Wind Power Prediction by a New Forecast Engine Composed of Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization

Nima Amjady, *Senior Member, IEEE*, Farshid Keynia, *Member, IEEE*, and Hamidreza Zareipour, *Senior Member, IEEE*

***Abstract*—Following the growing share of wind energy in electric power systems, several wind power forecasting techniques have been reported in the literature in recent years. In this paper, a wind power forecasting strategy composed of a feature selection component and a forecasting engine is proposed. The feature selection component applies an irrelevancy filter and a redundancy filter to the set of candidate inputs. The forecasting engine includes a new enhanced particle swarm optimization component and a hybrid neural network. The proposed wind power forecasting strategy is applied to real-life data from wind power producers in Alberta, Canada and Oklahoma, U.S. The presented numerical results demonstrate the efficiency of the proposed strategy, compared to some other existing wind power forecasting methods.**

***Index Terms*—Feature selection, forecasting engine, hybrid neural network, particle swarm optimization, wind power forecasting.**

I. Introduction

INTEGRATION of wind-powered generators into electric power systems is growing at a significant rate in many countries around the world. For example, the share of wind power generation in the United States has been increasing with an annual rate of 25% since 1990 [1]. It is estimated that by 2020, about 12% of the world's electricity will be supplied by wind generation [2]. Wind generation installation has reached considerable percentages (in the range of 5% to 20%) of the whole installed capacity in some European countries in recent years, such as Germany, Spain, and Denmark [3]. Thus, wind energy is becoming an important component in the supply mix to meet the growing demand for electric energy.

Wind power is proportional to the air density, the intercepting area, and the wind velocity to the third power. The air density is an exponential function of air pressure and temperature [4]. Thus, wind power is dependent on the weather situation and parameters, such as wind speed, wind direction, and temperature. In addition, wind power of each time interval has high relevance with its past values, known as the short run trend characteristic [5]. Hence, wind power at future time intervals may be partially explained by its values at the past time intervals.

Despite the environmental benefits of wind power, its variability makes it a nondispatchable resource, which can potentially put system reliability in jeopardy. Forecasting future variations of wind power is acknowledged as a tool for mitigating wind integration challenges. In [3], it has been discussed that wind power forecast has become a major concern for transmission system operators (TSOs), wind park owners and regulators all support efforts to develop better wind power prediction methods. In Alberta, Canada, establishing a central wind power forecasting center for facilitating wind power integration has been recognized and pursued by the Alberta Electric System Operator (AESO) [6]. However, wind power is a nonlinear function of wind characteristics, such as wind speed, air density, and turbulent kinetic energy [7]. A discussion about the complexities of wind power forecast (such as strong nonstationary and hard nonlinear behaviors, dependency on many exogenous variables, and high volatility of wind power time series) can be found in [2]–[4]. Therefore, reasonably accurate wind power prediction is a difficult task, requiring a forecasting method with high learning capability to extract any existing patterns in the available data.

Wind power forecasting methods can be generally classified into physical models and statistical or time-series models [8], [9]. The physical approaches attempt to estimate local wind speed using the physical laws governing atmospheric behavior and then the corresponding power generated at the wind farm level [10]. These models use physical considerations about the site, such as surface roughness, orography, obstacles, stratification of the atmosphere, pressure, and temperature to estimate the future wind speed and generated power [11]. The statistical approaches aim to determine the relationship between a set of explanatory variables and the power generated at the wind farm using historical data [10]. These models can also be divided in two categories, i.e., models that only employ time series data and predict future values taking into account the past history and models that use numerical weather predictions (such as forecasted values for wind speed and direction) in addition to the electric power time series data [3]. A review of wind power forecasting research works can be found in [4] and [11]–[13]. Our proposed wind power forecast method is a statistical model of the second category.

Although the existing methods have made considerable improvements over the years, more accurate and robust wind

Manuscript received July 01, 2010; revised December 17, 2010; accepted February 06, 2011. Date of publication February 14, 2011; date of current version June 22, 2011.

N. Amjady and F. Keynia are with the Department of Electrical Engineering, Semnan University, Semnan 35195-363, Iran (e-mail: amjady@tavanir.org.ir; keynia@yahoo.com).

H. Zareipour is with the Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, T2N 1N4, Canada (e-mail: h.zareipour@ucalgary.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2011.2114680

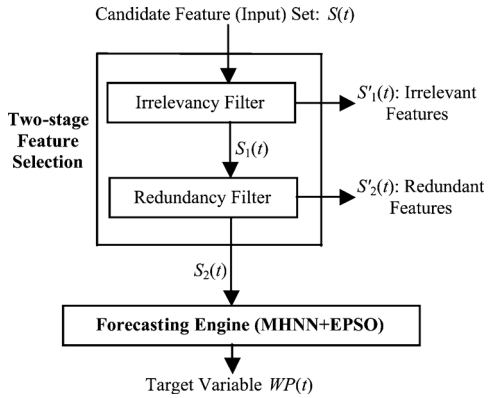

Fig. 1. Structure of the proposed wind power forecast strategy.

power forecast methods are still demanded. For instance, based on a study by Basin Electric Power Cooperative, a wholesale power supplier to several cooperatives in the Midwest U.S., the accuracy of next-hour forecasting is currently poor such that for about one third of the time, errors in the next-hour wind generation forecasts are greater than 50% [14]. On the other hand, improvement in the accuracy of wind power forecast is critical for both wind park owners and system operators.

In this paper, a new wind power forecasting strategy is proposed and its efficiency is demonstrated using several numerical experiments. Following many previous research works in the area, such as [2], [10], [12], and [15], and without losing generality, this paper focuses on hourly forecasts with a forecasting horizon of several hours ahead (e.g., 24 hours ahead). The main contributions of this paper can be summarized as follows:

1) A novel and efficient forecast method for wind power prediction is proposed. The forecast method is composed of a new enhanced particle swarm optimization technique (EPSO) and modified hybrid neural network (MHNN). It has high learning capability and can avoid the overfitting problem and trapping in local minima and dead bands.
2) Most of the previous research work in the area of wind power prediction focuses on the forecast methods, while the design of the input vector is usually carried out in a discretionary way, mainly based on heuristics or trial-and-error procedures. In this paper, an effective feature selection technique, based on the information-theoretic criteria, is applied to select the most informative candidate inputs for the proposed forecast engine. This feature selection technique can process any set of candidate inputs filtering out its irrelevant and redundant features.

The remaining parts of the paper are organized as follows. In Section II, the proposed wind power forecast strategy is introduced. In Section III, several other prediction approaches, presented in the literature, are implemented and compared with the proposed wind power forecast strategy. Moreover, we also compare the obtained results from the proposed strategy with the published literature figures in the same test conditions. Section IV concludes the paper.

II. Proposed Forecasting Strategy

The structure of the proposed wind power forecast strategy is shown in Fig. 1. Briefly, the proposed strategy is composed of a two-stage feature selection component and a wind power forecasting engine, which are introduced in Sections II-A and II-B, respectively.

A. The Employed Two-Stage Feature Selection

Wind power can be seen as a nonlinear mapping function of several exogenous meteorological variables and its past values. Assuming the past and forecast values of the exogenous variables, such as wind speed, wind direction, temperature, and humidity, are available at the wind farm location or a weather station close to the wind farm, a set of candidate forecasting features (inputs), say $S(t)$, can be constructed as follows:

$$\begin{aligned} S(t) = \{&\mathrm{WP}(t-1), \ldots, \mathrm{WP}(t-N_{\mathrm{WP}}), \mathrm{WS}(t), \\ &\mathrm{WS}(t-1), \ldots, \mathrm{WS}(t-N_{\mathrm{WS}}), \mathrm{WD}(t), \\ &\mathrm{WD}(t-1), \ldots, \mathrm{WD}(t-N_{\mathrm{WD}}), \\ &\mathrm{T}(t), \mathrm{T}(t-1), \ldots, \mathrm{T}(t-N_{\mathrm{T}}), \mathrm{H}(t), \\ &\mathrm{H}(t-1), \ldots, \mathrm{H}(t-N_{\mathrm{H}})\} \end{aligned} \quad (1)$$

where $\mathrm{WP}(t)$, $\mathrm{WS}(t)$, $\mathrm{WD}(t)$, $\mathrm{T}(t)$, and $\mathrm{H}(t)$ indicate wind power, wind speed, wind direction, temperature, and humidity at time interval t. The time interval depends on the desired forecasting step; for instance, for hourly wind generation forecast, t is measured in terms of hours. The candidate inputs $\mathrm{WP}(t-1), \ldots, \mathrm{WP}(t-N_{\mathrm{WP}})$ in (1) are the historical values of wind power, since wind power is dependent on its past values. $\mathrm{WS}(t)$ is the forecast value of wind speed for time interval t and $\mathrm{WS}(t-1), \ldots, \mathrm{WS}(t-N_{\mathrm{WS}})$ are its past values, considering that wind speed is an important driver for wind power. Similarly, the forecast and past values of wind direction, temperature, and humidity are included in the set of candidate features $S(t)$. In (1), N_{WP} indicates the order of back shift for the wind power candidate features, which sometimes is referred to as the order of the dynamical forecast system [2]. Similarly, N_{WS}, N_{WD}, N_{T}, and N_{H} are defined. From a data mining view point, these orders should be considered high enough so that no useful information is missed. However, a compromise is always necessary to avoid a too large set of candidate features. For instance, even by considering the low orders of $N_{\mathrm{WP}} = N_{\mathrm{WS}} = N_{\mathrm{WD}} = N_{\mathrm{T}} = N_{\mathrm{H}} = 24$ (only including the features of 24 hours ago for hourly wind power forecast), we reach 124 candidate features in (1). Moreover, if the exogenous variables related to the wind farm are measured in more than one place (e.g., in seven weather stations considered in [2]), the candidate inputs of the exogenous variables should be repeated by the number of measurement sites. However, such a large set of inputs is not directly applicable to a forecasting engine, since it may include ineffective features, which complicate the extraction of input/output mapping function of the prediction process for the forecast engine and degrade its performance. Thus, the set of candidate inputs $S(t)$ should be refined by a feature selection technique such that a minimum subset of the most informative features is selected and the other unimportant candidates are filtered out. This process is also shown in Fig. 1.

Correlation analysis has been proposed for the feature selection of wind power forecast in [1] and [2]. However, wind power is a nonlinear mapping function of several input variables, whereas correlation analysis is a linear feature selection

technique. Thus, it may not correctly evaluate the information value of the candidate inputs for wind power forecasting. Moreover, correlation analysis only considers the relevancy between the target variable (here, wind power of the next hour or $\mathrm{WP}(t)$) and candidate inputs to rank them. However, in feature selection, it has been recognized that the combinations of individually good features do not necessarily lead to good estimation performance. In other words, "the m best features are not the best m features" [16]. It should also be noted that relevant features in a forecasting process may include redundant information. Redundant features can potentially degrade the learning process of the forecasting engine, besides the associated unnecessary extra computation burden. Hence, an efficient nonlinear feature selection technique, which can evaluate both relevancy and redundancy of the candidate features, is necessary.

In our previous work [17], we proposed a nonlinear feature selection technique based on the information-theoretic criterion of mutual information (MI). This feature selection technique evaluates MI of each candidate input with the target variable and ranks the candidates based on their MI. The MI-based feature selection method of [18] can effectively evaluate relevance of the target variable to each candidate input and select relevant inputs. However, it does not consider the redundancy of candidate features. In [19], we enhanced the feature selection method of [17] and proposed an effective two-stage feature selection technique to filter out both irrelevant and redundant candidate inputs. This two-stage feature selection technique is applied to the wind power prediction problem in the present paper and its performance is compared with some other previously reported feature selection methods. Below, a short description of the employed feature selection technique is provided; further details may be found in [17]–[19].

1) First Stage (Irrelevancy Filter): Mutual information between each candidate input $x(t) \in S(t)$ and the target variable $\mathrm{WP}(t)$, denoted by $\mathrm{MI}[x(t), \mathrm{WP}(t)]$, is computed according to the binomial distribution technique of [17]. For instance, $x(t)$ can be any of candidate inputs shown in (1). A higher value of $\mathrm{MI}[x(t), \mathrm{WP}(t)]$ means $x(t)$ has more common information content with $\mathrm{WP}(t)$ and so $x(t)$ is a more relevant feature for forecasting $\mathrm{WP}(t)$. Thus, the candidate features of $S(t)$ are sorted based on their mutual information with the target feature such that a higher $\mathrm{MI}[x(t), \mathrm{WP}(t)]$ value results in a higher rank. Then, the candidate inputs with $\mathrm{MI}[x(t), \mathrm{WP}(t)]$ value greater than a relevancy threshold $\mathrm{TH1}$ are considered as the relevant features of the forecast process denoted by subset $S_1(t) \subset S(t)$. Other candidate inputs with $\mathrm{MI}[x(t), \mathrm{WP}(t)]$ lower than $\mathrm{TH1}$ are considered as irrelevant features, denoted by subset $S_1'(t) \subset S(t)$. Observe that $S_1 \cup (t) S_1'(t) = S(t)$. The subset $S_1(t)$ is retained for the next stage while $S_1'(t)$ is filtered out as shown in Fig. 1.

2) Second Stage (Redundancy Filter): In this stage, redundant features among the selected candidate inputs of $S_1(t)$ are found and filtered out. Higher value of mutual information between two selected candidates $x_k(t) \in S_1(t)$ and $x_m(t) \in S_1(t)$, i.e., higher $\mathrm{MI}[x_k(t), x_m(t)]$, means more common information between $x_k(t)$ and $x_m(t)$ and so these candidates have a higher level of redundancy. Therefore, the following redundancy criterion $\mathrm{RC}[.]$ is defined to measure the redundancy of each selected feature $x_k(t) S_1(t)$ with the other candidate inputs of $S_1(t)$

$$\mathrm{RC}\left[x_k(t)\right] = \max_{x_m(t) \in S_1(t) - \{x_k(t)\}} \left(\mathrm{MI}\left[x_k(t), x_m(t)\right]\right). \quad (2)$$

Observe that $x_m(t) S_1(t) - \{x_k(t)\}$ in (2), since each feature is fully redundant with itself and so we should exclude $x_k(t)$ from $S_1(t)$. We can rank the candidate features of $S_1(t)$ according to the redundancy measure of (2) such that a higher value of $\mathrm{RC}[x_k(t)]$ means $x_k(t)$ is a more redundant feature or equivalently a less informative candidate input. If $\mathrm{RC}[x_k(t)]$ becomes greater than a redundancy threshold $\mathrm{TH2}$, $x_k(t)$ is considered as a redundant candidate input and so between this candidate and its rival, one feature should be filtered out. For instance, suppose that $x_k(t)$ has the highest redundancy (mutual information) with $x_r(t)$ among all features of the subset $S_1(t) - \{x_k(t)\}$, i.e.,

$$\arg \max_{x_m(t) \in S_1(t) - \{x_k(t)\}} \left(\mathrm{MI}\left[x_k(t), x_m(t)\right]\right) = x_r(t). \quad (3)$$

In such a case, we call $x_r(t)$ the rival of $x_k(t)$. If $\mathrm{MI}[x_k(t), x_r(t)] > \mathrm{TH2}$, between $x_k(t)$ and its rival $x_r(t)$, one feature should be eliminated. For this purpose, the relevancy factors of these features, i.e., $\mathrm{MI}[x_k(t), \mathrm{WP}(t)]$ and $\mathrm{MI}[x_r(t), \mathrm{WP}(t)]$, are considered and the feature with lower relevancy factor (less relevant feature or less effective feature for the forecast process) is filtered out. The redundancy filtering process is repeated for all features of $S_1(t)$ until no redundancy measure of (2) becomes greater than $\mathrm{TH2}$. The remaining features of $S_1(t)$, owning $\mathrm{RC}[.]$ less than $\mathrm{TH2}$, constitute the selected features of the two-stage feature selection technique, denoted by $S_2(t)$ in Fig. 1. The subset $S_2(t)$ includes relevant and nonredundant candidate inputs. The redundant features filtered in this stage comprise the set of $S_2'(t)$ in Fig. 1. Tuning the values of the thresholds $\mathrm{TH1}$ and $\mathrm{TH2}$ is a trade-off between quality and number of selected features. Higher values of $\mathrm{TH1}$ and $\mathrm{TH2}$ result in the selection of more relevant and less redundant candidate features, but with the cost of having less selected inputs. Numerically adjusting the values of $\mathrm{TH1}$ and $\mathrm{TH2}$ and other adjustable parameters of the forecast engine is described in Section II-B. The selected candidate features in $S_2(t)$ are considered as the inputs of the wind power forecast engine as shown in Fig. 1. Thus, the forecast engine should implement the prediction process of $S_2(t) \rightarrow \mathrm{WP}(t)$.

Although this feature selection technique has been proposed in our previous work [19] for price spike analysis, demonstrating its efficiency in a wind power forecasting application is one of the contributions of the present paper. The previous research works on wind power prediction focus on forecasting models and model inputs are selected based on heuristics, trial-and-error, or at most correlation analysis. However, in this paper, an efficient information-theoretic feature selection method is presented for this purpose evaluating information value of the candidate inputs, from both relevancy and redundancy view points.

B. Proposed Wind Power Forecasting Engine

An efficient hybrid neural network (HNN), with the architecture shown in Fig. 2, has been proposed in our previous work

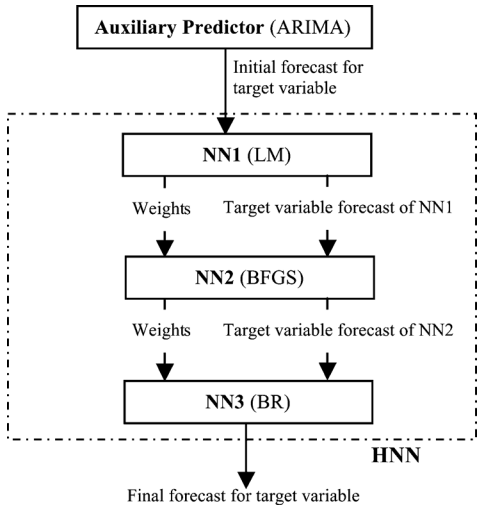

Fig. 2. Architecture of HNN and its auxiliary predictor.

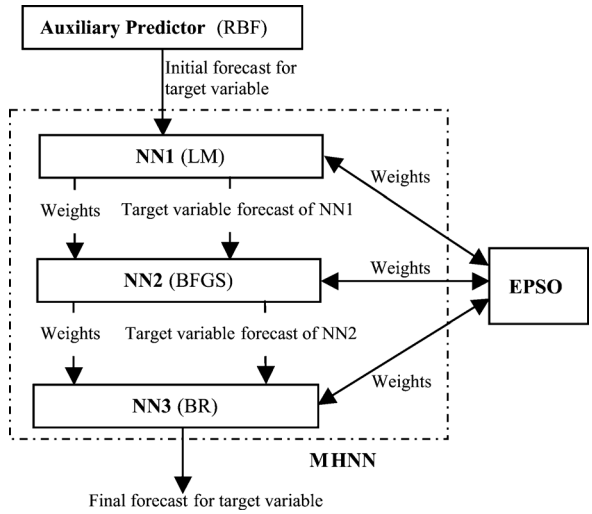

Fig. 3. Architecture of the proposed wind power forecast engine.

[18] for electricity market price forecasting. In the present paper, first, a modified version of the HNN, named MHNN, is introduced. Next, the MHNN is combined with a new particle swarm optimization (PSO) component, refereed to here as enhanced PSO (EPSO). This combination (MHNN+EPSO) results in the proposed wind power forecast engine of Fig. 1. In Section II-B1, first, the structure and philosophy of the HNN is briefly discussed, following by those of the MHNN.

1) HNN and Its Conversion to the MHNN: A well-designed combination of different neural networks (NNs) can potentially enhance their learning capability in modeling a complex process. For example, some parallel and cascaded structures for combining NNs with improved performance have been proposed in [2] and [20]–[22]. However, these structures share the input data among their building blocks. Thus, the knowledge of one block is not really shared with other blocks. In the HNN of [18], on the other hand, a knowledge transfer procedure is envisioned from one NN to another. The HNN of [18] is composed of three NNs having the same multilayer perceptron (MLP) structure. MLP is an efficient architecture for estimating neural networks. In addition, according to Kolmogorov's theorem, MLP can solve a problem by using one hidden layer, provided it has the proper number of neurons [17]. Thus one hidden layer has been considered in the MLP structure of all the NNs of the HNN. This choice has the advantage of having a minimum number of parameters to be adjusted numerically.

Fig. 2 shows that each NN of the HNN transfers two sets of information to the next NN. The first set includes the obtained values for weights. In other words, each NN transfers its obtained knowledge to the next one and so it can begin its learning process from the point that the previous NN terminated (instead of beginning from a random point). Only the first NN should begin with an initial set of random values for the weights. Since all NNs have the same MLP structure, these weights can be directly used by the next NN and then it can increase the obtained knowledge of the previous one. In addition, selecting a variety of MLP training algorithms of the NNs, the HNN benefits from wider learning capability. The three NNs of HNN have LM (Levenberg–Marquardt), BFGS (Broyden, Fletcher, Goldfarb, Shanno), and BR (Bayesian Regularization) learning algorithms. Further discussions justifying these choices can be found in [18].

The discussed HNN is an efficient forecast engine and provides good forecast accuracy for price forecast of electricity markets as shown in [18]. However, wind power usually is a more complex and volatile signal than electricity price. For instance, the periodic behavior (such as daily and weekly periodicities) usually seen in electricity price time series [17], is much weaker in wind power time series. Furthermore, the most important driver for electricity price is the demand, which in turn follows regular patterns and shows a strong periodic behavior [17]. On the other hand, the most important driver of wind power is wind speed [2], [14], which is a highly volatile signal as discussed in [11], with less predictable patterns.

In view of the above discussion, an improved version of the above HNN forecast engine is proposed for wind power prediction in this paper. As seen from Fig. 2, the second set of results transferred between NNs of the HNN is the forecast of target variable. In this way, each NN also has an initial prediction of target variable, which helps it to produce a more accurate forecast. However, there is no NN before the first one. Thus, the initial forecast for NN1, which is the initial forecast for the HNN, is produced by an auxiliary predictor as shown in Fig. 2. Thus, the auxiliary predictor is an important component for this forecast method, since a more accurate initial forecast of it results in a more accurate forecast of each NN and so a more accurate final forecast (provided by NN3 as shown in Fig. 2). In the HNN proposed in [18], a multivariate ARIMA, which is a linear forecaster, was used as the auxiliary predictor (Fig. 2). However, wind power is a nonlinear mapping function of its inputs and thus, a radial basis function (RBF) neural network, which is a nonlinear forecaster, is used instead of the multivariate ARIMA here. RBF is effective in exploring local characteristics of the input data, whereas MLP networks are good at capturing global data trends [21]. Therefore, by combining RBF as the auxiliary predictor with the MLP networks of HNN, the resultant forecast engine can capture both local and global behaviors of the target variable. Indeed, we examined different linear and nonlinear forecasters (such as multivariate ARIMA, support vector machine, different MLP neural networks, and RBF) as the auxil-

iary predictor and RBF resulted in the best performance among them (some of these comparative results will be presented in Section III). Here, the HNN with this modification in its auxiliary predictor is called the MHNN as shown in Fig. 3.

A second modification to the HNN of [18] here is to combine the proposed MHNN with a new stochastic search technique (EPSO). This combination is shown in the present work to increase the learning capability of the MHNN in extracting the input/output mapping function of a complex signal, such as wind power. When one of the NNs of the MHNN is trapped in a local minimum during the training phase, neither that NN nor the next ones may be able to escape from the local minimum. To remedy this problem, the EPSO component is added to each NN of the MHNN. Owing to the high exploration capability of the EPSO component, an NN trapped in a local minimum can be released. In Sections II-B2 and II-B3, the EPSO component is first described. Then, the combination of the MHNN and the EPSO component to construct the proposed forecast engine is detailed.

It should be noted that to implement the forecast process of $S_2(t) \rightarrow \mathrm{WP}(t)$ by MHNN, the selected candidate inputs of $S_2(t)$ are given to both the auxiliary predictor and all NNs of MHNN (each NN of MHNN also receives the initial forecast of the previous block) and the target variable is $\mathrm{WP}(t)$. Multiperiod forecast, e.g., prediction of wind power for the next 24 hours, is reached via recursion, i.e., by feeding input variables with the forecaster's outputs. For instance, predicted wind power for the first hour is used as $\mathrm{WP}(t-1)$ for the wind power forecast of the second hour provided that $\mathrm{WP}(t-1)$ is among the selected candidate inputs of $S_2(t)$.

2) Presentation of the EPSO: PSO is a population-based search method which deals with random particles in a solution space. The particles, i.e., trial solutions of the optimization problem, share their information with each other and run toward the best trajectory to find the optimum solution in an iterative process. A velocity vector is defined for each particle and the particle position depends on this velocity. In each iteration, the velocity and position of the particles are updated as follows:

$$\begin{aligned} V_{i,\mathrm{iter}+1} =& \omega V_{i,\mathrm{iter}} \\ &+ c_1 r_1 (P^{\mathrm{best}}_{i,\mathrm{iter}} - X_{i,\mathrm{iter}}) \\ &+ c_2 r_2 (G^{\mathrm{best}}_{\mathrm{iter}} - X_{i,\mathrm{iter}}) \end{aligned} \tag{4}$$

$$X_{i,\mathrm{iter}+1} = X_{i,\mathrm{iter}} + V_{i,\mathrm{iter}+1} \tag{5}$$

where $V_{i,\mathrm{iter}}$ and $X_{i,\mathrm{iter}}$ represent the velocity vector and position vector of the ith particle at iteration iter, respectively; $P^{\mathrm{best}}_{i,\mathrm{iter}}$ and $G^{\mathrm{best}}_{i,\mathrm{iter}}$ are personal best position of the ith particle and global best position of swarm until iteration iter, respectively; ω is inertia factor controlling the global and local exploration capabilities of particles; c_1 and c_2 are cognitive and social coefficients, respectively; r_1 and r_2 are two random numbers between 0 and 1. More details about classical PSO can be found in [23]. Here, an enhanced version of PSO is proposed.

In [24], a split-up in the cognitive behavior of PSO into $P^{\mathrm{best}}_{i,\mathrm{iter}}$ and $P^{\mathrm{best}}_{i,\mathrm{iter}}$ is proposed that changed (4) as follows:

$$\begin{aligned} V_{i,\mathrm{iter}+1} =& \omega V_{i,\mathrm{iter}} + c_{1b} r_1 (P^{\mathrm{best}}_{i,\mathrm{iter}} - X_{i,\mathrm{iter}}) \\ &+ c_{1w} r_2 (X_{i,\mathrm{iter}} - P^{\mathrm{worst}}_{i,\mathrm{iter}}) + c_2 r_3 (G^{\mathrm{best}}_{\mathrm{iter}} - X_{1,\mathrm{iter}}) \end{aligned} \tag{6}$$

where $P^{\mathrm{best}}_{i,\mathrm{iter}}$ is exactly the same as the cognitive component of the classical PSO and $P^{\mathrm{worst}}_{i,\mathrm{iter}}$ is personal worst position of the ith particle until iteration iter; c_{1b} and c_{1w} are the coefficients of the new cognitive part corresponding to $P^{\mathrm{worst}}_{i,\mathrm{iter}}$ and $P^{\mathrm{best}}_{i,\mathrm{iter}}$, respectively; r_1, r_2, and r_3 are random numbers between 0 and 1. That is, the particle is made to remember its worst position also. This modification helps to search the solution space more effectively compared with the classical PSO and has also been used in some other research works. However, our experience shows that after a few initial iterations, the swarm particles usually become better and better and so, $P^{\mathrm{worst}}_{i,\mathrm{iter}}$ of the ith particle approximately remains unchanged. In other words, $P^{\mathrm{worst}}_{i,\mathrm{iter}}$ nearly becomes a static limit and loses its dynamic behavior during the search process. Thus, the expected enhancement in the exploration capability of the PSO may not be obtained. To solve this problem, we propose the new idea of decomposing the cognitive part into the "Best" and "Not-best" components instead of splitting up into the "Best" and "Worst" components, as follows:

$$\begin{aligned} V_{i,\mathrm{iter}+1} =& \omega V_{i,\mathrm{iter}} + c_{1b} r_1 (P^{\mathrm{best}}_{i,\mathrm{iter}} - X_{i,\mathrm{iter}}) \\ &+ c_{1nb} r_2 (X_{i,\mathrm{iter}} - P^{\mathrm{not-best}}_{i,\mathrm{iter}}) \\ &+ c_2 r_3 (G^{\mathrm{best}}_{\mathrm{iter}} - X_{i,\mathrm{iter}}) \end{aligned} \tag{7}$$

$$\begin{cases} P^{\mathrm{best}}_{i,\mathrm{iter}} = X_{i,\mathrm{iter}}, & \text{if } \mathrm{OF}(X_{i,\mathrm{iter}}) < \mathrm{OF}(P^{\mathrm{best}}_{i,\mathrm{iter}-1}) \\ P^{\mathrm{best}}_{i,\mathrm{iter}} = P^{\mathrm{best}}_{i,\mathrm{iter}-1}, & \text{if } \mathrm{OF}(X_{i,\mathrm{iter}}) \geq \mathrm{OF}(P^{\mathrm{best}}_{i,\mathrm{iter}-1}) \end{cases} \tag{8}$$

$$\begin{cases} P^{\mathrm{not-best}}_{i,\mathrm{iter}} = X_{i,\mathrm{iter}}, & \text{if } \mathrm{OF}(X_{i,\mathrm{iter}}) \geq \mathrm{OF}(P^{\mathrm{best}}_{i,\mathrm{iter}-1}) \\ P^{\mathrm{not-best}}_{i,\mathrm{iter}} = P^{\mathrm{not-best}}_{i,\mathrm{iter}-1} & \text{if } \mathrm{OF}(X_{i,\mathrm{iter}}) < \mathrm{OF}(P^{\mathrm{best}}_{i,\mathrm{iter}-1}) \end{cases} \tag{9}$$

where $\mathrm{OF}(.)$ is the objective function of the optimization problem that should be minimized (without loss of generality, we assume a minimization problem); the coefficients of c_{1b} and c_{1nb} in (7) are related to the "Best" and "Not-best" components, respectively. Inertia factor and cognitive and social coefficients are user defined parameters of a PSO. In the proposed EPSO, ω, c_{1b}, c_{1nb}, and c_2 of (7) are adaptively set [25] to enhance the search ability and convergence behavior of the EPSO.

Equation (8) describes $P^{\mathrm{best}}_{i,\mathrm{iter}}$ according to the classical PSO. The new component of "Not-best" is defined in (9). If the ith particle in iteration iter can find a solution better than its best solution until the previous iteration, i.e., $P^{\mathrm{best}}_{i,\mathrm{iter}-1}$, the "Best" component of the current iteration $P^{\mathrm{best}}_{i,\mathrm{iter}}$ is set to $X_{i,\mathrm{iter}}$ as shown in (8), since $X_{i,\mathrm{iter}}$ is the newly found personal best position in this case, while the "Not-best" component $P^{\mathrm{not-best}}_{i,\mathrm{iter}}$ is replaced by its previous value $P^{\mathrm{not-best}}_{i,\mathrm{iter}-1}$ as shown in (9). On the other hand, if the ith particle in iteration iter cannot find a solution better than $P^{\mathrm{best}}_{i,\mathrm{iter}-1}$, $P^{\mathrm{not-best}}_{i,\mathrm{iter}}$ is updated by $X_{i,\mathrm{iter}}$ as shown in (9), since $X_{i,\mathrm{iter}}$ in this case is considered as the new "Not-best" solution, while $P^{\mathrm{best}}_{i,\mathrm{iter}}$ is replaced by its previous value $P^{\mathrm{best}}_{i,\mathrm{iter}-1}$ as shown in (8). In other words, the limits $P^{\mathrm{best}}_{i,\mathrm{iter}}$ and $P^{\mathrm{not-best}}_{i,\mathrm{iter}}$ are revised in the opposite directions in each iteration. In this way, the limit is continuously updated (even in the last iterations of the PSO) and saves its dynamic behavior throughout the search process of the PSO enhancing its exploration capability. The PSO depicted in (7)–(9) is called EPSO in this paper.

3) Combination of the MHNN and the EPSO to Construct the Proposed Wind Power Forecasting Engine: The hybridization approach of MHNN and EPSO to construct the proposed

wind power forecast engine is depicted in Fig. 3. NN1 is trained first by the LM learning algorithm. The training phase is terminated based on the early stopping condition to avoid the overfitting problem [17]. Then, the obtained weight values are transferred to the EPSO as shown in Fig. 3. The EPSO continues the training process of NN1 by modeling this process as an optimization problem. The objective function of the optimization problem, which is $\mathrm{OF}(.)$ in (8) and (9), is the error function of NN1 that should be minimized. Error of validation samples or validation error [26] is adopted as the error function in the training process for both the learning algorithms of the MHNN (LM, BFGS, and BR) and EPSO. Note that the NNs of the MHNN and the EPSO component have the same training samples and validation samples such that the weights can be transferred between them. Consequently, the EPSO tries to further minimize the validation error of NN1 after the LM learning algorithm is terminated. Although LM, BFGS, and BR are computationally efficient learning algorithms, they search the solution space in a specific direction (such as steepest descent [27]) and thus, these learning algorithms may be trapped in a local minimum to learn the nonlinear prediction process of $S_2(t) \rightarrow \mathrm{WP}(t)$, without being able to escape out. On the contrary, the EPSO algorithm, with its enhanced exploration capability, can widely search the solution space in various directions and hence, the training process is more likely to escape out of the local minimum.

To continue the training process of NN1 (further minimize the validation error), in general, the position vector of the ith particle of the EPSO or X_i includes the weights of NN1 (each dimension of the position vector is a weight of NN1). In other words, the particles of the EPSO, containing the decision variables of the optimization problem, are potential solutions for the weight vector of NN1. To begin the search process of the EPSO, at first, its particles should be initialized. Usually, particles of a PSO are randomly initialized. Here, to also transfer the obtained knowledge of the LM learning algorithm to the EPSO, one of its particles is initially set to the obtained solution of the LM for the weights and the positions of the other particles are randomly initialized:

$$\text{The initial Swarm of the EPSO} = X_{1,0}, X_{2,0}, \ldots, X_{\mathrm{NP},0} \tag{10}$$

$$X_{1,0} = W_{\mathrm{LM}} \tag{11}$$

$$X_{2,0}, \ldots, X_{\mathrm{NP},0}\text{: Randomly initialized weight vectors} \tag{12}$$

where NP is number of particles of the EPSO. In (10), $X_{1,0}, X_{2,0}, \ldots, X_{\mathrm{NP},0}$ indicate the initial positions of the NP particles of the EPSO (the second subscript or iter is zero). In (11), W_{LM} is the obtained result for the weight vector of NN1 by the LM learning algorithm. Also, the initial velocity vectors of the NP particles $\{V_{1,0}, V_{2,0}, \ldots, V_{\mathrm{NP},0}\}$ are randomly set within the allowable ranges and the "Best" and "Not-best" limits of the particles are initialized as follows:

$$P_{i,0}^{\mathrm{best}} = P_{i,0}^{\mathrm{not-best}} = X_{i,0},\ 1, \ldots, \mathrm{NP}. \tag{13}$$

After the initialization, the particles of the EPSO move and search the solution space, iteratively. For this purpose, (7), (5), (8), and (9) are executed ($i = 1, \ldots, \mathrm{NP}$) and the velocity vector, position vector, "Best" limit, and "Not-best" limit of the NP particles are updated, respectively. Then, iter is incremented and the cycle is repeated until the stopping condition of the EPSO is satisfied. Here, if the best particle of the EPSO (owning the minimum value of the validation error or $\mathrm{OF}(.)$ among all particles) does not change in three successive iterations, its search process is terminated. Then, the position vector (weight vector) of the best particle of the EPSO is returned to NN1, which are considered as the final weights of NN1. At this point, the training process of NN1 is completed. Next, to begin the training process of NN2, it receives the final weights of NN1 as the initial weight values as shown in Fig. 3. Starting from this initial point, the training process BFGS $\leftrightarrow$ EPSO of NN2 is executed similar to that of the LM $\leftrightarrow$ EPSO of NN1. NN2 is trained by the BFGS learning algorithm until it is stopped based on the early stopping condition (like the LM learning algorithm of NN1). Then the weight values obtained by the BFGS are transferred to the EPSO and it searches for a better weight vector as previously described. The best weight vector found by the EPSO (the position vector of the best particle) constitutes the final weights of NN2. After completing the training process of NN2, its final weights are sent to NN3. The training process BR $\leftrightarrow$ EPSO of NN3, starting from these weight values, is executed similar to that of the BFGS $\leftrightarrow$ EPSO of NN2. After completing the training process of NN3, all NNs of MHNN are trained. The auxiliary predictor, i.e., RBF, is also trained [26] by the same historical data of the NNs of the MHNN.

At this stage, the whole proposed wind power forecast engine is trained and ready for the prediction of the future values of wind power. In the prediction phase, at first, RBF predicts the value of wind power for the next time interval, i.e., $\mathrm{WP}(t)$. For this purpose, the inputs of RBF are the selected candidate features of $S_2(t)$. Its prediction is submitted as the initial forecast for NN1 of the MHNN as shown in Fig. 3. Therefore, the inputs of NN1 include the selected candidate features of $S_2(t)$ plus the initial forecast of the auxiliary forecaster. Using these inputs, NN1, based on its final weight values (obtained from the training process LM $\leftrightarrow$ EPSO), generates a better forecast for $\mathrm{WP}(t)$, which is submitted for NN2. Similarly, NN2 and NN3 produce their predictions until the final forecast of $\mathrm{WP}(t)$ is obtained from NN3 of MHNN.

The only remaining part of the proposed wind power forecast strategy is the numerical refinement of its adjustable parameters, including TH1 and TH2 of the two-stage feature selection technique and the number of nodes in the single hidden layer of the NNs of MHNN. As described in Section II-B1, all NNs of MHNN have the same MLP structure with one hidden layer. Here, these adjustable parameters are fine tuned by the line search procedure of [17], which is a computationally efficient cross-validation technique.

III. Numerical Results

The proposed wind power forecasting strategy, including the two-stage feature selection technique and the wind power forecasting engine, is tested by the real data from wind farms in Alberta, Canada and Blue Canyon wind farm in southwestern Oklahoma, the United States. The Alberta test case includes Castle River wind farm (owning 60 turbines and 39.6-MW total

capacity), McBride Lake wind farm (owning 114 turbines and 75-MW total capacity), and Summerview wind farm (owning 38 turbines and 68-MW total capacity). For this test case, we focus on the prediction of the aggregated wind power of the three wind farms with weather data obtained from the Pincher Creek weather station. The Blue Canyon wind farm test case includes 45 turbines with a nameplate capacity of 74 MW [2]. The required wind power and weather data are obtained from [28] and [29] for the first test case and from [30] and [31] for the second test case, respectively. The numerical weather predictions used for the Alberta and Oklahoma test cases have been originated from the Global Environmental Multiscale (GEM) model of the Canadian Meteorological Centre (CMC) [29] and the high-resolution Global Forecast System (GFS) of the U.S. National Centers for Environmental Prediction (NCEP) [31], respectively.

The numerical experiments of this paper are designed to evaluate the effectiveness of the employed two-stage feature selection technique for wind power forecasting in a comparative manner, to demonstrate the capabilities of the proposed wind power forecasting engine, and finally to show the forecasting performance of the entire wind power forecasting strategy, i.e., feature selection technique + forecasting engine. The obtained results from these numerical experiments are presented in Sections III-A–III-C, respectively. The choice of hourly forecasting step makes it possible to compare the proposed method with some of the existing literature. Observe that in all numerical experiments of this paper, both the feature selection process and the training process of the forecasting engine are preformed using a set of training data, separate from the out-of-sample test data.

A. Evaluating the Feature Selection Component

The set of candidate inputs used in all numerical experiments of this paper is $S(t)$ shown in (1) with $N_{\rm WP} = N_{\rm WS} = N_{\rm WD} = N_{\rm T} = N_{\rm H} = 50$, except that prediction of wind direction or $\rm WD(\it t)$ is not considered, since its data are not available in the mentioned references. A large set of candidate inputs, including 253 features (50 lagged values of wind power, wind speed, wind direction, temperature, and humidity plus three forecast features of $\rm WS(\it t)$, $\rm T(\it t)$, and $\rm H(\it t)$), are considered such that no likely informative feature is missed. However, this set is too large for applying to a forecast engine and should be refined by a feature selection technique as described in Section II-A. Sample results of the proposed feature selection technique, including selected features among the set of candidate inputs $S(t)$ and rank of them, for the Alberta test case in April 10, 2007 are shown in Table I. Here, $\rm WP(\it t)$ indicates aggregated wind power of the three wind farms of this test case for hour t. The first stage of the feature selection technique (irrelevancy filter) filters out the irrelevant candidate inputs of $S(t)$ such that its output $S_1(t)$ includes 93 relevant features. Since this set is too large, it is not represented here. Then, the second stage (redundancy filter) filters out redundant candidate inputs of $S_1(t)$ and selects the subset $S_2(t)$ including 20 relevant and nonredundant features shown in Table I. Thus, the filtering ratio of the first stage, second stage, and whole proposed feature selection technique is $253/93 = 2.72$, $93/20 = 4.65$, and $253/20 = 2.72 \times 4.65 = 12.65$, respectively. The great filtering ratio of the proposed technique, i.e., 12.65, indicates its effectiveness to filter out irrelevant and redundant candidate inputs and select a minimum subset of the most informative features. The thresholds TH1 and TH2 of the technique for this test case are set to 0.68 and 0.91, respectively, by the cross-validation procedure of [17].

TABLE I
SELECTED FEATURES, I.E., $S_2(t)$, FOR THE ALBERTA TEST CASE ON APR. 10, 2007

Selected Feature	Rank	Selected Feature	Rank	Selected Feature	Rank
WP(*t*-1)	1	*WD*(*t*-2)	8	*WP*(*t*-7)	15
WS(*t*)	2	*T*(*t*-4)	9	*WP*(*t*-15)	16
WS(*t*-1)	3	*WP*(*t*-23)	10	*WD*(*t*-12)	17
WD(*t*-1)	4	*WP*(*t*-12)	11	*H*(*t*-1)	18
WP(*t*-3)	5	*T*(*t*-12)	12	*WP*(*t*-48)	19
T(*t*-1)	6	*WD*(*t*-4)	13	*T*(*t*-8)	20
WS(*t*-3)	7	*WS*(*t*-12)	14		

As seen from Table I, $S_2(t)$ includes seven lagged values of WP, indicating dependency of wind power time series on its historical values (auto-regression behavior). Wind speed and wind direction are important drivers for wind power and so eight features of WS and WD are seen among the selected features of $S_2(t)$. Some temperature and humidity features are also seen in $S_2(t)$, but with lower ranks with respect to WS and WD features. No specific periodic behavior is observed from the selected features as discussed in Section II-B1. Table I shows that most of high ranked selected features concentrate in the vicinity of the target forecast hour t (such as selection of previous hour or $t-1$ candidate inputs), representing the effect of short-run trend of wind power time series. We increased $N_{\rm WP}$, $N_{\rm WS}$, $N_{\rm WD}$, $N_{\rm T}$, and $N_{\rm H}$ from 50 and repeated the feature selection analysis. However, the selected features of Table I did not change, indicating low information value of far candidate inputs with high back shifts (from both wind power and exogenous variables) for wind power forecast of this test case. Finally, note that the selected features depend on the considered system. Moreover, $S_2(t)$ may change from a forecast day to another. Hence, it is better to separately perform the feature selection analysis for each test system and each forecast horizon (e.g., each forecast day).

To better illustrate the efficiency of the employed feature selection technique for wind power forecasting, the obtained prediction errors with the proposed technique and some other feature selection methods are shown in Table II. Forecast errors in Table II are in terms of root mean square error (RMSE) and normalized mean absolute error (NMAE) defined as follows:

$$\rm RMSE = \sqrt{\frac{1}{NH}\sum_{\it t=1}^{\rm NH}[WP(\it t) - WPF(\it t)]^2} \qquad (14)$$

$$\rm NMAE(\%) = \frac{1}{NH}\sum_{\it t=1}^{\rm NH}\frac{|WP(\it t) - WPF(\it t)|}{\it P_N}\times 100 \qquad (15)$$

where $\rm WP(\it t)$ and $\rm WPF(\it t)$ represent actual and forecast values of wind power in hour t, respectively; P_N is the nameplate capacity; NH indicates number of hours in the forecast horizon. Here, $\rm NH = 24$ for day ahead wind power forecast. In other words, the proposed wind power forecast strategy proceeds step by step (with hourly steps) along the forecast horizon until

TABLE II
RMSE (MWh) AND NMAE (%) RESULTS OF DIFFERENT FEATURE SELECTION METHODS FOR WIND GENERATION FORECAST OF THE ALBERTA TEST CASE

Test Week	NSA		PCA		CA		MI		Two-stage MI	
	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE
Feb.	6.72	2.69	6.03	2.32	6.02	2.45	4.34	2.33	4.18	2.12
May	7.20	2.80	6.87	2.71	6.68	2.64	4.68	2.49	4.68	2.26
Aug.	6.68	2.57	6.42	2.56	5.67	2.41	5.45	2.37	5.21	2.22
Nov.	7.53	3.02	6.52	2.55	5.43	2.47	4.39	2.35	4.31	2.25
Average	7.03	2.77	6.46	2.53	5.95	2.49	4.72	2.38	4.60	2.21

it reaches the end of the horizon. The length of the forecast horizon for the Alberta case study is 24 hours or 24 forecast steps. Four test weeks corresponding to four seasons of year 2007 (including the third weeks of February, May, August, and November) are considered for this numerical experiment indicated in the first column of Table II. This is to represent the whole year in the numerical experiments. The RMSE and NMAE values for each test week in Table II is the average of seven RMSE and NMAE values of its corresponding forecast days. Also, the average results of the four test weeks is shown in the last row of Table II. The benchmark methods of Table II include numerical sensitivity analysis (NSA), principal component analysis (PCA), correlation analysis (CA), and mutual information (MI) proposed for the feature selection of electricity load and price forecasts in the previous research works such as [17], [32], and [33]. CA is also used for wind power prediction in [1] and [2]. For the sake of a fair comparison, all feature selection techniques of Table II have the same training data (including 50 days previous to each forecast day), test data (the mentioned four test weeks), and proposed wind power forecast engine. Observe that the two-stage MI-based technique, presented in Section II-A, outperforms all other feature selection methods of Table II. The employed technique has both the lowest average RMSE and NMAE (the last row) and the lowest RMSE and NMAE of each test week. NSA, PCA, and CA are linear feature selection methods, while wind power is a nonlinear mapping function of its input variables. Thus, these methods may not correctly evaluate the impact of each candidate input on $\mathrm{WP}(t)$. MI is a nonlinear feature analysis method similar to the one employed here. However, neither MI nor NSA, PCA, and CA analyze the redundancy of candidate inputs performed in the second stage of the proposed technique.

The selected features for the Blue Canyon case study in a sample day (June 14, 2005) are shown in Table III. At first, observe that 21 features are selected for the Blue Canyon test case in Table III while 20 features are selected for the Alberta test case in Table I. Different thresholds TH1 and TH2 are set by the line search procedure for the feature selection technique in these two test cases leading to selection of different numbers of features. Moreover, comparing the selected features of Tables I and III, it is seen that 12 features are common between the two test cases, which are $\{\mathrm{WP}(t-1), \mathrm{WP}(t-12), \mathrm{WS}(t), \mathrm{WD}(t-1), \mathrm{WS}(t-1), \mathrm{WD}(t-2), \mathrm{T}(t-1), \mathrm{T}(t-4), \mathrm{H}(t-1), \mathrm{WS}(t-12), \mathrm{WP}(t-15), \mathrm{WP}(t-48)\}$. However, 9 features out of 21 selected features of Table III are different from those of Table I, which are $\{\mathrm{WP}(t-4), \mathrm{WS}(t-2), \mathrm{WS}(t-4), \mathrm{WS}(t-23), \mathrm{WP}(t-8), \mathrm{WP}(t-20), \mathrm{WP}(t-18), \mathrm{WP}(t-27), \mathrm{WS}(t-14),\}$. Thus, a significant difference $(9/21 \approx 43\%)$ is seen between the selected features of these two test cases. Besides, most of the 12 common features have different ranks in Tables I and III.

TABLE III
SELECTED FEATURES FOR THE BLUE CANYON TEST CASE ON JUNE 14, 2005

Selected Feature	Rank	Selected Feature	Rank	Selected Feature	Rank
WP(*t*-1)	1	*WS*(*t*-2)	8	*WP*(*t*-8)	15
WP(*t*-4)	2	*T*(*t*-1)	9	*WP*(*t*-15)	16
WP(*t*-12)	3	*T*(*t*-4)	10	*WP*(*t*-20)	17
WS(*t*)	4	*WS*(*t*-4)	11	*WP*(*t*-18)	18
WD(*t*-1)	5	*WS*(*t*-23)	12	*WP*(*t*-27)	19
WS(*t*-1)	6	*H*(*t*-1)	13	*WP*(*t*-48)	20
WD(*t*-2)	7	*WS*(*t*-12)	14	*WS*(*t*-14)	21

TABLE IV
SELECTED FEATURES FOR THE BLUE CANYON TEST CASE ON JUNE 13, 2005

Selected Feature	Rank	Selected Feature	Rank	Selected Feature	Rank
WP(*t*-1)	1	*WP*(*t*-16)	9	*WP*(*t*-29)	17
WP(*t*-4)	2	*T*(*t*-1)	10	*WS*(*t*-22)	18
WP(*t*-12)	3	*WS*(*t*-4)	11	*WP*(*t*-25)	19
WS(*t*)	4	*WP*(*t*-48)	12	*WP*(*t*-18)	20
WS(*t*-1)	5	*WS*(*t*-12)	13	*WP*(*t*-15)	21
WP(*t*-8)	6	*H*(*t*-1)	14	*WS*(*t*-48)	22
WD(*t*-1)	7	*WD*(*t*-4)	15	*WD*(*t*-12)	23
WP(*t*-23)	8	*T*(*t*-4)	16		

Selected features for the Blue Canyon test case on June 13, 2005 are shown in the newly added Table IV. Although some top ranked features are common between Tables III and IV (since they are highly relevant features for the forecast process), the other selected features of these tables have considerable differences. Accurately, 8 out of 23 features of Table IV are different from the features of Table III, which are $\{\mathrm{WP}(t-23), \mathrm{WP}(t-16), \mathrm{WD}(t-4), \mathrm{WP}(t-29), \mathrm{WS}(t-22), \mathrm{WP}(t-25), \mathrm{WS}(t-48), \mathrm{WD}(t-12)\}$. Thus, $8/23 \approx 35\%$ of the selected features for these two consecutive days are different. Moreover, some of the 15 common features have different ranks in Tables III and IV. These differences can be explained considering that the value of different candidate inputs for wind power forecasting may change with time, e.g., by changing atmospheric situation. For instance, in a stable atmospheric situation, more lagged candidate inputs with higher back-shifts may be informative compared with an unstable atmospheric situation for which closer candidate inputs in the vicinity of the forecast hour may be more appropriate. Thus, the rank of the candidate inputs and even the selected inputs for a wind power forecast process may change with time.

TABLE V
RMSE (MWh) AND NMAE (%) RESULTS OF HNN, MHNN, MHNN+GA, MHNN+PSO, MHNN+D,E AND MHNN+EPSO (PROPOSED FORECAST ENGINE) FOR WIND GENERATION FORECAST OF THE ALBERTA TEST CASE

Test Week	HNN		MHNN		MHNN+GA		MHNN+PSO		MHNN+DE		MHNN+EPSO	
	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE
Feb.	6.71	2.68	6.51	2.59	5.12	2.48	5.13	2.46	4.75	2.28	4.18	2.12
May	6.23	2.44	6.19	2.43	5.37	2.57	5.36	2.53	5.15	2.50	4.68	2.26
Aug.	6.55	2.57	6.48	2.54	5.69	2.51	5.65	2.67	5.53	2.70	5.21	2.22
Nov.	5.99	2.55	5.78	2.52	5.35	2.36	5.36	2.26	4.79	2.32	4.31	2.25
Average	6.37	2.56	6.24	2.52	5.38	2.48	5.37	2.47	5.06	2.45	4.60	2.21

TABLE VI
RMSE (MWh) AND NMAE (%) RESULTS OF RBF (AUXILIARY PREDICTOR), RBF+NN1, RBF+NN1+NN2, AND RBF+NN1+NN2+NN3 (PROPOSED FORECAST ENGINE) FOR WIND GENERATION FORECAST OF THE ALBERTA TEST CASE

Test Week	RBF		RBF+NN1		RBF+NN1+NN2		RBF+NN1+N2+NN3	
	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE
Feb.	13.24	5.34	7.44	3.41	6.75	3.33	4.18	2.12
May	13.17	5.38	7.32	3.38	6.44	3.28	4.68	2.26
Aug.	14.18	5.46	7.59	3.47	6.50	3.29	5.21	2.22
Nov.	13.22	5.33	7.76	3.57	6.83	3.48	4.31	2.25
Average	13.45	5.38	7.53	3.46	6.63	3.35	4.60	2.21

TABLE VII
RMSE (MWh) AND NMAE (%) RESULTS OF MULTIVARIATE ARIMA, MLP WITH LM, MLP WITH BFGS, MLP WITH BR AND PROPOSED FORECAST ENGINE FOR WIND GENERATION FORECAST OF THE ALBERTA TEST CASE

Test Week	Multivariate ARIMA		MLP with LM		MLP with BFGS		MLP with BR		Proposed (MHNN+EPSO)	
	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE	RMSE	NMAE
Feb.	15.32	6.24	8.49	3.65	13.62	5.43	11.14	4.56	4.18	2.12
May	16.56	7.02	9.18	3.89	11.64	4.88	12.68	5.02	4.68	2.26
Aug.	15.78	6.53	9.41	4.25	14.55	6.01	10.25	4.41	5.21	2.22
Nov.	14.84	6.11	9.32	4.12	12.75	4.99	13.59	5.97	4.31	2.25
Average	15.62	6.47	9.10	3.98	13.14	5.33	11.92	4.99	4.60	2.21

B. Evaluating the Performance of the Proposed Forecasting Engine

In this subsection, the effectiveness of the proposed EPSO, MHNN, and their combination as the wind power forecast engine is evaluated, respectively. For this purpose, the obtained RMSE and NMAE results from HNN, MHNN, MHNN+GA (genetic algorithm), MHNN+PSO, MHNN+DE (differential evolution), and MHNN+EPSO (proposed forecast engine) are shown in Table V. All numerical experiments of this subsection are performed on the same test weeks of Table II for day-ahead wind power forecasting. In the first benchmark method of Table V, both the proposed modifications are removed such that its forecast engine is the original HNN shown in Fig. 2. The forecast method of the second benchmark method is MHNN (including the auxiliary predictor of RBF) without the new EPSO module. In the next three benchmark methods, GA, PSO (classical PSO), and DE, which are three well-known stochastic search techniques, are added to MHNN instead of the EPSO component, respectively. Table V shows that the proposed forecast engine (MHNN+EPSO) outperforms all other alternatives. As seen, the first alternative, i.e., the original HNN, has the lowest accuracy in all test weeks. By adding the first modification (replacing the multivariate ARIMA with the RBF), we reach MHNN with a better prediction accuracy. By adding a stochastic search technique to MHNN, based on the hybridization approach of Section II-B3, much more improvement in the forecast accuracy is observed. As discussed before, this is mainly because the resulting hybrid forecast engine can better learn the input/output mapping function of wind power forecast process. Among the four hybrid forecast engines of Table V, MHNN+EPSO obtains the highest accuracy due to high exploration capability of EPSO to search the solution space and release the forecast engine from local minima. This numerical experiment validates both the proposed hybridization approach to combine MHNN with a stochastic search technique and the effectiveness of EPSO among different stochastic search techniques to enhance the accuracy of the suggested wind power forecast engine. Observe that all prediction methods presented in the numerical experiments of this subsection (such as those presented in Table V) have the same training data, test data and feature selection technique, since the purpose of these numerical experiments is comparison of the efficiency of different forecast engines.

In the next step, to validate the efficiency of the proposed structure for the forecast engine, another numerical experiment, reported in Table VI, is performed. In this table, obtained results from single RBF (auxiliary predictor shown in Fig. 3), RBF plus NN1 of MHNN, RBF plus NN1 and NN2 of MHNN, and RBF plus NN1, NN2, and NN3 of MHNN are represented. In this experiment, NN1, NN2, and NN3 are trained by LM+EPSO, BFGS+EPSO, and BR+EPSO as described in Section II-B3. Table VI shows that the wind power forecast errors improve step by step from the RBF to the proposed forecast engine (RBF+NN1+NN2+NN3). By adding each NN of MHNN to the forecast engine, its accuracy significantly enhances. For instance, average RMSE of the four test weeks reduces from 13.45 for RBF to 4.60 for the proposed forecast engine showing the great improvement of $(13.4 - 4.60)/13.45 \approx 66\%$.

In the last numerical experiment of this subsection, reported in Table VII, the proposed forecast engine (MHNN+EPSO)

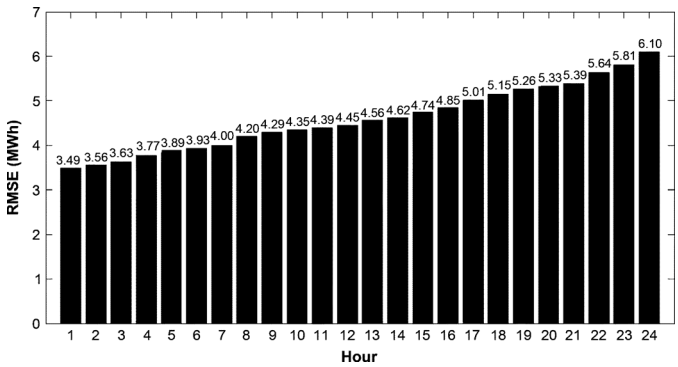

Fig. 4. RMSE for 24 hours of the forecast horizon of the Alberta case study.

is compared with multivariate ARIMA and three MLP neural networks trained by LM, BFGS, and BR learning algorithms, respectively. These four benchmark methods are well-known forecast techniques frequently used in the literature for prediction of electrical load demand, electricity price, and wind power. Table VII shows that the proposed forecast engine has significantly better wind generation forecast accuracy than the four other prediction methods. ARIMA is a linear predictor. The three NN based forecast methods of Table VII are nonlinear forecasters. However, they do not have the EPSO module in their training mechanisms, while the proposed forecast engine can benefit from its high global search ability to find better solutions in the solution space and avoid being trapped in local minima. Besides, the four other prediction methods of Table VII rely on single forecasters, whereas the proposed forecast engine combines different NNs, based on the data flow of MHNN, to enhance the obtained knowledge about the forecast process.

In Fig. 4, RMSE for each of 24 hours of the forecast horizon of the Alberta case study is shown. The RMSE shown for hour j in Fig. 4 $(1 \leq j \leq 24)$ is computed based on the forecast error of hour j in the 28 test days of the 4 test weeks of this case study as follows:

$$\mathrm{RMSE}_j = \sqrt{\frac{1}{28}\sum_{n=1}^{28}[\mathrm{WP}_{n,j} - \mathrm{WPF}_{n,j}]^2} \qquad (16)$$

where $\mathrm{WP}_{n,j}$ and $\mathrm{WPF}_{n,j}$ represent actual and forecast values of wind power in hour j of the nth test day $(1 \leq n \leq 28)$, respectively; RMSE_j is the RMSE value of hour j shown in Fig. 4. Observe from Fig. 4 that the hourly RMSE values increase from 3.49 MWh for hour 1 to 6.10 MWh for hour 24 due to the cumulative error effect. Despite the increasing behavior of the 24 hourly RMSE values of Fig. 4, even the worst RMSE of hour 24 has a still low value (6.10 MWh), which is, for instance, much better than the average RMSE of all other forecast methods of Table VII.

To also evaluate the effectiveness of the proposed wind power forecast strategy in a longer period, its obtained RMSE and NMAE results for four consecutive test months of May to August, 2007, are shown in Table VIII. This table shows that the error criteria for the four test months are consistent with those values presented in Tables V–VII for the four test weeks.

TABLE VIII
RESULTS OF THE PROPOSED STRATEGY FOR WIND GENERATION FORECAST OF THE ALBERTA TEST CASE IN MAY, JUNE, JULY, AND AUGUST 2007

Test Period	May	June	July	August	Whole four months
RMSE (MWh)	4.65	3.71	3.14	3.98	3.86
NMAE (%)	1.97	1.47	1.25	1.64	1.59

TABLE IX
NMAE (%) AND NRMSE (%) FOR WIND GENERATION PREDICTION OF THE BLUE CANYON WIND FARM WITH DIFFERENT FORECAST HORIZONS

Test conditions		Persistence method [2]	CA+BCD+SVR strategy [2]	Proposed strategy
Forecast horizon	Error criterion			
Hour ahead	NMAE	7.84	6.65	4.12
Hour ahead	NRMSE	11.93	10.54	7.52
24-hour ahead	NMAE	21.24	14.38	7.90
24-hour ahead	NRMSE	29.84	19.74	12.60
48-hour ahead	NMAE	25.42	15.73	10.51
48-hour ahead	NRMSE	34.81	21.24	16.58

C. Evaluating the Performance of the Complete Wind Power Forecasting Strategy

In the numerical experiment of Table IX, the whole proposed wind power forecast strategy, including the two-stage feature selection technique and the wind power forecast engine, is compared with the persistence method and an efficient wind generation prediction strategy presented in [2]. Persistence (also called the naïve predictor) is the benchmark method most frequently used for comparison in the wind power forecast literature [34]. In this method, the forecast for all future time intervals in the forecast horizon is set to the last measured value. The wind generation forecasting strategy of [2], shown by CA+BCD+SVR in Table IX, has CA as the feature selection part and a combination of Bayesian clustering by dynamics (BCD) and support vector regression (SVR) as the forecast engine. The results of persistence and CA+BCD+SVR methods are directly quoted from [2]. In this reference, the results of CA+BCD+SVR for different input variables are provided and its best obtained results are given in Table IX. For the sake of a fair comparison, the same test case (the Blue Canyon wind farm), the same test period (the month of June, 2005), and the same error criteria (NMAE and NRMSE) of [2] are also considered for the proposed wind power forecast strategy. NRMSE (normalized RMSE) is defined as follows:

$$\mathrm{NRMSE}(\%) = \sqrt{\frac{1}{\mathrm{NH}}\sum_{t=1}^{\mathrm{NH}}\left(\frac{\mathrm{WP}(t) - \mathrm{WPF}(t)}{P_N}\right)^2} \times 100 \qquad (17)$$

where $\mathrm{WP}(t)$, $\mathrm{WPF}(t)$, NH, and P_N are as defined for (15). NMAE and NRMSE are given for each method in Table IX for hour ahead, 24-hour ahead, and 48-hour ahead wind generation predictions. Observe that the proposed forecast strategy outperforms persistence and CA+BCD+SVR methods in all considered forecast horizons. Besides, for this test case (the Blue Canyon wind farm), CA+BCD+SVR in [2] has more detailed data obtained from seven weather stations, while we only have

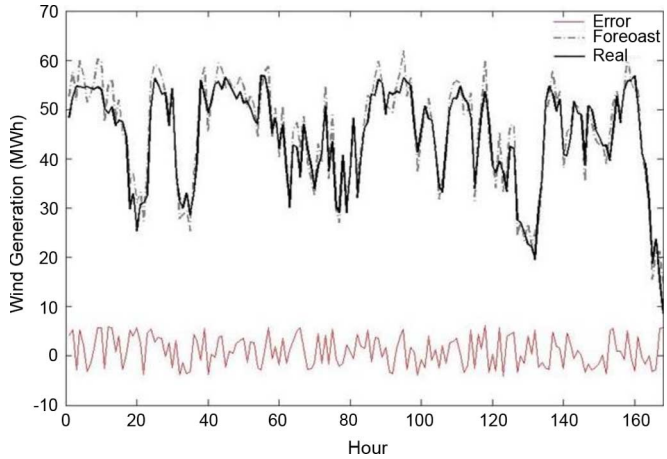

Fig. 5. Curves of real values, forecast values, and forecast errors $(\mathrm{WPF}(t)\mathrm{WP}(t))$ for the first test week of Tables II–VII.

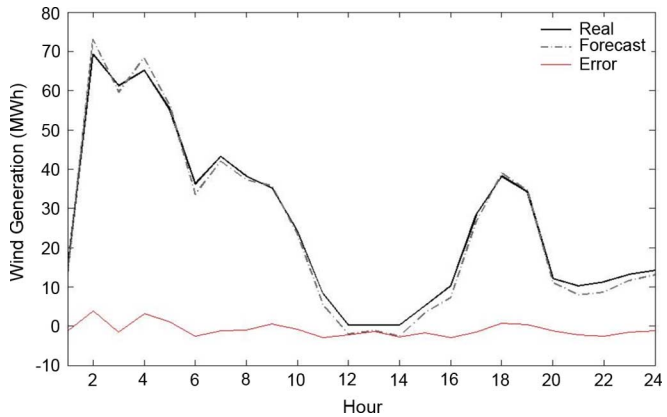

Fig. 6. Curves of real values, forecast values and forecast errors $(\mathrm{WPF}(t)\mathrm{WP}(t))$ for June 25, 2005 of the test set of Table IX.

access to publicly available weather data of one weather station [31] for the proposed forecast strategy. Despite less available data, both NMAE and NRMSE of the proposed forecast strategy are considerably lower than NMAE and NRMSE of CA+BCD+SVR for each forecast horizon in Table IX indicating efficiency of the proposed strategy for wind power forecasting.

To also give a graphical view about the wind power prediction accuracy of the proposed forecasting strategy, its obtained 24-hour ahead prediction results for the first test week of Tables II and V–VII (one wind farm) and June 25, 2005 of the test period of Table IX are shown in Figs. 5 and 6, respectively. From both figures, it is seen that the forecast curve accurately follows the trend of the real curve and only small errors are seen in the prediction of the proposed strategy.

Total computation time required for the setup of the proposed wind power forecast strategy including execution of the feature selection technique, training process of the wind power forecast engine, and adjustment of the adjustable parameters is about 20 min for the test cases of this paper. When the actual values of all 24 predicted hourly wind generations become available, the setup process of the proposed strategy is performed once a day considering the new data. After the setup process, the response time of the proposed strategy to produce wind power forecast of the next hour is very low and less than one second for our test cases, since it only requires the forward propagation of the NNs of the proposed forecast engine, which has very low computation burden. We do not repeat the setup process after each hour, since inclusion of only one new training sample in the historical data slightly changes the results of the setup process. Thus, if the user is interested in 24-hour-ahead wind power prediction, the setup process and generation of 24-hour -ahead forecasts can be performed in about 20 min. However, if the user is interested in one-hour-ahead wind power prediction, the forecast of the first hour is prepared in about 20 min and forecast of the next 23 hours is generated in less than one second. Hence, the setup time, measured on a simple hardware set of Pentium P4 3.2 GHz with 4-GB RAM, is acceptable within a day-ahead and even hour ahead decision making framework.

IV. Conclusion

In this paper, a new wind power forecast strategy is proposed. The proposed strategy is composed of an efficient two-stage feature selection technique and a novel forecasting engine. The presented feature selection technique can handle nonlinearities of a forecast process and filter out both irrelevant and redundant candidate inputs. Based on a minimum subset of the most informative selected features, the proposed forecasting engine implements the input/output mapping function of wind power forecast process. The proposed forecast engine is composed of MHNN and EPSO. MHNN, combining RBF and MLP networks, can capture both local and global behaviors of the target variable. By hybridizing MHNN with the enhanced stochastic search technique, i.e., the EPSO, the proposed forecasting engine can benefit from good global search ability of the EPSO, avoiding being trapped in local minima, in addition to high convergence rate of the NN learning algorithms. Obtained results from extensive testing of the employed feature selection technique, the proposed forecasting engine and the complete wind power forecasting strategy confirm the validity of the developed approach.

References

[1] K. Methaprayoon, C. Yingvivatanapong, W. J. Lee, and J. R. Liao, "An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty," *IEEE Trans. Ind. Appl.*, vol. 43, no. 6, pp. 1441–1448, Nov./Dec. 2007.

[2] S. Fan, J. R. Liao, R. Yokoyama, L. Chen, and W.-J. Lee, "Forecasting the wind generation using a two-stage network based on meteorological information," *IEEE Trans. Energy Convers.*, vol. 24, no. 2, pp. 474–482, Jun. 2009.

[3] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1657–1666, Nov. 2009.

[4] Y. K. Wu and J. S. Hong, "A literature review of wind forecasting technology in the world," in *Proc. IEEE Power Tech. Conf.*, Jul. 2007, pp. 504–509.

[5] N. Amjady, F. Keynia, and H. Zareipour, "A new hybrid iterative method for short-term wind speed forecasting," *Eur. Trans. Electrical Power*, DOI: 10.1002/etep.463.

[6] Alberta Electric System Operator (AESO), Wind Power Integration [Online]. Available: http://www.aeso.ca/gridoperations/13902.html

[7] F. Ö. Thordarson, H. Madsen, H. A. Nielsen, and P. Pinson, "Conditional weighted combination of wind power forecasts," *Wind Energy*, vol. 13, no. 8, pp. 751–763, Nov. 2010.

[8] G. Sideratos and N. Hatziargyriou, "Using radial basis neural networks to estimate wind power production," in *IEEE Power Engineering Society General Meeting*, 2007, DOI: 10.1109/PES.2007.385812.

[9] G. Giebel, The State-Of-The-Art in Short-Term Prediction of Wind Power a Literature Overview 2003 [Online]. Available: http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf

[10] I. J. Ramirez-Rosado, L. A. Fernandez-Jimenez, C. Monteiro, J. Sousa, and R. Bessa, "Comparison of two new short-term wind-power forecasting systems," *Renewable Energy*, vol. 34, no. 7, pp. 1848–1854, Jul. 2009.
[11] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, May 2009.
[12] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable Sustain. Energy Rev.*, vol. 12, no. 6, pp. 1725–1744, Aug. 2008.
[13] M. Lange and U. Fucken, *Physical Approach to Short-Term Wind Power Prediction*. Berlin, Heidelberg: Springer, 2006.
[14] U.S. Federal Energy Regulatory Commission (FERC), Assessing the state of wind energy in wholesale electricity markets: Comments of Basin Electric Power Co-operative FERC Docket AD04-13-000.
[15] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Trans. Energy Convers.*, vol. 24, no. 1, pp. 125–136, Mar. 2009.
[16] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[17] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 306–318, Feb. 2009.
[18] N. Amjady, A. Daraeepour, and F. Keynia, "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network," *IET Gener. Trans. Distrib.*, vol. 4, no. 3, pp. 432–444, Mar. 2010.
[19] N. Amjady and F. Keynia, "Electricity market price spike analysis by a hybrid data model and feature selection technique," *Electr. Power Syst. Res.*, vol. 80, no. 3, pp. 318–327, Mar. 2010.
[20] L. Zhang, P. B. Luh, and K. Kasiviswanathan, "Energy clearing price prediction and confidence interval estimation with cascaded neural network," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 99–105, Feb. 2003.
[21] J. J. Guo and P. B. Luh, "Improving market clearing price prediction by using a committee machine of neural networks," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1867–1876, Nov. 2004.
[22] A. M. Gonzalez, A. M. S. Roque, and J. G. Gonzalez, "Modeling and forecasting electricity prices with input/output hidden markov models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 13–24, May 2005.
[23] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Conf. Neural Netw.*, 1995, vol. 4, pp. 1942–1948.
[24] A. I. Selvakumar and K. Thanushkodi, "A new particle swarm optimization solution to nonconvex economic dispatch problems," *IEEE Trans. Power Syst*, vol. 22, no. 1, pp. 42–51, Feb. 2007.
[25] K. T. Chaturvedi, M. Pandit, and L. Srivastava, "Self-organizing hierarchical particle swarm optimization for nonconvex economic dispatch," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1079–1087, Aug. 2008.
[26] D. R. Hush and B. G. Horne, "Progress in supervised neural networks," *IEEE Signal Process.Mag.*, vol. 10, no. 1, pp. 8–39, Jan. 1993.
[27] J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electric Power Syst. Res.*, vol. 77, pp. 1297–1304, Aug. 2007.
[28] Alberta Wind Energy Corporation Aug. 2010 [Online]. Available: http://www.albertawindenergy.net/index.html
[29] Canada's National Climate Archive Aug. 2010 [Online]. Available: http://climate.weatheroffice.gc.ca/climateData
[30] Oklahoma Wind Power Initiative Aug. 2010 [Online]. Available: http://www.seic.okstate.edu/owpi
[31] National Climate Data Center Aug. 2010 [Online]. Available: http://www.ncdc.noaa.gov/oa/ncdc.html
[32] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. Hoboken, NJ: Wiley, 2002.
[33] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, Nov. 2007.
[34] ANEMOS Project [Online]. Available: http://www.anemos-project.eu

Nima Amjady (SM'09) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1992, 1994, and 1997, respectively.

At present, he is a full Professor with the Electrical Engineering Department, Semnan University, Semnan, Iran. He is also a Consultant with the National Dispatching Department of Iran. His research interests include forecast processes and control of power systems, operation of electricity markets, artificial intelligence, and data mining and its applications to the problems of power systems.

Farshid Keynia (M'09) received the B.S. and M.S. degrees in electrical engineering from Shahid Baahonar University, Kerman and Semnan University, Semnan, Iran, in 1996 and 2001, respectively. Currently he is working toward the Ph.D. degree in the Electrical Engineering Department, Semnan University.

His research interests are application of forecast methods, feature selection, and classification algorithms in the operations of deregulated electricity markets.

Hamidreza Zareipour (SM'09) received the B.Sc. degree in electrical engineering from K. N. Toosi University of Technology, Tehran, Iran, the M.Sc. degree in electrical engineering from Tabriz University, Tabriz, Iran, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1995, 1997, and 2006, respectively.

He currently is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB, Canada. His research focuses on economics, planning, and management of intelligent electric energy systems in a competitive electricity market environment.