Applied Energy 182 (2016) 80–93

Contents lists available at ScienceDirect

Applied Energy

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

Deep belief network based deterministic and probabilistic wind speed forecasting approach

H.Z. Wang, G.B. Wang *, G.Q. Li, J.C. Peng, Y.T. Liu

College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China

HIGHLIGHTS

- For the first time, deep belief network is designed for wind speed forecast (WSF).
- The nonlinear features in wind speed are used to improve forecast accuracy.
- The uncertainties of wind speed are evaluated using quantile regression.
- The competitive performance and high-stability of the proposed method were proved.

ARTICLE INFO

Article history:
Received 28 March 2016
Received in revised form 18 July 2016
Accepted 17 August 2016
Available online 25 August 2016

Keywords:
Deep belief network
Wind speed forecast
Deep learning
Quantile regression
Wavelet transform

ABSTRACT

With the rapid growth of wind power penetration into modern power grids, wind speed forecasting (WSF) plays an increasingly significant role in the planning and operation of electric power and energy systems. However, the wind speed time series always exhibits nonlinear and non-stationary features, making it very difficult to be predicted accurately. Recognizing this challenge, a novel deep learning based approach is proposed for deterministic and probabilistic WSF. The approach is a hybrid of wavelet transform (WT), deep belief network (DBN) and spine quantile regression (QR). WT is employed to decompose raw wind speed data into different frequency series with better behaviors. The nonlinear features and invariant structures of each frequency are completely extracted by layer-wise pre-training based DBN. Then, the uncertainties in wind speed are statistically synthesized via the QR method. Case studies using real wind farm data from China and Australia have been presented. The comparative results demonstrate that the high-level nonlinear and non-stationary feature in the wind speed series can be learned better, and competitive performance can thus be obtained. It is therefore convinced that the proposed method has a high potential for practical applications in electric power and energy systems.

1. Introduction

To mitigate climate change and reduce environmental pollution, new regulatory acts that encourage the use of renewable energy have been established worldwide [1]. Among renewable energy sources, wind power, as an alternative to fossil fuel-generated electricity, has been accepted as one of the most promising sources because it is clean, widely distributed, and emits no greenhouses gas during operation. Meanwhile, together with its mature and competitive technologies, wind power has gone through an unexpectedly high growth in recent years. It is, therefore, essential and desirable for high-accurate WSF to maximize the benefits of the high penetration of wind energy into electric power and energy systems [2] because the optimal solutions for almost every problem in power systems, e.g., the operating strategies, unit commitment, capacity planning and load balancing, are more or less associated with the accuracy of WSF [3]. However, the weather system always exhibits a chaotic nature, making it very difficult to achieve accurate and reliable WSF.

Traditionally, wind speed is predicted by deterministic forecast, short for deterministic point forecast. To date, four types of methodologies have been proposed in the literature for deterministic WSF, including the persistence method, physical modelling methods, statistical models, and soft-computing based approaches. The persistence method operates an assumption that any future wind speed value will be equal to the last known value due to the high autocorrelation shown in wind speed series [3]. Thus, the persistence method is generally applied as a benchmark for WSF to baseline the performance of newly developed forecasters [4]. Physical

* Corresponding author.
E-mail addresses: wanghz@szu.edu.cn (H.Z. Wang), wanggb@szu.edu.cn (G.B. Wang), ligq@szu.edu.cn (G.Q. Li), jcpeng@szu.edu.cn (J.C. Peng), liuyt@szu.edu.cn (Y.T. Liu).

http://dx.doi.org/10.1016/j.apenergy.2016.08.108
0306-2619/

methods manage to utilize several meteorological parameters, such as the temperature, pressure, and orography, to establish a mathematical model of WSF [5]. However, due to the high computational cost, these methods may not be suitable for practical real-time WSF in electric power and energy systems. Statistical models, such as Autoregressive Integrated Moving Average (ARIMA), grey predictor, and exponential smoothing, try to approximate the WSF model by using historical samples with error minimization [6,7]. In addition, with the development of soft-computing techniques, various new WSF methods, such as artificial neural network (ANN) [8], neuro-fuzzy inference system [9], and Kalman filtering [10], have been proposed. Generally, the soft-computing based methods always provide a more competitive performance than the other three types of methodologies because of their potential abilities for data-mining and feature-extracting [11].

Due to the volatile and erratic nature of weather systems, the individual methods mentioned above may not comprehensively represent the inner stochastic traits of wind speed series. As a remedy, many hybrid methods in combination with different forecast models have been mooted for deterministic point WSF. In [12], a novel hybrid approach based on ANN and fifth generation mesoscale model was proposed. The numerical results demonstrated that this method outperforms physical methods and pure ANN. In [13], a hybrid architecture based on an unscented Kalman filter and support vector regression was presented. In [14], a novel hybrid approach based on wavelet packet transform (WPT), particle swarm optimization (PSO), and fuzzy inference system was proposed to mitigate the volatility and intermittency in wind speed series, and its efficiency was demonstrated in wind farms in Portugal. In [15], a general framework in combination with extreme learning machine and PSO was developed to approximate the WSF model. In [16], a multi-layer feed-forward neural network method proposed to extract the stochastic feature in wind speed series was well-trained by the perturbation stochastic approximation algorithm. In [17], the first/second order adaptive coefficient was employed to establish the WSF model, and its forecast error was minimized by PSO. In [18], to achieve better performance, WPT and least square support vector machine were utilized to effectively alleviate the randomness and uncertainties exhibited in wind speed series.

However, deterministic forecasts, i.e., point forecasts, fail to estimate the uncertainties associated with a given prediction of WSF [4]. These uncertainties play a key role in improving the economic benefits of day-ahead energy bidding and reserve scheduling [4]. Therefore, probabilistic WSF that can describe the uncertainties involved in wind speed data has attracted much more attentions in recent years. Generally, probabilistic uncertainty in wind speed series can be effectively formulated using an ensemble of point forecasts. In [19], an ensemble of mixture density ANN was proposed for probabilistic WSF, and the uncertainties in model mis-specification were numerically evaluated. In [20,21], the uncertainties in wind power series were fully captured by modelling probabilistic forecasts as an ANN-based optimal prediction interval (PI) problem. In [22], the optimal PI was directly constructed via an extreme learning machine-based ensemble approach and was solved by PSO. The main demerit of ensemble-based probabilistic forecasts is their high computational cost, which may cause a problem for real-time implementation. Another mainstream method for obtaining probabilistic information is to extend the point forecast into a probabilistic form by using parametric method [23], non-parametric method [24] and quantile regression [25]. In [26], an adaptive sparse Bayesian model for probabilistic wind power forecast was proposed and its computational cost was effectively mitigated via a sparsification method. An overview of probabilistic WSF and wind power prediction was presented in [27].

However, the deterministic and probabilistic approaches for WSF presented above usually adopt shallow models as their core of learning principle. As noted in [28], the deep nonlinear features of data series may not be fully extracted by these shallow models. Furthermore, the widespread use of environmental sensors and other relevant technologies drives us into the era of big data, which makes it even harder to extract the deep variability and volatility features in wind speeds [29]. Therefore, the unsatisfactory feature mining of shallow models inspires us to rethink the WSF problem based on deep learning architecture.

Recently, deep learning, as a new branch of machine learning, has been growing rapidly and has been employed in a variety of fields, including classification tasks, data mining, dimensionality reduction and image processing [30]. Previous studies have proven that compared with the shallow models, deep learning can discover the inherent abstract features and hidden invariant structures in data from the lowest level to the highest level. Considering the wind speed series complicated in nature, deep learning can represent the inner structures and features without any prior knowledge. Therefore, the performance of deep learning exhibits superiority and higher accuracy in WSF problems, and the characteristics specific to feature extraction make deep learning much more attractive [31]. The classical deep learning algorithm mainly includes stacked auto-encoder (SAE), and DBN. Earlier implementation of SAE tailored for WSF was reported in [28,32]. Nevertheless, to date, DBN designed for WSF has not yet been considered in the published literature. Therefore, this work, which investigates a deep WSF framework and a hybrid intelligent approach based on WT, DBN and QR, is originally proposed to enhance WSF performance and prediction efficiency. The main contributions of this paper are presented as follows:

- For the first time, deep belief network is introduced and tailored to comprehensively extract the deep invariant structures and hidden high-level nonlinear features exhibited in any wind speed frequency.
- A novel deterministic WSF approach in combination with DBN and WT is proposed to mitigate the effects of invariant structure and nonlinearity features that exist in wind speed series on prediction accuracy.
- A probabilistic WSF framework is formulated based on WT, DBN and QR to accurately evaluate the randomness and uncertainty in wind speed series from the perspective of sharpness, reliability and overall skills.

The proposed hybrid deep architecture for WSF has been thoroughly tested and benchmarked on real wind speed data from China and Australia under various time-scales and operation scenarios.

2. Deep belief network

In this section, the DBN architecture is presented. DBN was initially invented by Hinton [33] and has been implemented successfully in feature learning, classification and collaborative filtering [30]. DBN mainly contains an unsupervised learning subpart using restricted Boltzmann machines (RBMs) as its building blocks and a logistic regression layer for prediction.

2.1. Restricted Boltzmann machine

Restricted Boltzmann machine is a stochastic neural network that can learn a distribution over its set of inputs. The network generally consists of one layer of binary-valued visible neurons and one layer of Boolean hidden units. In a RBM, no connections

exist between neurons in the same layer, but full connections exist between neurons in different layers. Meanwhile, these connections are bidirectional and symmetrical, as shown in Fig. 1.

RBM tries to learn a probability distribution from the visible layer to the hidden layer so that its configuration can exhibit desirable properties. The learning process is achieved via an energy function. Given the visible units v_i, hidden units h_j, and their connection weights W_{ij} (size $n_v \times n_h$), as well as the offset a_i for v_i and the bias weight b_j for h_j, the energy function $E(v, h)$ of a certain configuration can be defined as follows

$$E(v,h) = -\sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v}\sum_{j=1}^{n_h} h_j W_{j,i} v_i \quad (1)$$

Then, the probability distributions $P(v, h)$ over visible and hidden layers can be defined in terms of the energy function:

$$P(v,h) = e^{-E(v,h)}/Z \quad (2)$$

$$Z = \sum_{v}\sum_{h} e^{-E(v,h)} \quad (3)$$

where Z is the partition function.

Consequently, the individual activation probability of h_i given v, or of v_i given h, can be deduced as follows:

$$P(v_i = 1|h) = sigm\left(a_i + \sum_{j=1}^{n_h} W_{j,i} h_j\right) \quad (4)$$

$$P(h_j = 1|v) = sigm\left(b_j + \sum_{i=1}^{n_v} W_{j,i} v_i\right) \quad (5)$$

where *sigm* denotes the logistic sigmoid function.

With the aid of certain pre-training principles, the unknown RBM parameters, i.e., $\boldsymbol{W}$, $\boldsymbol{a}$ and $\boldsymbol{b}$, can be determined in a unsupervised manner, as discussed in Section 3.2.

2.2. Deep belief network

Stacking the binary RBM layer by layer hierarchically and adding a logistic regression to the end of the stacks creates a DBN.

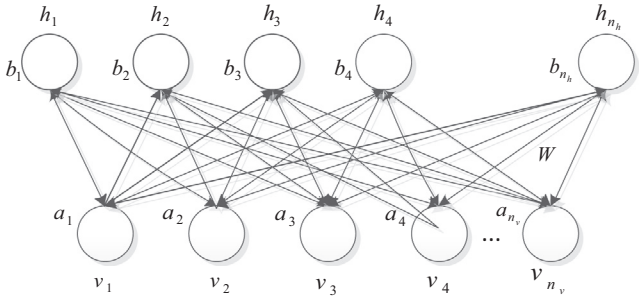

Fig. 1. Restricted Boltzmann machine.

More specifically, the first layer is pre-trained as an independent RBM, with training datasets as its inputs. Once the parameters of the first hidden layer are required, the output of the hidden layer found on the layer below is taken as the input of the hidden layer above, as illustrated in Fig. 2. Then, these two hidden layers can be considered to be a new RBM, and are thus trained in the same manner. Finally, a standard predictor, i.e., logistic regression, is added on the topmost layer and trained in a supervised manner to apply the DBN network for WSF. The training procedure of the predictor, termed as fine-tuning, is simply a back propagation or gradient descent algorithm, aiming to slightly adjust the parameters throughout the whole network [33].

3. DBN based hybrid WSF framework

A novel hybrid approach for deterministic and probabilistic WSF is originally proposed in this section to mitigate the effects of variability and volatility in wind speed series on prediction accuracy. This approach is a combination of the data filtering technique WT, layer-wise pre-training-based DBN and spine QR method. The details of each subpart are given below.

3.1. Wavelet decomposition

Raw wind speed series contain a variety of fluctuations, spikes and other types of nonlinearity and non-stationarity [4,34]. These features are among the attributes for deteriorating WSF accuracy. As presented in [35], WT can be used as an effective mathematical tool to decompose the wind speed signal into several frequencies with better behavior in terms of data variance and outliers. Thus, the impact of nonlinearity and non-stationarity in wind speed on the prediction accuracy can be appropriately mitigated.

WT can be defined in a discrete form for efficiency improvement, as follows

$$W(m,n) = 2^{-(m/2)}\sum_{t=0}^{T-1} f(t)\phi\left[(t - n2^m)/2^m\right] \quad (6)$$

where ϕ is the mother wavelet, m and n denote two integer variables that determine the parameters of scaling and translation of ϕ, t is the discrete time index, and T is the length of the signal $f(t)$.

In this research, a fast discrete WT algorithm termed the Mallat [36] is adopted. This algorithm is based on four filters: decomposition filters of low and high pass and reconstruction filters of low and high pass. Consequently, the original wind speed signal can be decomposed into one approximation (A_n) and several details (D_n) via a multi-level decomposition process based on Mallat's algorithm, as depicted in Fig. 3.

3.2. Layer-wise pre-training based DBN

Conventional NN adopts back-propagation (BP) as one of core of training principles, which may readily fall into a local optimum [28]. This drawback becomes apparent when the NN architecture

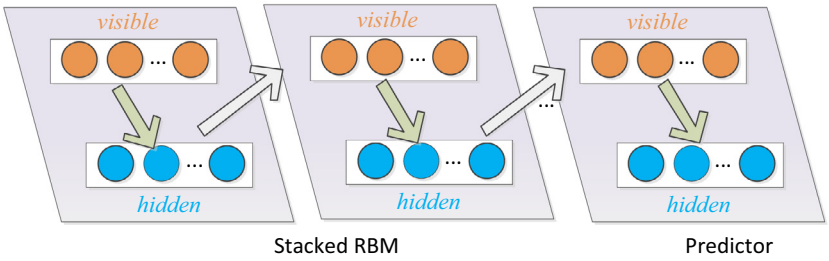

Fig. 2. DBN architecture.

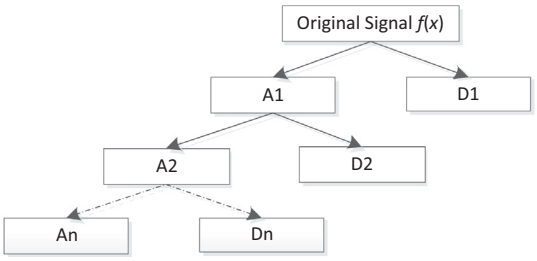

Fig. 3. Multilevel decomposition process.

goes deep because there is a large number of parameters to be optimized in this situation. One promising way to alleviate the local minimum dilemma of deep NN is to initialize the parameters to the highest possible degree [37]. As a result, the chance to find the global optimum increases substantially if the parameters are initialized very close to the optimal states in the search spaces [32]. Regarding DBN, the training process consists of a layer-wise pre-training process and a fine-tuning process. The former is used to provide good initial values for all parameters, while the latter is used to search the optimum based on the given initial states of the network.

Layer-wise pre-training method demands that the network should be pre-trained in a greedy layer-by-layer form, as depicted in Section 2.2. In other words, an independent RBM is pre-trained each time, and the related parameters, i.e., $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{W}$, can thus be obtained. The pre-training process is achieved by a stochastic gradient ascent on the objective function of RBM, defined as the log-likelihood of $P(v)$. Here, $P(v)$ is the probability of the visible vector over all of the hidden units, as follows

$$P(v) = \sum_{h} e^{-E(v,h)}/Z \tag{7}$$

Thus, the objective function takes shape as follows,

$$L_{\theta,S} = \sum_{v \in S} \log P(v,\theta) \tag{8}$$

where $\theta \in \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{W}\}$, and S is the training dataset.

According to the Bayesian statistics theory [33], maximizing the objective function (9) via stochastic gradient ascent algorithm generates a stable RBM with good initial states. The gradient ascent algorithm indicates that the parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{W}$ in RBM are updated based on the derivative of the objective function L, as follows,

$$\partial \log P(v)/\partial W_{i,j} = E_P[h_j v_i] - E_{\widehat{P}}[h_j v_i] \tag{9}$$

$$\partial \log P(v)/\partial a_i = v_i - E_{\widehat{P}}[v_i] \tag{10}$$

$$\partial \log P(v)/\partial b_j = E_P[h_j] - E_{\widehat{P}}[h_j] \tag{11}$$

where $E[\cdot]$ stands for expectation operation, and E_P and $E_{\widehat{P}}$ denote the original-data-driving and reconstructed-data-driving probabilities over v, respectively.

The first terms of (9)–(11) can be easily calculated using (4) and (5) on the training dataset. However, the computation process of the second terms in (9)–(11) is much more complicated since it is the expectation over distribution $\widehat{P}$ that is learned by the DBN systems. One feasible strategy is to apply alternating Gibbs sampling [33] on any stochastic states of the visible units until certain convergence criterion, such as k-steps, is satisfied. And thus, the expectations over $\widehat{P}$ can be estimated analytically. However, as presented in [38], the sampling strategy is unacceptable time-consuming, and so it is rarely implemented in real-life's task. As a remedy, a fast learning approach, termed contrastive divergence (CD), was proposed in [33]. This approach adopts two tricks to speed up the sampling process. One was to initialize the Markov chain with a training sample, and the other was to acquire samples after only k-steps of Gibbs sampling, termed CD-k. Experimental results demonstrate that CD performs satisfactorily for model recognition even $k = 1$.

In this research, CD-1 is adopted to estimate the expectations concerning $E_{\widehat{P}}$. Therefore, the update rule for parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{W}$ can be inferred from (9)–(11), as follows,

$$W^{t+1} = W^t + \eta(P(h|v^{(0)})[v^{(0)}]^T - P(h|v^{(1)})[v^{(1)}]^T) \tag{12}$$

$$a^{t+1} = a^t + \eta(v^{(0)} - v^{(1)}) \tag{13}$$

$$b^{t+1} = b^t + \eta(P(h|v^{(0)}) - P(h|v^{(1)})) \tag{14}$$

where the superscript t stands for time steps. η is the learning rate and 0.9 is adopted in this research.

In summary, the layer-wise pre-training process for DBN is illustrated in Fig. 4.

3.3. Supervised fine-tuning process

As depicted in Section 3.2, all of the parameters in DBN are appropriately initialized based on layer-wise pre-training method. These parameters are required to be slightly adjusted in a supervised manner until the loss function of DBN reaches its minimum. In this paper, BP is adopted for this type of task due to its effectiveness and efficiency. During the fine-tuning process, BP periodically works in a top-down manner. One period means that all of the parameters are updated one time, resulting in smaller forecasting errors. The errors are then back-propagated through the training set to re-correct the parameters of the DBN towards their optimal states. More details on the BP algorithm in DBN can be found in [39]. Therefore, after certain BP periods, the optimal states of all

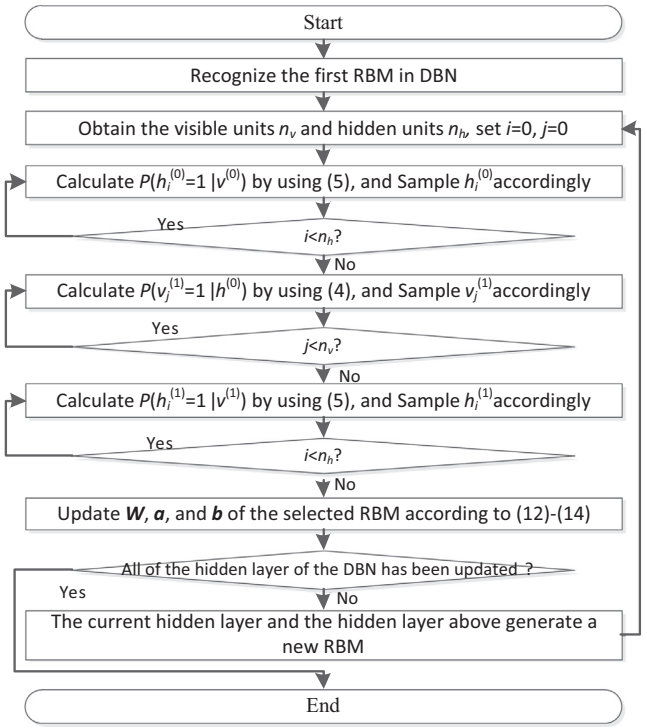

Fig. 4. The flow chart of the lay-wise pre-training for DBN.

of the parameters can be found, and thus, the training process of DBN has been completed.

3.4. Probabilistic WSF

Generally, conventional methods for modelling probabilistic WSF from a deterministic forecasting method are based upon a man-made hypothesis that the obtained errors follow a certain distribution, which may be incorrect. Thus, it can be beneficial to model the errors without any hypothesis. QR is one of such approaches, and has been discussed extensively in [25,40]. Roughly speaking, QR is based on a linear model,

$$y = \beta_0(\tau) + \beta_1(\tau)x_1 + \cdots + \beta_p(\tau)x_p + r = \hat{Q}(\tau;x) + r \tag{15}$$

where x is p known explanatory variable, and β is an unknown vector which is to be estimated based on x and τ. Given N samples, the vector β can be statistically determined,

$$\hat{\beta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{N} \rho_\tau[y_i - \hat{Q}(\tau;x)] \tag{16}$$

where ρ_τ is a piecewise linear loss function, described as follows,

$$\rho_\tau = \begin{cases} \tau r, & \text{if } r \geqslant 0 \\ (\tau - 1)r, & \text{if } r < 0 \end{cases} \tag{17}$$

Thereupon, the uncertainties of wind speed can be estimated as a set of forecast quantiles, and each quantile is modelled as an aggregation of a group of nonlinear smooth functions. Here, the smooth functions are approximated using cubic B-splines. Then, the forecast quantile with a proportion of the forecast errors can be established, as follows,

$$Q(\tau;f_e) = \beta_{0,t} + f_t(f_p,\tau) = \beta_{0,t} + \sum_{j=1}^{K-1} b_j(f_p)\beta_{j,t}(\tau) \tag{18}$$

where f_e, f_p denote the forecast error and result from the given deterministic WSF model, K and b are the parameters required in cubic B-spline basis function. In Section 5, the performance of the proposed probabilistic WSF has been benchmarked on real wind speed data collected from wind farm in China and Australia.

3.5. Implementation of the approach

The proposed deterministic and probabilistic WSF approach is a combination of WT, DBN and QR. The input of the hybrid approach is the wind speed data series. Using WT, the wind speed data are first decomposed into four frequencies, including one approximation and three details. Here, the 4th Daubechies function is selected as the mother wavelet because it provides a proper balance between wavelength and smoothness [14]. Then, an independent DBN is designed for each wind speed frequency. In this research, a 4-layer DBN was adopted to forecast the approximation signal. The input parameters of the 4-layer DBN are wind speed at the current time step WS_t and the previous 8 values $WS_{t-1}, \ldots, WS_{t-8}$. The numbers of the hidden neural in two hidden layers are 50 and 20. The output of the 4-layer DBN is the wind speed at the next time step WS_{t+1}. Similarly, for each detailed signal, a 10-layer DBN is properly designed. The number of the input neural is 20, and these of the hidden neurals in each layer are 65, 30, 45, 15, 60, 10, 20 and 15. These parameters are determined in a trial-and-error manner. The DBN for the detail signal is deeper than the DBN for approximation because the detail signal contains more spikes and fluctuations and thus requires more layers to forecast the uncertainties. All of the parameters, including $\boldsymbol{W}$, $\boldsymbol{a}$ and $\boldsymbol{b}$, are randomly initialized and updated based on layer-wise pre-training process and fine-tuning process until they all are converged. Consequently, aggregating the entire set of prediction frequencies via wavelet reconstruction and denormalization generates the final deterministic forecast result. Finally, the probabilistic WSF model is statistically established based on the spine QR, and the WSF uncertainties can thus be probabilistically represented as a set of quantiles ranging from 5% to 99%. The parameters in WT and QR adopt the default values in its MATLAB toolbox since the results are not sensitive to these parameters. The developed hybrid intelligent WSF approach is implemented using MATLAB, and the overall schematic diagram is illustrated in Fig. 5.

4. Performance criterion

To comprehensively assess the overall performance of the proposed deep approach, several errors for describing deterministic and probabilistic performance are employed, as follows.

4.1. Errors for deterministic performance

(1) *Mean Absolute Error (MAE):* This index is to assess the prediction capacity of the proposed algorithm,

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |WS_i^a - WS_i^f| \times 100\% \tag{19}$$

where WS_i^a is the true wind speed for the time step i, and WS_i^f represents the wind speed forecast.

(2) *Root Mean Square Error (RMSE):* This statistics is to measure the variety degree of the dataset,

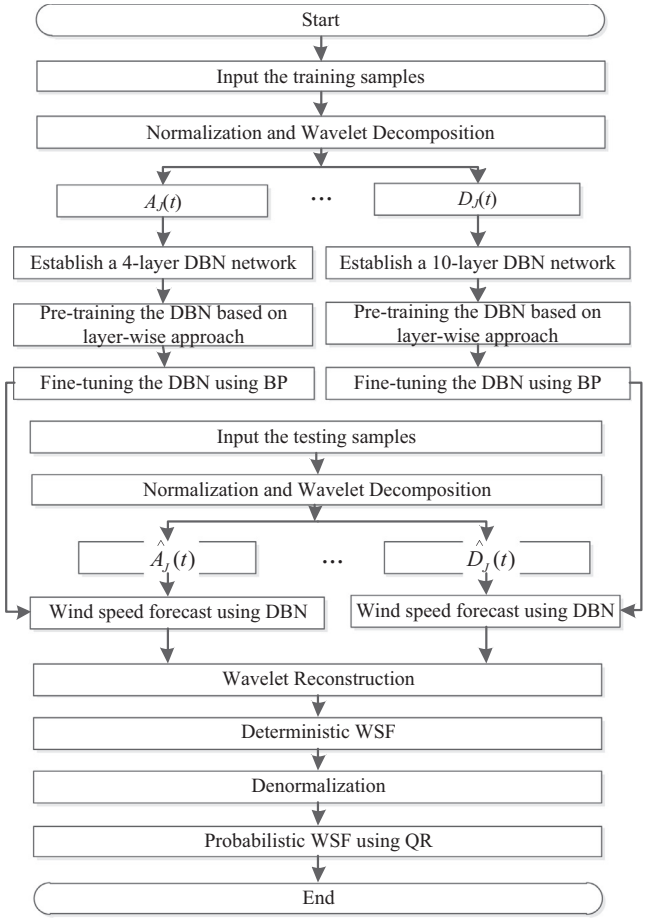

Fig. 5. Schematic diagram of the proposed deep framework for WSF.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(WS_i^a - WS_i^f\right)^2} \times 100\% \quad (20)$$

(3) *Mean Absolute Percentage Error (MAPE):* This is to quantify the forecast deviation of the proposed approach,

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|WS_i^a - WS_i^f|}{\overline{WS}_i^{a,N}} \times 100\% \quad (21)$$

where $\overline{WS}_i^{a,N}$ is the average of actual wind speed for the N^{th} hour.

4.2. Errors for probabilistic performance

(1) *Average Coverage Error (ACE):* Given the prediction interval nominal confidence (PINC [34]) $100(1-\alpha)\%$, this index measures how well the forecast quantiles match the observed values,

$$ACE = \frac{1}{N}\sum_{i=1}^{N} c_i \times 100\% - \text{PINC} \quad (22)$$

where c_i is the indicator of PI coverage probability and is defined as,

$$c_i = \begin{cases} 0, & WS_i^a \in I_i^a \\ 1, & WS_i^a \notin I_i^a \end{cases} \quad (23)$$

where I_i^α is the prediction interval for the time step i given PINC.

(2) *Interval Sharpness (IS):* This index is to comprehensively assess the sharpness of the PI, by rewarding the narrow PI and penalizing the wide one, as follows,

$$IS = \frac{1}{N}\sum_{i=1}^{N}\begin{cases} -2\alpha\delta_i^\alpha - 4[L_i^\alpha - WS_i^a], & \text{if } WS_i^a < L_i^\alpha \\ -2\alpha\delta_i^\alpha & \text{if } WS_i^a \in I_i^\alpha \\ -2\alpha\delta_i^\alpha - 4[WS_i^a - U_i^\alpha], & \text{if } WS_i^a > U_i^\alpha \end{cases} \quad (24)$$

where U_i^α, L_i^α are the upper and lower bound of PI, and δ_i^a is the width of the PI at time step i, and can be computed as $(U_i^\alpha - L_i^\alpha)$.

(3) *Continuous Rank Probability Score (CRPS):* This index simultaneously considers both reliability and sharpness, and is widely used for probabilistic forecasts [41]. Given the cumulative distribution function CDF_i and the measurement y_i over the testing sample, the average value of CRPS can be calculated as [42]

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{y=0}^{\infty}[CDF_i - H(y - y_i)]^2 dy \quad (25)$$

where $H(y-y_i)$ represents indicative function. The value of $H(y-y_i)$ is 0 if $y < y_i$, otherwise, the value is 1.

5. Case studies

5.1. Experimental data description

In this paper, the proposed hybrid approach based on WT, DBN and QR has been comprehensively tested and benchmarked on wind speed datasets at two wind farms, the Shangchuan Island wind farm (SIWF) in Guangdong Province, China, and the Cathedral Rocks wind farm (CRWF) in Australia [22]. The wind speed data collected from SIWF covers the period from January 2013 to December 2013 with a one-hour resolution. Datasets from CRWF were recorded in 5-min intervals for the whole year of 2011, and twelve 5-min intervals over one hour were averaged to obtain the hourly wind speed data used in this paper. For each dataset, four different prediction models with the same architecture are adopted to fulfill the seasonal forecast task, as detailed in Section 3.5 because the weather conditions and the wind speed data vary significantly in different seasons. In each season, the wind speed data are divided into the training dataset and testing dataset. The training dataset covers the days from the 1st to the 25th of each month, and the rest of the data constitutes the testing dataset. The WSF models are well-trained using the training dataset to completely explore the deep nonlinear and non-stationary features exhibited in the training dataset. The testing dataset is used to test the performance of the proposed approach. To fully validate the effectiveness of the proposed algorithm, the results are compared with the Auto-Regressive and Moving Average Model (ARMA), the well-tuned Back-propagation Neural Network (BPNN), and the Morlet Wavelet Neural Network (MWNN).

5.2. Deterministic forecast results

Considering SIWF, the 1-h ahead seasonal deterministic prediction results obtained from ARMA, BPNN, MWNN and the proposed hybrid approach are presented in Table 1. The MAE index obtained from the proposed algorithm varies from a low of 0.3415 to a high of 0.5155, with an average of 0.4282. Compared to ARMA, BPNN, and MWNN, the MAE index has been averagely improved by 44.18%, 48.99%, and 48.47%, respectively. Similarly, the RMSE has been evenly improved by 45.84%, 50.13%, and 48.47%, respectively, and MAPE by 42.48%, 46.84%, and 46.97%, respectively. Moreover, to graphically demonstrate the privilege of the proposed approach, Figs. 6 and 7 present the estimated wind speed using three benchmarks and the measured wind speed. From these two figures, it is clear that the results from the proposed approach and the real values almost overlap, which indicates that the estimated values are closest to the real data. Therefore, the comparative results demonstrate that the proposed hybrid algorithm exhibits the best point forecast performances in all four seasons and thus shows the best prediction capability over the benchmarks. In addition, the results also show that the ARMA outperforms the BPNN and MWNN in 1-h ahead forecasts, which is consistent with the results presented in [28]. This result is due to the high non-linearity, complexity, and non-smoothness exhibited in short-time wind speed series. These dynamics cannot be extracted effectively by shallow NN models, such as BPNN and MWNN.

The seasonal 1-h ahead forecasting performance considering CRWF is presented in Table 2 and intuitively shown in Figs. 8 and 9. The proposed hybrid algorithm generates the least MAE,

Table 1
Deterministic 1-h ahead forecasting error at SIWF.

Season	Error	ARMA	BPNN	MWNN	Proposed method
Spring	MAE	0.6183	0.6330	0.6869	**0.3609**
	RMSE	0.8009	0.8152	0.8549	**0.4771**
	MAPE	11.02%	11.28%	12.24%	**6.43%**
Summer	MAE	0.6719	0.7850	0.7392	**0.3415**
	RMSE	0.9328	1.0685	0.9647	**0.4391**
	MAPE	10.94%	12.78%	12.03%	**5.56%**
Fall	MAE	1.0606	1.2232	1.1396	**0.4947**
	RMSE	1.3935	1.5917	1.5143	**0.6671**
	MAPE	10.20%	11.77%	10.96%	**4.76%**
Winter	MAE	0.7176	0.7166	0.7583	**0.5155**
	RMSE	0.9304	0.9309	0.9722	**0.6142**
	MAPE	12.27%	12.25%	12.96%	**8.81%**
Average	MAE	0.7671	0.8394	0.8310	**0.4282**
	RMSE	1.0144	1.1016	1.0765	**0.5494**
	MAPE	11.11%	12.02%	12.05%	**6.39%**

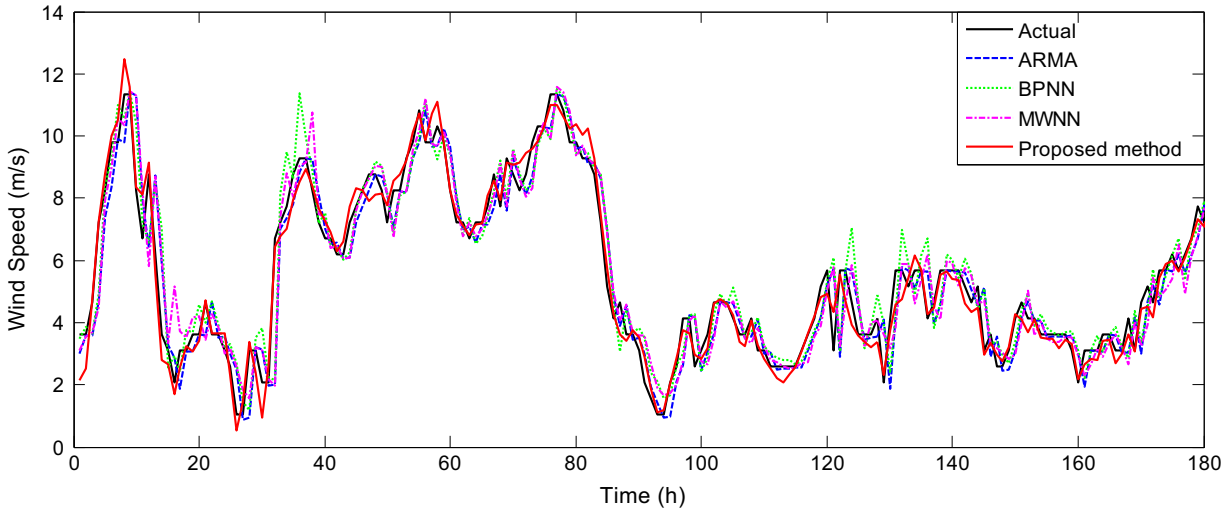

Fig. 6. 1-h ahead WSF results in summer 2013 at SIWF.

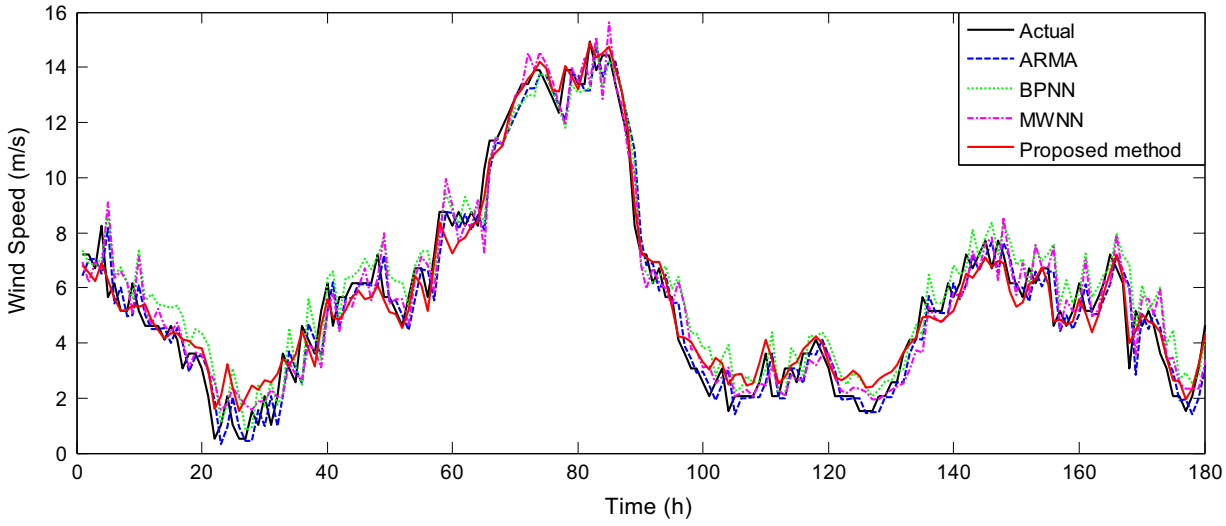

Fig. 7. 1-h ahead WSF results in winter 2013 at SIWF.

RMSE and MAPE errors in all four seasons, which further reveals the robustness of the proposed WT+DBN model. The high-accuracy and superiority of the hybrid approach mainly come from the deep NN architecture, which provides an effective way to approximate the inherent invariant features and structures. Thus, the high-level non-linear, non-stationary and non-smoothness features that are exhibited in the wind speed datasets can be better extracted. Moreover, with reference to Tables 1 and 2, it can be found that the deterministic point prediction errors, i.e., MAE, RMSE and MAPE, are slightly worse in CRWF than these in SIWF, which is understandable because the weather environments in CRWF are relatively more chaotic and thus more unpredictable.

The performance of the soft-computing based prediction model is generally well-known to change from one simulation to another because of the variability and mutability of the model parameters. Therefore, to make the results and conclusions convincible, the simulations above were repeated 20 times for each case. Fig. 10 presents the related statistics of MAPE in spring. The MAPE in two datasets obtained from the proposed hybrid approach remains unchanged over time because once the deep network topology is determined, the model parameters in each simulation stay global-optimal, and therefore will not be changed anymore.

Table 2
Deterministic 1-h ahead forecasting error at CRWF.

Season	Error	ARMA	BPNN	MWNN	Proposed method
Spring	MAE	0.7957	0.8426	0.9086	**0.4097**
	RMSE	1.0170	1.0834	1.1395	**0.5510**
	MAPE	11.94%	12.65%	13.64%	**6.15%**
Summer	MAE	0.6714	0.9036	0.7426	**0.3987**
	RMSE	0.9251	1.1662	0.9799	**0.5200**
	MAPE	10.99%	14.79%	12.15%	**6.53%**
Fall	MAE	0.7323	0.8062	0.8584	**0.5397**
	RMSE	0.9748	1.0424	1.1083	**0.7061**
	MAPE	7.18%	7.90%	8.41%	**5.29%**
Winter	MAE	1.0526	1.3597	1.1552	**0.6222**
	RMSE	1.4675	1.8815	1.5467	**0.7829**
	MAPE	17.27%	22.30%	18.95%	**10.21%**
Average	MAE	0.8130	0.9780	0.9162	**0.4926**
	RMSE	1.0961	1.2934	1.1936	**0.6400**
	MAPE	11.84%	14.41%	13.29%	**7.04%**

Considering SIWF, the MAPE from BPNN varies from a minimum of 11.53% to a maximum of 18.75%, with an average of 13.22% and a variance of 3.13%. The MAPE obtained from MWNN varies

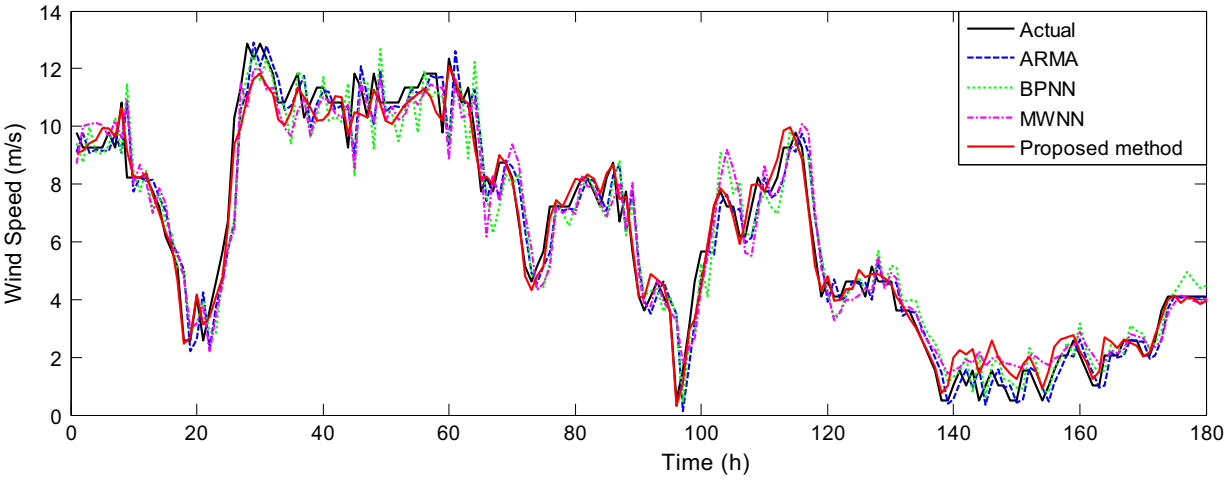

Fig. 8. 1-h ahead WSF results in summer 2011 at CRWF.

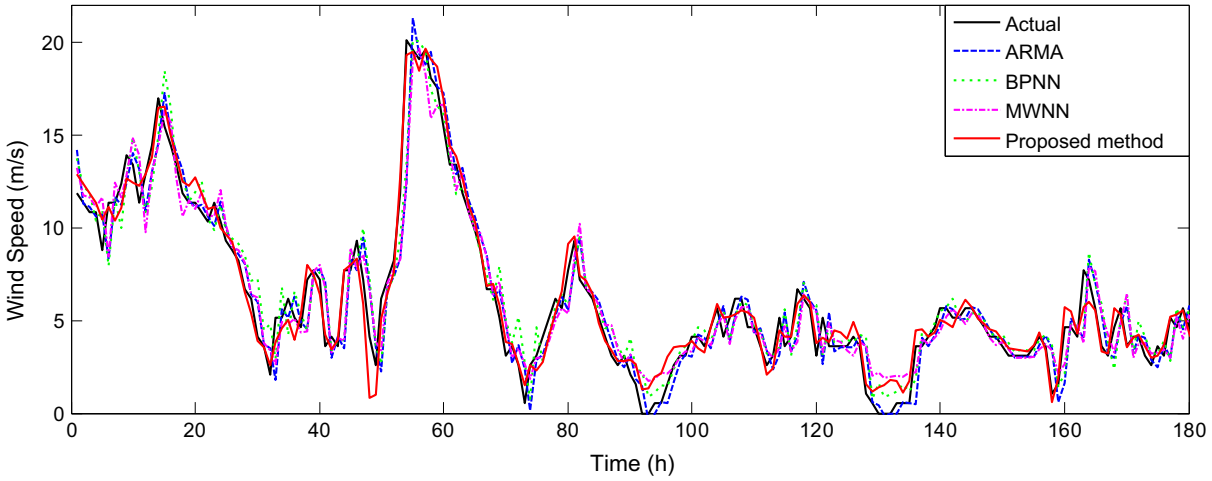

Fig. 9. 1-h ahead WSF results in winter 2011 at CRWF.

from 11.37% to 14.45%, with an average of 12.54% and a variance of 0.71%. Similarly, regarding CRWF, the MAPE index from BPNN has an average of 13.01% and a variance of 0.69%, and the MAPE index from MWNN exhibits an average of 13.33% and a variance of 2.62%. The fluctuations of MAPE show that the model parameters of BPNN and MWNN change from one simulation to another, i.e., they are local optimum and thus exhibit large instability. Therefore, the results presented demonstrate that the proposed method can not only significantly improve the prediction performance, but also remain high-stability. Thus the superiority and potentiality of the proposed approach are further illustrated.

To further investigate the effectiveness of the proposed WT +DBN approach, Figs. 11 and 12 sketch the annual average of MAPE in terms of various forecast horizons. The forecast horizon in SIWF varies from 1-h ahead to 8-h ahead due to the 1-h data resolution, while that in CRWF varies from 15-min ahead to 8-h ahead. The operational time-span of the electric power dispatch in which the deterministic point prediction performance can be used accounts for the reason for the selection of these prediction horizons. From the results, it is clear that the MAPE index obtained from the proposed approach is significantly better than the benchmarks in all of the prediction horizons, which confirms the high efficiency and capability of the proposed algorithm to solve short-term WSF problems. In addition, as the prediction hour increases, the MAPE index generally deteriorates significantly. This numerical performance is expected due to the higher uncertainty and randomness exhibited in long-term wind speed data series. Moreover, considering CRWF, the value of MAPE for 15-min ahead is slightly worse than the value of MAPE for 30-min ahead. This difference can be explained by the chaotic nature that is found in very short-term wind speed. Therefore, from the above results, it is appropriate to conclude that the proposed deterministic WT + DBN model exhibits the most stable and robust structure, and has the best generalization properties. We can therefore conclude that among the benchmarks, the proposed approach is the best model to fulfill the WSF tasks with high forecast accuracy.

5.3. Probabilistic forecast results

This section evaluates the effectiveness of the proposed probabilistic WT + DBN + QR approach. According to Section 4.1, three indices, i.e., ACE, IS, and CRPS, are used in the literature to measure the probabilistic performance. Regarding ACE and IS, high-confidence levels of PINC $100(1-\alpha)\%$ ranging from 90% to 99% are generally considered because of the high reliability required in power system optimization and operation. These two probabilistic indices on two wind farms as well as their corresponding PICP are tabulated in Tables 3 and 4, respectively. In all four seasons,

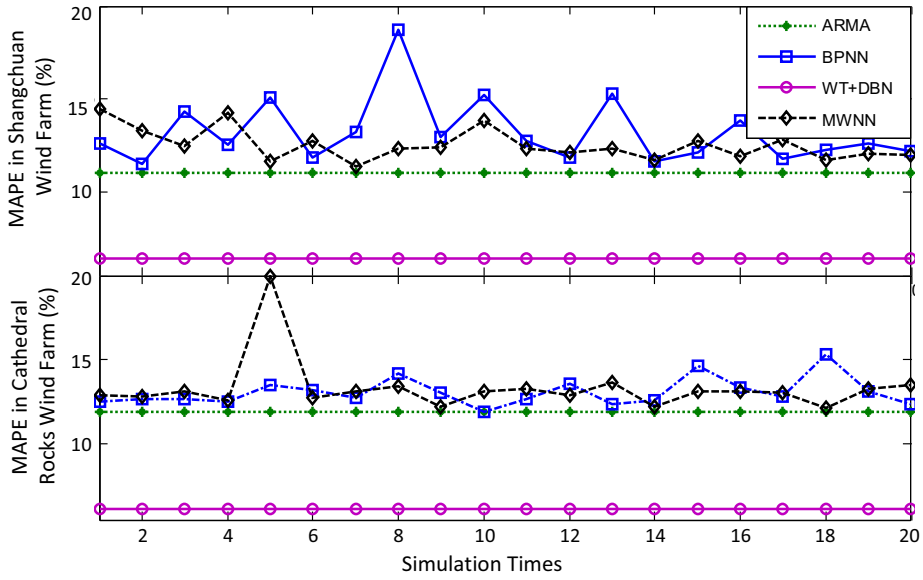

Fig. 10. MAPE statistics in two wind farms at spring season.

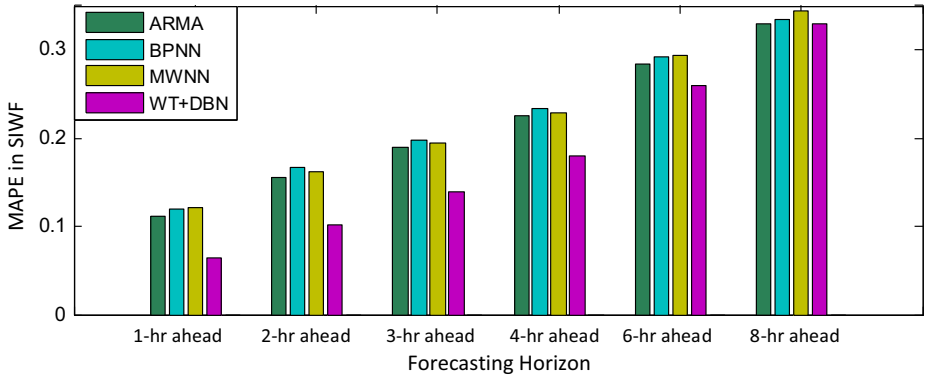

Fig. 11. MAPE statistics in terms of various forecasting horizons at SIWF.

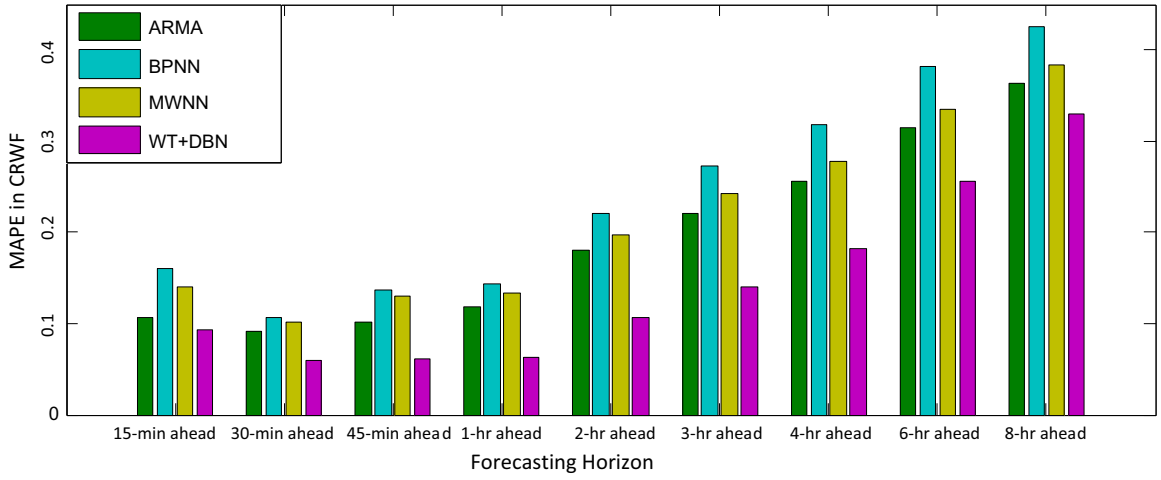

Fig. 12. MAPE statistics in terms of various forecasting horizons at CRWF.

the ACE of the QR estimator from the proposed WSF approach exhibits the lowest deviations to the nominal confidence levels, especially for the case of higher confidence levels of 95% and 99%. Quantitatively, considering SIWF, the average of ACE using the proposed approach is 0.6742%, which has been improved by 23.96%, 26.79%, and 26.12% compared to ARMA, BPNN and MWNN. Considering CRWF, the average value using the proposed approach is 0.5958% and the corresponding improvements are 41.44%,

36.16% and 43.75%. These results prove that the reliability errors between the observed probability and nominal confidence from the proposed approach are at a minimum, and further indicate that the proposed approach exhibits higher prediction reliability. Moreover, the average value of ACE in SIWF is better than the average of ACE in CRWF due to the weather conditions, consistent with the observations from Tables 1 and 2. In addition, the QR estimator obtained from ARMA outperforms the QR estimator from BPNN and MWNN in all cases because the deterministic results from ARMA exhibit better performance than that from the other two benchmarks.

From the perspective of sharpness, the IS of the quantiles obtained from the proposed approach ranges from −0.53 to −0.05 in SIWF, with an average of −0.23. Obviously, the averaged IS is the best among the benchmarks. In addition, the IS using the proposed approach in CRWF ranges from −0.62 to −0.05, with an average of −0.29. Concretely, the sharpness of the proposed approach are improved by 46.29%, 49.86% and 49.20% compared to ARMA, BPNN and MWNN, respectively. Hence, from the numerical results, we can easily conclude that the proposed approach exhibits the most desirable performance for probabilistic WSF. Moreover, with reference to Tables 3 and 4, the IS performance in SIWF is reduced by 11.66% compared to the IS in CRWF. This numerical result is consistent with the conclusions from the deterministic point prediction results and therefore is expected because of the less chaotic weather conditions in SIWF.

Figs. 13–16 graphically illustrates the constructed PIs with PINC 90% obtained by the proposed approach as well as the corresponding actual measured wind speeds. It can be seen that the measured data are within the established lower and upper bounds for a great percentage. Meanwhile, the shapes of the three lines in each graph are very similar to each other. The high PI coverage and similarity of the lines indicate that the probabilistic performance criteria for the samples are very satisfactory using the proposed approach. In addition, it is clear that the line trends in each graph have strong fluctuations, which show the nonlinear and non-stationary characteristics in wind speed in different seasons and regions. Furthermore, the PIs for SIWF are narrower compared to the PIs for CRWF because, as mentioned above, of the more irregular patterns and uncertainties exhibited in the wind speed data for SIWF. Thus, the proposed approach has the ability to construct high-quality PIs for a wind speed dataset under various confidence levels.

To further demonstrate the effectiveness and feasibility of the proposed probabilistic WSF approach, a series of prediction simulations using different time resolutions has been carried out and the numerical results are tabulated in Table 5. Here, the results of ARMA are presented because of its relatively superior performance in the above simulations. The time resolutions for SIWF cover from 1-h to 8-h with a step of 1 h, and the time resolutions for CRWF are from 5-min to 2-h. Moreover, the training dataset of the simulations using different time resolutions are obtained by interval sampling on the original wind speed dataset. In Table 5, the performances for probabilistic WSF are evaluated in terms of CRPS, which is an overall scoring rule that addresses reliability and sharpness simultaneously and is generally accepted as a proper assessment index in the published literature [42]. Statistically, the CRPS for SIWF obtained from the proposed approach ranges from 0.2293 to 0.9460 in spring, from 0.2083 to 0.9627 in summer, from 0.3324 to 1.4536 in fall, and from 0.3309 to 1.4097 in winter. Compared to CRPS from ARMA, the improvement grows from 20.46% to 44.47% in spring, from 22.04% to 53.67% in summer, from 7.71% to 53.68% in fall, and from 1.69% to 29.91% in winter. The CRPS for CRWF from the proposed approach has a minimum of 0.1892 and a maximum of 1.2133 for the whole year, which obviously outperform ARMA a lot. Therefore, it is clear from these numerical results that the proposed approach outperforms ARMA in terms of seasons and time resolutions.

ACE, IS and CRPS are three typical metrics that are used to evaluate the effectiveness and are satisfactory for probabilistic prediction. Based on the above results, it is evident that the proposed probabilistic WSF approach outperforms the three benchmarks not only from the viewpoint of reliability and sharpness but also from the perspective of overall skills.

5.4. Practical application of the proposed approach

The in-depth numerical results demonstrate that the proposed WSF approach exhibits a more accurate and reliable performance for short-term deterministic and probabilistic WSF in comparison with the three benchmarks of the state of the art. More accurate WSF prediction results mean fewer uncertainties and fluctuations on the estimation of wind power, which is significant for the operation and planning of electric power and energy systems. For instance, deterministic minutes-ahead WSF prediction results with high-accuracy can be implemented in wind turbine control [43] to effectively reduce the response time of the maximum power point tracking (MPPT) controller. Also for example, probabilistic 1-h ahead WSF results can be employed to quantify the impacts of wind speed uncertainties on total generation cost in economic dispatch [44]. In this process, four steps are required. Firstly, the obtained wind speed quantiles are transformed to the corresponding probability density function (PDF). Then, the PDF for wind power is calculated by applying the aerodynamic behavior of a wind turbine on the PDF of the wind speed. Then, the impact of not using the available wind power on the system generation cost and the impact of overestimating wind power on the generation

Table 3
Probabilistic 1-h ahead forecasting error at SIWF.

Season	PINC (%)	ARMA			BPNN			MWNN			Proposed method		
		PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS
Spring	90	90.48	0.48	−0.67	90.48	0.48	−0.67	90.48	0.48	−0.73	90.31	**0.31**	**−0.34**
	95	96.19	1.19	−0.39	96.19	1.19	−0.38	95.71	0.71	−0.42	95.24	**0.24**	**−0.19**
	99	99.52	0.52	−0.09	100.00	1.00	−0.09	99.52	0.52	−0.10	99.52	**0.52**	**−0.05**
Summer	90	90.99	0.99	−0.82	91.42	1.42	−0.90	91.42	1.42	−0.83	90.99	**0.99**	**−0.34**
	95	96.14	1.14	−0.48	95.71	0.71	−0.51	96.57	1.57	−0.49	95.71	**0.71**	**−0.20**
	99	99.57	0.57	−0.11	100.00	1.00	−0.12	99.57	0.57	−0.12	99.57	**0.57**	**−0.05**
Fall	90	90.64	0.64	−1.17	90.64	0.64	−1.33	90.64	0.64	−1.25	90.64	**0.64**	**−0.53**
	95	95.74	0.74	−0.70	95.74	0.74	−0.75	96.17	1.17	−0.73	95.74	**0.74**	**−0.31**
	99	100.00	1.00	−0.16	100.00	1.00	−0.17	100.00	1.00	−0.16	100.00	**1.00**	**−0.07**
Winter	90	90.91	0.91	−0.79	90.91	0.91	−0.82	90.91	0.91	−0.82	90.41	**0.41**	**−0.43**
	95	96.46	1.46	−0.45	95.96	0.96	−0.47	95.96	0.96	−0.47	95.96	**0.96**	**−0.23**
	99	100.00	1.00	−0.10	100.00	1.00	−0.11	100.00	1.00	−0.11	100.00	**1.00**	**−0.05**

Table 4
Probabilistic 1-h ahead forecasting error at CRWF.

Season	PINC (%)	ARMA			BPNN			MWNN			Proposed method		
		PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS	PICP (%)	ACE (%)	IS
Spring	90	91.00	1.00	−0.86	91.00	1.00	−0.95	91.00	1.00	−0.90	90.35	**0.35**	**−0.46**
	95	95.73	0.73	−0.48	95.73	0.73	−0.54	96.21	1.21	−0.50	95.21	**0.21**	**−0.27**
	99	100.00	1.00	−0.12	100.00	1.00	−0.13	100.00	1.00	−0.11	100.00	**1.00**	**−0.06**
Summer	90	91.62	1.62	−0.81	91.10	1.10	−0.88	91.10	1.10	−0.84	90.58	**0.58**	**−0.39**
	95	95.81	0.81	−0.49	96.34	1.34	−0.49	96.34	1.34	−0.48	95.81	**0.81**	**−0.22**
	99	99.48	0.48	−0.13	100.00	1.00	−0.11	100.00	1.00	−0.12	99.42	**0.42**	**−0.05**
Fall	90	91.09	1.09	−0.80	91.09	1.09	−0.82	91.09	1.09	−0.89	91.09	**1.09**	**−0.56**
	95	96.04	1.04	−0.46	96.04	1.04	−0.48	96.04	1.04	−0.50	95.54	**0.54**	**−0.33**
	99	100.00	1.00	−0.15	99.50	0.50	−0.16	99.50	0.50	−0.15	99.50	**0.50**	**−0.08**
Winter	90	91.10	1.10	−1.27	90.58	0.58	−1.38	91.62	1.62	−1.36	90.57	**0.57**	**−0.62**
	95	96.34	1.34	−0.74	96.34	1.34	−0.81	95.81	0.81	−0.81	95.66	**0.66**	**−0.35**
	99	100.00	1.00	−0.17	99.48	0.48	−0.19	100.00	1.00	−0.19	99.42	**0.42**	**−0.09**

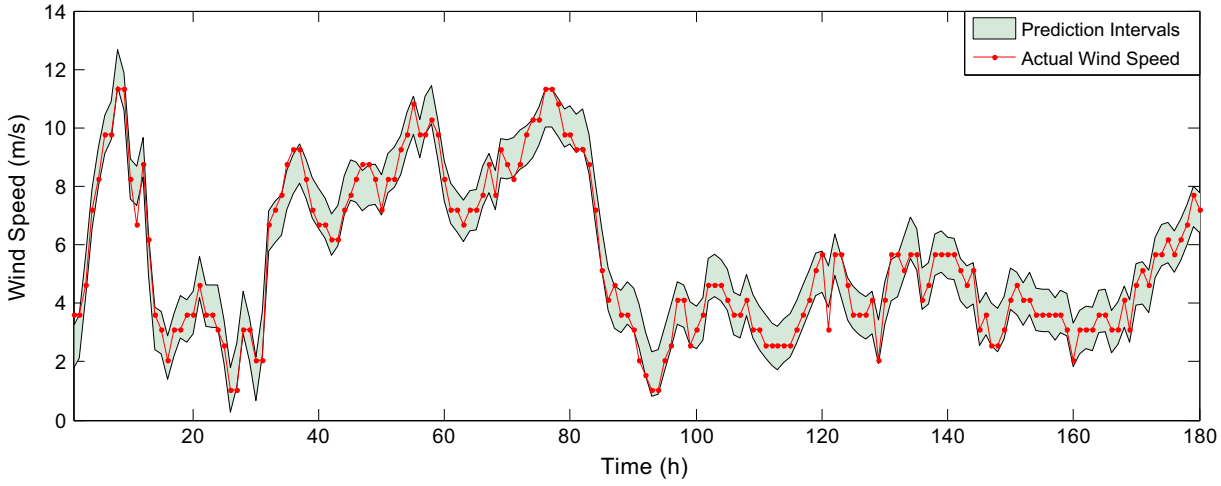

Fig. 13. PIs with PINC 90% in summer 2013 at SIWF obtained from the proposed approach.

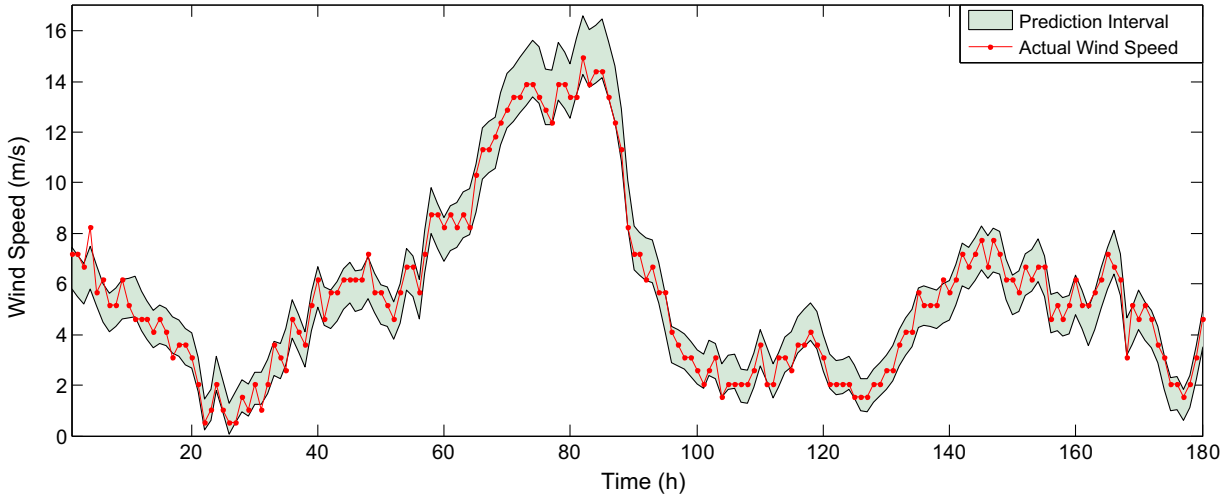

Fig. 14. PIs with PINC 90% in winter 2013 at SIWF obtained from the proposed approach.

cost can be assessed accordingly. Lastly, synthesizing these two impacts generates the total generation cost. To intuitively demonstrate the privilege of the implementation of the proposed method on economic dispatch, a series of simulations on two-wind-turbine-embedded IEEE-30 bus system are carried out and the results are given in Fig. 17. The system parameters are detailed in [45]. The wind power penetration in the system is 10%. From Fig. 17, it is obvious that the proposed approach leads to the lowest total generation cost of the system. The promising results attribute to the most accurate quantiles from the proposed approach for probabilistic WSF. In addition, it can be seen that the generation cost obtained from the Weibull sampling approach [44] exhibits

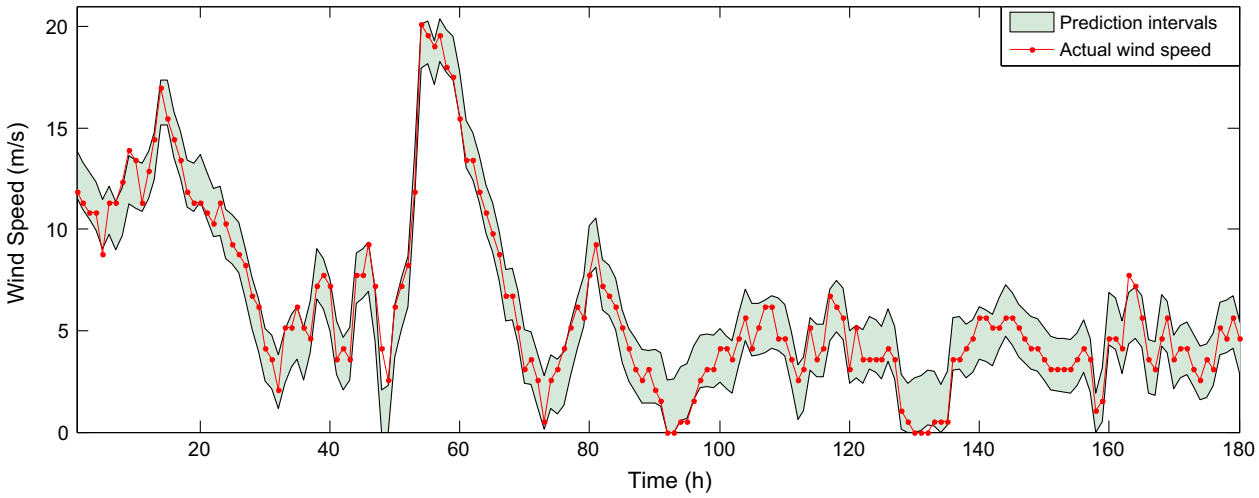

Fig. 15. PIs with PINC 90% in summer 2011 at CRWF obtained from the proposed approach.

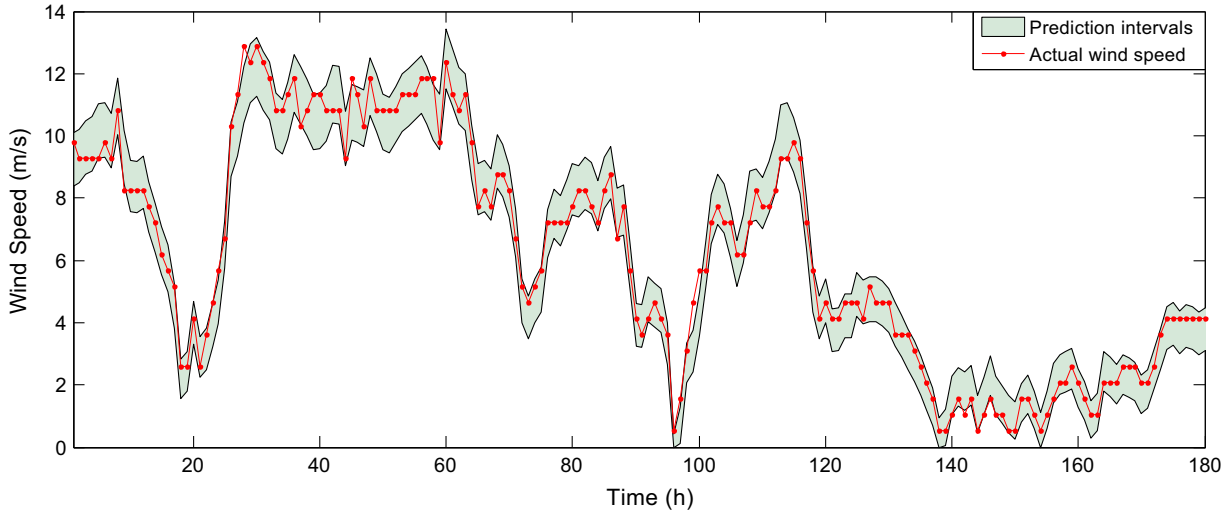

Fig. 16. PIs with PINC 90% in winter 2011 at CRWF obtained from the proposed approach.

Table 5
CRPS in different time resolution.

Wind Farm	Time Resolution	Spring		Summer		Autumn		Winter	
		ARMA	Proposed Approach	ARMA	Proposed Approach	ARMA	Proposed Approach	ARMA	Proposed Approach
SIWF	1-h	0.4129	0.2293	0.4496	0.2083	0.7176	0.3324	0.4721	0.3309
	2-h	0.6067	0.3605	0.6772	0.3183	0.9151	0.5180	0.6390	0.5281
	3-h	0.7532	0.4700	0.8122	0.4137	1.1005	0.6517	0.7580	0.7452
	4-h	0.8765	0.5778	0.9486	0.5341	1.2493	0.8273	0.9874	0.9181
	5-h	0.9838	0.6600	1.0499	0.6467	1.3656	0.9906	1.2774	1.0712
	6-h	1.0708	0.7568	1.1430	0.7707	1.4476	1.1466	1.3868	1.1928
	7-h	1.1440	0.8544	1.1923	0.8748	1.4917	1.3190	1.5074	1.3251
	8-h	1.1893	0.9460	1.2349	0.9627	1.5750	1.4536	1.6139	1.4097
CRWF	5-min	0.2511	0.2401	0.2751	0.2381	0.3751	0.3499	0.1939	0.1892
	15-min	0.4038	0.3332	0.4407	0.3123	0.5646	0.4895	0.3085	0.2691
	30-min	0.5381	0.4239	0.5359	0.3891	0.7612	0.6213	0.4338	0.3520
	45-min	0.6259	0.5172	0.6286	0.4847	0.9579	0.7462	0.5581	0.4375
	60-min	0.7390	0.6082	0.7274	0.5697	1.1050	0.8624	0.6495	0.5281
	75-min	0.8287	0.7101	0.8587	0.6608	1.2085	0.9802	0.7619	0.6220
	90-min	0.9135	0.8104	0.9606	0.7659	1.2999	1.1000	0.8586	0.7219
	120-min	0.9950	0.9027	1.0461	0.8604	1.3853	1.2133	0.9532	0.8129

irregularity and strong randomness. This result is due to the lack of wind speed prediction in this approach. Therefore, we can conclude that the proposed WSF approach has the ability to decrease the operation uncertainties of wind-embedded electric power and energy systems, making it very appealing for practical applications.

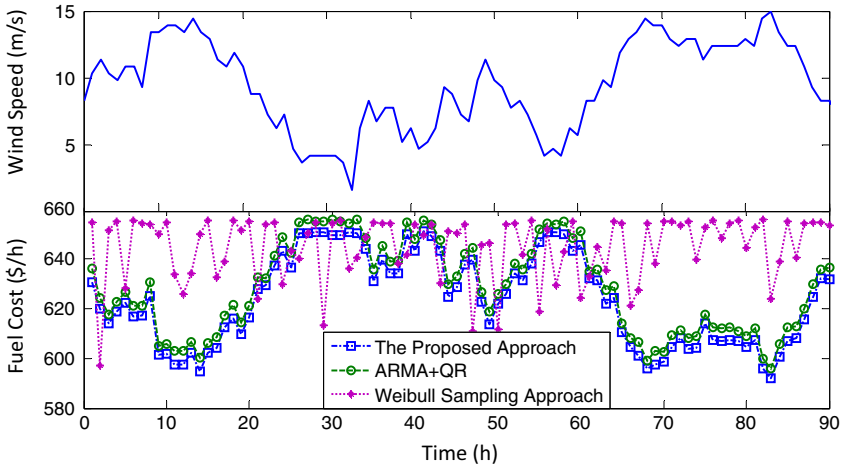

Fig. 17. Total generation cost of two-wind-turbine-embedded IEEE-30-bus system using different algorithms.

Table 6
Run time per training of various algorithms.

Number of training samples	ARMA	BPNN	MWNN	Proposed method
1800	3.5841(s)	5.1514(s)	6.1100(s)	11.4264(s)

A comprehensive study of the average computation time per train over 50 independent runs is tabulated in Table 6. All of the prediction algorithms were implemented in MATLAB R2014a and carried out on a personal computer with an Intel(R) Xeon(R) E3-1225 V2 3.2-GHz CPU and 16.00 GB of RAM. The training samples for each season are $24 \times 25 \times 3 = 1800$. In Table 6, the runtime on 1800 samples is presented. As seen, the average runtime of the proposed WSF approach was not the most efficient from the overall viewpoint. This result is understandable because the DBN-based WSF architecture contains a signal decomposition process and four independent DBN networks, each of which would take many time. However, the proposed approach counts for only 11.4264 s. This type of time span indicates a high potential in on-line application because it is far less than the WSF prediction interval ranging from several minutes to several hours.

6. Conclusions

To date, wind energy has gone through an unexpected growth due to its benefits on climate change mitigation and environmental pollution reduction. As to maximize these benefits, high-accurate wind speed prediction techniques are becoming an essential necessity for the operation and planning of electric power and energy systems. Therefore, this paper originally proposes a deep approach for short-term deterministic and probabilistic WSF. This approach is a hybrid of WT, DBN and QR. The following are the main benefits of the proposed approach: (1) For each wind speed frequency, an independent deep framework based on DBN driven by layer-wise pre-training rule is designed for structure mining and feature extraction. (2) A deterministic approach for WSF based on WT and DBN is proposed to improve the prediction capability by taking advantage of the invariant structures and high-level features. Then, the proposed approach for deterministic WSF is extended to a deep framework which can accurately quantify the randomness and uncertainties exhibited in wind speed. (3) The proposed approach exhibit high-stability which means the prediction results will not change from one simulation to another.

The developed deterministic and probabilistic approaches have been successfully applied in short-term WSF problems on SIWF and CRWF. Compared with the benchmark algorithms on various performance criteria, the in-depth numerical studies have confirmed the superior high-accuracy of the proposed approach for WSF in terms of various seasons and prediction horizons. It is, therefore, convinced that the proposed deterministic and probabilistic WSF approach has a high potential for practical applications in electric power and energy systems.

Acknowledgement

This research was supported in part by National Natural Science Foundation of China (Nos. 51477104, 51507103), Natural Science Foundation of Guangdong Province (No. 2015A030310316), the Foundations of Shenzhen Science and Technology Committee (JCYJ201505250929), Shenzhen University Research and Development Startup Fund (Nos. 2016035, 201530), and National Basic Research Program (973 Program) (No. 2013CB228202).

References

[1] Miguel GL, Ismael S. Regional wind power forecasting based on smoothing techniques, with application to Spanish peninsular system. IEEE Trans Power Syst 2012;27(4):1990–7.
[2] Liu H, Tian HQ, Pan DF, Li YF. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. Appl Energy 2013;107:191–208.
[3] Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. Renewable Energy 2016;85:790–809.
[4] Haque AU, Hashem NM, Mandal P. A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting. IEEE Trans Power Syst 2014;29(4):1663–72.
[5] Salcedo-Sanz S, Perez-Bellido AM, Ortiz-Garcia EG, Portilla-Figueras A, Prieto L, Paredes D. Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. Renew Energy 2009;34(6):1451–4157.
[6] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34(5):1388–93.
[7] Yang M, Fan S, Lee WJ. Probabilistic short-term wind power forecast using componential sparse Bayesian learning. IEEE Trans Ind Appl 2013;49(6):2783–92.
[8] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87(7):2313–20.
[9] Pinson P, Kariniotakis G. Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. Proc 2003 IEEE power tech conf, Bologna, Italy, vol. 2. p. 8–13.
[10] Louka P et al. Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering. J Wind Eng Ind Aerodyn 2008;96(12):2348–62.
[11] Wang Y, Wang JZ, Wei X. A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: a case study of wind farms in northwest China. Energy 2015;91:556–72.

[12] Salcedo-Sanz S, Pérez-Bellido Ángel M, Ortiz-García EG, Portilla-Figueras A, Prieto L, Paredes D. Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. Renew Energy 2009;34(6):451–1457.
[13] Chen K, Yu J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. Appl Energy 2014;113:690–705.
[14] Catalão JPS, Pousinho HMI, Mendes VMF. Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal. IEEE Trans Sustain Energy 2011;2(1):50–9.
[15] Wan C, Xu Z, Pinson P. Optimal prediction intervals of wind power generation. IEEE Trans Power Syst 2014;29(3):1166–74.
[16] Hong Ying-Yi, Chang Huei-Lin, Chiu Ching-Sheng. Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. Energy 2010;35:3870–6.
[17] Zhang Wenyu, Wu Jie, Wang Jianzhou, Zhao Weigang, Shen Lin. Performance analysis of four modified approaches for wind speed forecasting. Appl Energy 2012;99:324–33.
[18] Wang Jian-Zhou, Wang Yun, Jiang Ping. The study and application of a novel hybrid forecasting model – a case study of wind speed forecasting in China. Appl Energy 2015;143:472–88.
[19] Men Z, Yee E, Lien F, Wen D, Chen Y. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. Renew Energy 2016;87:203–11.
[20] Khosravi A, Nahavandi S. Combined nonparametric prediction intervals for wind power generation. IEEE Trans Sustain Energy 2013;4(4):849–56.
[21] Khosravi A, Nahavandi S, Creighton D. Prediction intervals for short term wind farm power generation forecasts. IEEE Trans Sustain Energy 2013;4 (3):602–10.
[22] Wan Can, Zhao Xu, Dong ZY, Wong KP. Probabilistic Forecasting of wind power generation using extreme learning machine, IEEE Trans. Power. System 2014;29(3):1033–44.
[23] Tastu J, Pinson P, Trombe P, Madsen H. Probabilistic forecasts of wind power generation accounting for geographically dispersed information. IEEE Trans Smart Grid 2014;5(1):480–9.
[24] Shrivastava N, Lohia K, Panigrahi B. A multiobjective framework for wind speed prediction interval forecasts. Renew Energy 2016;87:903–10.
[25] Bremnes JB. Probabilistic wind power forecasts using local quantile regression. Wind Energy 2004;7(1):47–54.
[26] Kou Peng, Gao Feng, Guan Xiaohong. Sparse online warped Gaussian process for wind power probabilistic forecasting. Appl Energy 2013;108:410–28.
[27] Giebel G. The state of the art in short term prediction of wind power-A literature overview. Deliverable 1.2b of the ANEMOS. Plus project, vol. 4; 2011.
[28] Hu Qinghua, Zhang Rujia, Zhou Yucan. Transfer learning for short-term wind speed prediction with deep neural networks. Renew Energy 2016;85:83–95.
[29] Lv Yisheng, Duan Yanjie, Kang Wenwen, Li Zhengxi, Wang Fei-Yue. Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transport Syst 2015;16(2):865–73.
[30] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, Yanfeng Gu. Deep learning-based classification of hyperspectral data. IEEE J Sel Top Appl Earth Obs Remote Sens 2014;7(6):2094–107.
[31] Rashwan Mohsen AA, Al Sallab Ahmad A, Raafat Hazem M, Rafea Ahmed. Deep learning framework with confused sub-set resolution architecture for automatic Arabic diacritization. IEEE Trans Audio, Speech, Lang Process 2015;23(3):505–16.
[32] Zhang Chun-Yang, Philip Chen CL, Gan Min, Chen Long. Predictive deep Boltzmann machine for multiperiod wind speed forecasting. IEEE Trans Sustain Energy 2015;6(4):1416–25.
[33] Geoffrey SO, Hinton E, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54.
[34] Tascikaraoglu Akin, Borhan Sanandaji M, Poolla Kameshwar, Varaiya Pravin. Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform. Appl Energy 2016;165:735–47.
[35] Tascikaraoglu A, Uzunoglu M, Vural B. The assessment of the contribution of short-term wind power predictions to the efficiency of stand-alone hybrid systems. Appl Energy 2012;94:156–65.
[36] Mallat S. A theory for multiresolution signal decomposition—the wavelet representation. IEEE Trans Pattern Anal Mach Intell 1989;11(7):674–93.
[37] Bengio Y. Learning deep architectures for AI. Found Trends Mach Learn 2009;2 (1):1–127.
[38] Papaa JP, Scheirerb W, Coxb DD. Fine-tuning deep belief networks using harmony search. Appl Soft Comput 2015. http://dx.doi.org/10.1016/j.asoc.2015.08.043.
[39] Haykin S. Neural networks and learning machines. 3rd ed. Englewood Cliffs (NJ, USA): Prentice-Hall; 2008.
[40] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy 2006;9:95–108.
[41] Pinson P, Reikard G, Bidlot J-R. Probabilistic forecasting of the wave energy flux. Appl Energy 2012;93:364–70.
[42] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J Roy Stat Soc 2007;69(2):243–68.
[43] Narayana M, Putrus G, Jovanovic M, Leung PS. Predictive control of wind turbines by considering wind speed forecasting techniques. In: Universities power engineering conference. p. 1–4.
[44] Hetzer J, Yu DC, Bhattarai K. An economic dispatch model incorporating wind power. IEEE Trans Energy Convers 2008;23(2):603–11.
[45] Farag A, AI-Baiyat S, Cheng TC. Economic load dispatch multiobjective optimization procedures using linear programming techniques. IEEE Trans Power Syst 1995;10(2):731–8.