FOR(t)}}}{\mathrm{WP_{Cap}}}\right| \times 100 \tag{13}$$

where $\mathrm{WP_{ACT(t)}}$ and $\mathrm{WP_{FOR(t)}}$ indicate the actual and forecast values of wind power for hour t. Also, NH indicates number of hours, which is 168 for test weeks, and $\mathrm{WP_{Cap}}$ is the total wind power capacity of aggregated wind farms, which are 861, 941, 941 and 1087 MW in

March, June, September and December of year 2012, respectively, due to growth of wind power capacity in Alberta.

4.1. Numerical results with 6-h updates

For these forecasts, at the end of each 6-h window, when the wind power values of 6 h become available, the historical data is updated to perform the wind power prediction of the next 6 h. Thus, each forecast horizon includes 6 forecast steps. However, one week or 168 h are considered as the evaluation period for the error criteria of nRMSE and nMAE to better evaluate the performance of the method over a longer period.

The results obtained from the proposed forecasting method, i.e. WNN with Morlet wavelet function, ICSA training algorithm and MCC training criterion, in comparison with the same model but consists of MSE training criterion instead of MCC are shown in Table 1. We have also generated the forecasts based on some other popular models, i.e., the persistence method, and RBF and MLP neural networks. A brief description of these models is presented in Appendix B. For the sake of a fair comparison, all of these methods have the same historical data except the persistence method that does not require training samples. Observe from Table 1 that WNN with MSE as the training measure outperforms the three other forecast methods. For instance, the average nRMSE and average nMAE of WNN with MSE are (14.95–14.17)/14.95 = 5.2% and (10.71–10.26)/10.71 = 4.2% lower than those of the persistence method.

Moreover, as seen from the results of Table 1, considering MCC as the training error measure can significantly improve the forecasting accuracy of wind power prediction. For instance, the average nRMSE and average nMAE for the proposed method are respectively 6.6% and 5.5% lower than those for WNN with MSE, and 11.4% and 9.4% lower than those for the persistence method, which clearly show the advantage of using MCC error measure in training phase of a wind power forecasting model. Note that while persistent forecasts are a useful benchmark, especially for one-step-ahead forecasts, they do not provide any information on variations and ramps in a multi-step-ahead forecasting practice. Thus, despite their average errors being relatively close to the proposed method, they do not contain ramping information, and thus, less useful.

Fig. 4 graphically shows the forecast results of different forecasting models for June 10, 2012, in which there is a sharp downward ramp, and June 12, 2012, in which there is an upward ramp. Observe from this figure that the proposed method, shown in red color, can satisfactorily follow the trend and ramps of the measured wind power, shown in dotted black color, for both days. Neural network based models, e.g., MLP and RBF shown in marked blue color, can also follow ramps of wind power to some extent although they might miss the correct direction sometimes. However, persistence model shown in marked green color, cannot provide any ramp information or good forecast values, especially when there is a ramp in the time series; hence, this model can only be useful for very short term prediction, e.g., up to one hour-ahead.

To illustrate the effectiveness of Morlet wavelet function as the activation function of WNN, the proposed method, i.e., WNN with Morlet wavelet function, ICSA training algorithm and MCC training criterion, is compared with the same model but with the Mexican hat wavelet function (i.e., WNN with Mexican hat wavelet function, ICSA training algorithm and MCC training criterion) in Table 2. Improved performance of the proposed method can be seen from Table 2 such that wind power forecast accuracy of the proposed method is better than the WNN with Mexican hat wavelet in all test weeks.

In the next numerical experiment, the effectiveness of the proposed training strategy is evaluated. For this purpose, the proposed ICSA is replaced with several other well-known stochastic search techniques including Simulated Annealing (SA), Particle Swarm Optimization (PSO), Differential Evolution (DE), and CSA, while the other parts of the suggested wind power forecasting engine, i.e. WNN with Morlet wavelet function and MCC training criterion, are kept unchanged. As seen from Table 3, the proposed ICSA leads to the lowest wind forecasting errors in all test weeks among all stochastic search techniques including SA, PSO, DE and CSA by 20.7%, 25.1%, 20.8% and 9.8% improvement of average nRMSE, and by 22.4%, 25.0%, 22.5% and 9.4% improvement of average nMAE, respectively, demonstrating the effectiveness of the proposed training strategy.

Finally, we applied the proposed method to predict wind power for the year 2012 so as to have a comprehensive evaluation of its performance. Hence, 10 months from March to December 2012 have been considered in the last numerical experiment and monthly errors are reported in Table 4. It is noted that the data for the first two months is used as training samples for prediction of month March, and therefore, no prediction result can be reported for January and February. According to this table, monthly errors for months March and December are very close to those earlier presented for the associated test weeks in these months. For instance, monthly nRMSE and nMAE are respectively 11.52% and 7.80% for the month December, and weekly nRMSE and nMAE are 11.58% and 8.22%, respectively, for the test week of December. Errors related to September test month shown in Table 4 are considerably lower than those for the test week of September presented in previous tables. The reason is that there are more sever ramps in the second week of September selected as the test week compared with other weeks in this month, and therefore, the average error of the month is lower than the error associated with the second week of this month. On the contrary, forecasting errors

Table 1
nRMSE (%) and nMAE (%) results for the four test weeks of year 2012.

Model	Error	Test week				Average
		Mar.	Jun.	Sep.	Dec.	
Persistence	nRMSE	13.71	15.14	18.44	12.49	**14.95**
	nMAE	10.08	10.79	13.11	8.84	**10.71**
RBF	nRMSE	18.32	14.57	18.62	14.11	**16.40**
	nMAE	13.32	10.45	13.77	10.24	**11.95**
MLP	nRMSE	15.36	15.62	19.80	12.32	**15.78**
	nMAE	12.42	11.56	14.54	9.02	**11.89**
WNN with MSE	nRMSE	12.38	14.99	17.66	11.65	**14.17**
	nMAE	9.36	10.64	12.49	8.53	**10.26**
Proposed method	nRMSE	12.23	12.48	16.68	11.58	**13.24**
	nMAE	9.22	9.64	11.73	8.22	**9.70**

The average of errors are in bold.

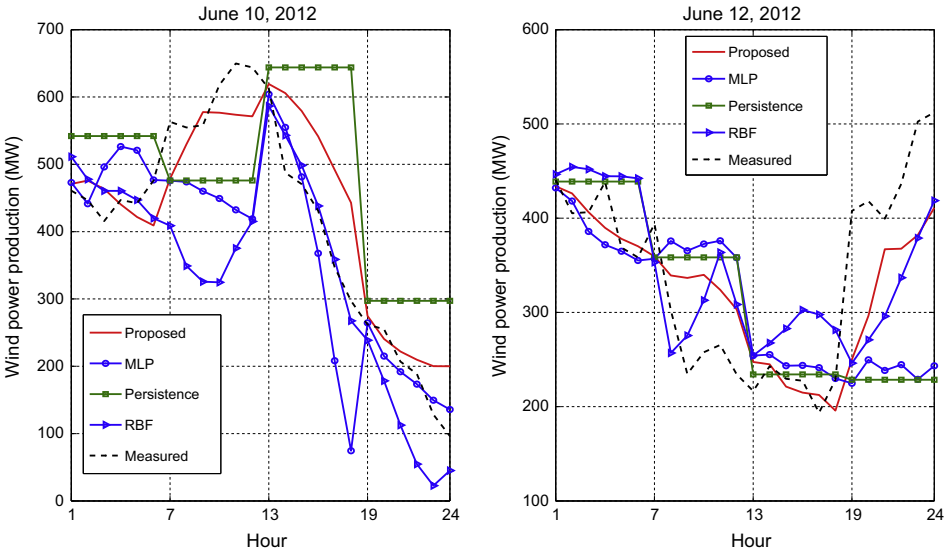

Fig. 4. Forecast results of different models for two different days.

Table 2
Comparison of the proposed method and a WNN with Mexican hat wavelet.

Model	Error	Test week				Average
		Mar.	Jun.	Sep.	Dec.	
WNN with Mexican hat	nRMSE	13.77	13.23	17.12	12.22	**14.08**
	nMAE	10.13	9.80	12.54	8.71	**10.30**
Proposed method	nRMSE	12.23	12.48	16.68	11.58	**13.24**
	nMAE	9.22	9.64	11.73	8.22	**9.70**

The average of errors are in bold.

Table 3
Comparison of the proposed training strategy, i.e. ICSA, with SA, PSO, DE and CSA.

Algorithm	Error	Test week				Average
		Mar.	Jun.	Sep.	Dec.	
SA	nRMSE	19.95	16.23	16.96	13.70	**16.71**
	nMAE	14.73	11.57	13.53	10.18	**12.50**
PSO	nRMSE	18.33	17.69	20.70	14.01	**17.68**
	nMAE	14.18	13.38	14.37	9.88	**12.95**
DE	nRMSE	17.51	17.43	17.94	13.97	**16.72**
	nMAE	13.39	13.34	13.08	10.25	**12.52**
CSA	nRMSE	13.79	13.47	18.29	13.22	**14.69**
	nMAE	10.49	10.29	12.80	9.21	**10.70**
Proposed method (ICSA)	nRMSE	12.23	12.48	16.68	11.58	**13.24**
	nMAE	9.22	9.64	11.73	8.22	**9.70**

The average of errors are in bold.

related to test month of June presented in Table 4 are higher than those earlier presented for the test week of June. Here, wind power data for the second week of June has been more predictable than other weeks in this month. Moreover, the average errors presented in Table 4, i.e., 12.10% and 8.41% in terms of nRMSE and nMAE, respectively, are close and even lower than ones presented for the four test weeks, i.e., 13.24% and 9.70%. As a result, it validates the comparative results presented in Tables 1–3 between the proposed method and other methods based on the test weeks' consideration. Furthermore, considering forecasting errors presented in Tables 1–3, no matter which test week is considered (e.g., with high or low predictability), the proposed method demonstrated its higher wind power forecast accuracy compared with other benchmark models.

Table 4
Wind power prediction error results of the proposed method for 10 months.

Test month	Error	
	nRMSE	nMAE
Mar.	11.89	8.32
Apr.	11.98	8.46
May	12.32	9.26
Jun.	13.69	9.74
Jul.	10.71	7.29
Aug.	12.08	8.05
Sep.	13.26	8.78
Oct.	11.35	7.78
Nov.	12.21	8.64
Dec.	11.52	7.80
Average	**12.10**	**8.41**

The average of errors are in bold.

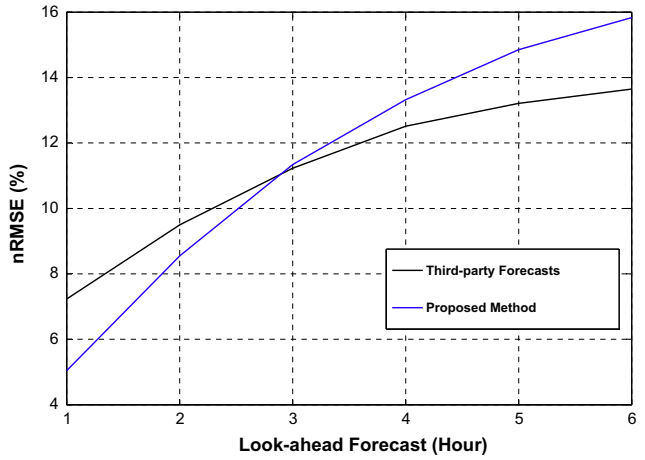

Fig. 5. Average nRMSE errors of two forecasting models in different look ahead forecasts.

4.2. Numerical results with hourly updates

Prediction errors can also be calculated for each look-ahead forecast distinctly (i.e., each forecast hour). In some of the mainly academic literature, the average of prediction errors for 1-h ahead up to the last hour of the forecast horizon is calculated. For instance, in 6-h ahead prediction, each forecasting window includes 1 to 6-h ahead forecasts, and the error is the average of the prediction errors of these 6 forecast values, as performed in Section 4.1. However, in most commercial forecasting tools, the forecasts are updated after each forecast interval, and forecast accuracy is evaluated for each look-ahead interval individually. In other words, the values of wind power for the first 6 h, i.e., $WP(t+1),\ldots,WP(t+6)$, are predicted at time t. When the observed value of wind power for the hour $t+1$ becomes available, the data is updated and forecasts for the next 6 h, i.e., $WP(t+2),\ldots,WP(t+7)$, are predicted at time $t+1$.

In other words, the data is updated every hour, while the forecast horizon is still 6-h-ahead. In the following numerical experiment, the proposed method is applied for aggregated wind power forecast of Alberta in 10 test weeks, including the second weeks of March to December 2012, such that the prediction accuracy of different look-ahead forecasts within the 6-h window is evaluated. Fig. 5 demonstrates the results of this numerical experiment including the curves of average nRMSE for the six look-ahead forecasts, obtained from the proposed approach. We also present the same measures of accuracy for the third-party forecasts currently employed by the Alberta Electric Systems Operator [4]. Observe from this figure that forecast accuracy for both models decreases as the look ahead forecast increases due to higher cumulative errors. For instance, the average forecast error of the proposed method for 1-h-ahead forecast, i.e. forecast values generated 1 h before the real time, is 5.04%, while for 6-h-ahead, i.e. forecast values generated 6 h before the real time, is 15.94%.

Compared to the third-party forecasts, the proposed method results in better or comparable forecast accuracy up to 3 h ahead. For instance, the proposed model improves the accuracy for 1-h-ahead forecast by 30.29%. Note that at any settlement interval, the next three hours are particularly important to the system operator in Alberta. This is because; given the real-time nature of the market, the next hour or two need to be monitored properly to ensure system security and supply–demand balance. Strategic market participants also need to watch the next three hours because it is just outside the T-2 Window. For the three longer look-ahead windows, the third party forecasts outperform the ones generated by our model. This is mainly because the third party forecasts include NWP data. Thus, a combination of the forecasts generated by our model for the first three hours and the third-party forecasts for the last three hours would provide a more accurate picture of future wind generation.

5. Conclusion

In this paper, a new wind power prediction strategy is proposed. A WNN with multi-dimensional Morlet wavelet as the activation function of the hidden neurons and MCC as the training criterion is applied as the forecasting engine to implement the input/output mapping function of wind prediction process. A new stochastic search technique, named ICSA, which is the improved version of Clonal selection algorithm, is proposed and adapted as the training procedure to optimize the free parameters of the forecasting engine. Effectiveness of the whole proposed wind power forecasting strategy as well as the effectiveness of its main components including the suggested WNN and training procedure is extensively evaluated by real-world data for wind power prediction. As regards the training algorithm, the proposed ICSA outperforms the other stochastic search algorithms including SA, PSO, DE and CSA in terms of both nRMSE and nMAE, illustrating the effectiveness of the proposed training strategy. Moreover, the suggested Morlet wavelet function results in more accurate wind power forecast than Mexican-hat wavelet function, and MCC training criterion leads to lower wind power prediction errors than the traditional training error measure of MSE.

Appendix A. Terms and definitions

Here, all terms mentioned in this paper and their definitions are listed in alphabetical order:

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ARMA	Auto Regressive Moving Average
CSA	Clonal Selection Algorithm
DE	Differential Evolution
GE	General Electric
ICSA	Improved Clonal Selection Algorithm
MCC	Maximum Correntropy Criterion
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
nMAE	Normalized Mean Absolute Error

nRMSE	Normalized Root Mean Squared Error
NYISO	New York Independent System Operator
NYSERDA	New York State Energy Research and Development Authority
NWP	Numerical Weather Prediction
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
SA	Simulated Annealing
SVR	Support Vector Regression
SVM	Support Vector Machine
WNN	Wavelet Neural Network

Appendix B. Benchmark models

Here, three forecasting methods used as benchmarks in Table 1, i.e., Persistence, MLP and RBF, are briefly described.

Persistence is the simplest forecasting model as it assumes the forecast value at a certain time in the future is the same as the last measured value. Therefore, this naive method is useful for very short-term prediction purposes, while its forecast error remarkably increases as the forecast horizon increases.

Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) are famous Artificial Neural Networks (ANNs), which have successfully been applied to forecasting problems in power systems. MLP is a feed-forward Artificial Neural Network model capable of creating a mapping function between sets of input data and a set of corresponding outputs. It consists of multiple layers of nodes, so-called neurons in neural networks, and each layer is connected to the other one. Neurons are the processing elements of the network composed of activation functions, such as linear, logarithmic sigmoid and tangent hyperbolic sigmoid functions. The last one is used as the activation function of neurons in this paper, as it results in better performance. Weights connecting the neurons of layers in the network and, biases connected to each neuron are free parameters that should be adjusted in the training phase.

In RBF neural networks, Radial Basis Functions are used as the activation functions of neurons in the hidden layer and linear functions for the neurons of the output layer. In this network, the vector distance between the input weights vector and the input vector is calculated (using the dot product of the two) and then multiplied by the bias. Afterwards, the result is transferred to the Radial Basis Function. The output of the first layer is then transferred to the second layer in which a linear function is the activation function of the output neuron. More details about MLP and RBF neural networks can be found in [28,45,49].

References

[1] Montoya FG, Manzano-Agugliaro F, Lopez-Marquez S, Hernandez-Escobedo Q, Gil C. Wind turbine selection for wind farm layout using multi-objective evolutionary algorithms. Expert Syst Appl 2014;41:6585–95.
[2] Manzano-Agugliaro F, Alcayde A, Montoya F, Zapata-Sierra A, Gil C. Scientific production of renewable energies worldwide: an overview. Renew Sustain Energy Rev 2013;18:134–43.
[3] World Wind Energy Association, URL <www.wwindea.org>.
[4] Alberta Electric System Operator, URL <https://www.aeso.ca>.
[5] Hernandez-Escobedo Q, Manzano-Agugliaro F, Zapata-Sierra A. The wind power of mexico. Renew Sustain Energy Rev 2010;14:2830–40.
[6] Hernandez-Escobedo Q, Manzano-Agugliaro F, Gazquez-Parra JA, Zapata-Sierra A. Is the wind a periodical phenomenon? the case of mexico. Renew Sustain Energy Rev 2011;15:721–8.
[7] Xu Y, Dong Z-Y, Xu Z, Meng K, Wong KP. An intelligent dynamic security assessment framework for power systems with wind power. IEEE Trans Ind Inform 2012;8:995–1003.
[8] Hu P, Karki R, Billinton R. Reliability evaluation of generating systems containing wind power and energy storage. IET Gen Transm Distrib 2009;3:783–91.
[9] Hernandez-Escobedo Q, Saldana-Flores R, Rodriguez-Garcia E, Manzano-Agugliaro F. Wind energy resource in northern mexico. Renew Sustain Energy Rev 2014;32:890–914.
[10] Rohrig K, Lange B. Improvement of the power system reliability by prediction of wind power generation. In: Power engineering society general meeting, 2007. USA: IEEE; 2007. p. 1–8.
[11] Cardell J, Anderson L, Tee CY. The effect of wind and demand uncertainty on electricity prices and system performance. In: 2010 IEEE PES transmission and distribution conference and exposition; 2010. p. 1–4.
[12] Black M, Strbac G. Value of bulk energy storage for managing wind power fluctuations. IEEE Trans Energy Convers 2007;22:197–205.
[13] New York Independent System Operator, URL <www.nyiso.com>.
[14] Chen N, Qian Z, Nabney I, Meng X. Wind power forecasts using gaussian processes and numerical weather prediction. IEEE Trans Power Syst 2013:1–10.
[15] Khalid M, Savkin A. A method for short-term wind power prediction with multiple observation points. IEEE Trans Power Syst 2012;27:579–86.
[16] Kou P, Liang D, Gao F, Gao L. Probabilistic wind power forecasting with online model selection and warped gaussian process. Energy Convers Manage 2014;84:649–63.
[17] Ramirez-Rosado IJ, Fernandez-Jimenez LA, Monteiro C, Sousa J, Bessa R. Comparison of two new short-term wind-power forecasting systems. Renew Energy 2009;34:1848–54.
[18] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[19] Gonzalez-Mingueza C, Munoz-Gutierrez F. Wind prediction using weather research forecasting model (wrf): a case study in peru. Energy Convers Manage 2014;81:363–73.
[20] Fan S, Liao J, Yokoyama R, Chen L, jen Lee W. Forecasting the wind generation using a two-stage network based on meteorological information. IEEE Trans Energy Convers 2009;24:474–82.
[21] Milligan M, Schwartz M, Wan Y. Statistical wind power forecasting models: results for u.s. wind farms, URL <http://www.nrel.gov/docs/fy03osti/33956.pdf>; 2003.
[22] Methaprayoon K, Yingvivatanapong C, jen Lee W, Liao J. An integration of ann wind power estimation into unit commitment considering the forecasting uncertainty. IEEE Trans Ind Appl 2007;43:1441–8.
[23] Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. Wind power forecasting: state-of-the-art 2009, URL <http://www.dis.anl.gov/pubs/65613.pdf>.
[24] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[25] Sfetsos A. A novel approach for the forecasting of mean hourly wind speed time series. Renew Energy 2002;27:163–74.
[26] Banos R, Manzano-Agugliaro F, Montoya F, Gil C, Alcayde A, Gomez J. Optimization methods applied to renewable and sustainable energy: a review. Renew Sustain Energy Rev 2011;15:1753–66.
[27] Barbounis T, Theocharis J. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 2006;69:466–96.
[28] Sideratos G, Hatziargyriou N. Using radial basis neural networks to estimate wind power production. In: Power engineering society general meeting, 2007. IEEE; 2007. p. 1–7.
[29] Potter C, Negnevitsky M. Very short-term wind forecasting for tasmanian power generation. IEEE Trans Power Syst 2006;21:965–72.
[30] Pousinho H, Mendes V, Catalao J. A hybrid PSO-ANFIS approach for short-term wind power prediction in portugal. Energy Convers Manage 2011;52:397–402.
[31] Amjady N, Keynia F, Zareipour H. A new hybrid iterative method for short-term wind speed forecasting. Eur Trans Electr Power 2011;21:581–95.
[32] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34:243–54.
[33] Nason G. Wavelet methods in statistics with R. Springer; 2008.
[34] Catalão J, Pousinho HMI, Mendes V. Hybrid intelligent approach for short-term wind power forecasting in portugal. IET Renew Power Gen 2011;5:251–7.
[35] Faria D, Castro R, Philippart C, Gusmao A. Wavelets pre-filtering in wind speed prediction. In: International conference on power engineering, energy and electrical drives, 2009. POWERENG '09; 2009. p. 168–73.
[36] de Aquino RRB, Lira MMS, de Oliveira J, Carvalho M, Neto O, de Almeida G. Application of wavelet and neural network models for wind speed and power generation forecasting in a brazilian experimental wind park. In: International joint conference on neural networks, 2009. IJCNN 2009; 2009. p. 172–8.
[37] Chen P, Chen H, Ye R. Chaotic wind speed series forecasting based on wavelet packet decomposition and support vector regression. In: IPEC, 2010 conference proceedings; 2010. p. 256–61.
[38] Ricalde LJ, Catzin G, Alanis AY, Sanchez EN. Higher order wavelet neural networks with kalman learning for wind speed forecasting. In: 2011 IEEE symposium on computational intelligence applications in smart grid (CIASG); 2011. p. 1–6.
[39] 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;2:265–76.
[40] Pindoriya N, Singh S, Singh S. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans Power Syst 2008;23:1423–32.
[41] Wu L, Shahidehpour M. A hybrid model for day-ahead price forecasting. IEEE Trans Power Syst 2010;25:1519–30.
[42] de Castro L, Von Zuben F. Learning and optimization using the clonal selection principle. IEEE Trans Evol Comput 2002;6:239–51.

[43] Wang Q, Wang C, Gao X. A hybrid optimization algorithm based on clonal selection principle and particle swarm intelligence. In: Sixth international conference on intelligent systems design and applications, 2006. ISDA '06. vol. 2; 2006. p. 975–9.
[44] Liao GC. Application of an immune algorithm to the short-term unit commitment problem in power system operation. IEE Proc Gen Transm Distrib 2006;153:309–20.
[45] Mathworks, URL <www.mathworks.com>.
[46] Slowik A, Bialko M. Training of artificial neural networks using differential evolution algorithm. In: 2008 conference on human system interactions; 2008. p. 60–5.
[47] Bessa R, Miranda V, Gama J. Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting. IEEE Trans Power Syst 2009;24:1657–66.
[48] Liu W, Pokharel PP, Principe JC. Correntropy: properties and applications in correntropy: properties and applications in non-gaussian signal processing. IEEE Trans Signal Process 2007;55:5286–98.
[49] Velo R, Lopez P, Maseda F. Wind speed estimation using multilayer perceptron. Energy Convers Manage 2014;81:1–9.