Renewable Energy 148 (2020) 135–149

Contents lists available at ScienceDirect

Renewable Energy

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

Multi-distribution ensemble probabilistic wind power forecasting

Mucun Sun, Cong Feng, Jie Zhang*

The University of Texas at Dallas, Richardson, TX, 75080, USA

A R T I C L E I N F O

Article history:
Received 6 February 2019
Received in revised form
24 August 2019
Accepted 26 November 2019
Available online 29 November 2019

Keywords:
Probabilistic wind power forecasting
Optimization
Surrogate model
Pinball loss
Ensemble forecasting

A B S T R A C T

Ensemble methods have shown to be able to improve the performance of deterministic wind forecasting. In this paper, an improved multi-distribution ensemble (MDE) probabilistic wind power forecasting framework is developed to explore the advantages of different predictive distributions. Both competitive and cooperative strategies are applied to the developed MDE framework to generate 1–6 h ahead and day-ahead probabilistic wind power forecasts. Three probabilistic forecasting models based on Gaussian, gamma, and laplace predictive distributions are adopted to form the ensemble model. The parameters of the ensemble model (i.e., weights and standard deviations) are optimized by minimizing the pinball loss at the training stage. A set of surrogate models are built to quantify the relationship between the unknown optimal parameters and deterministic forecasts, which can be used for online forecasting. The effectiveness of the proposed MDE framework is validated by using the Wind Integration National Dataset (WIND) Toolkit. Numerical results of case studies at seven locations show that the developed MDE probabilistic forecasting methodology has improved the pinball loss metric score by up to 20.5% compared to the individual-distribution models and benchmark ensemble models.

1. Introduction

Wind power is considered to be one of the most promising renewable energy sources in modern power systems due to the great economic, technological, and environmental incentives for wind development [1]. However, the uncertain and variable nature of wind power creates a big challenge for reliable and economic power system operations. Wind power forecasting has been developed and improved in recent years to assist grid integration of wind power. Both deterministic and probabilistic wind forecasting models have been developed in the literature. Probabilistic forecasts that provide quantitative uncertainty information associated with wind power become extremely important for reliable and economic power system operations and planning with an ever-increasing wind energy penetration. A number of wind power forecasting projects including probabilistic forecasts have been conducted in recent years, such as wind forecast improvement project (WFIP) [2], Alberta Electric System Operator (AESO) wind power forecasting pilot project [3], and ANEMOS project [4].

1.1. Literature review

A number of probabilistic wind power forecasting methods have been developed in the literature to assist power system operations. For example, probabilistic wind forecasting has been used for unit commitment and economic dispatch [5–7], electricity market [8–10], operating reserves [11–13], and energy storage sizing [14,15]. Probabilistic wind power forecasts usually take the form of prediction intervals, quantiles, or predictive distributions. Methods of probabilistic wind power forecasting can be classified into nonparametric and parametric approaches [16]. Nonparametric approaches are distribution free, and their predictive distributions are estimated through observations. Quantile regression [17] and kernel density estimation [18] are two of the most popular used methods for probabilistic forecasting. For example, Wan et al. [19] combined quantile regression and extreme learning machine with a linear programming optimization model to generate probabilistic forecasts. Bessa et al. [20] proposed a novel time-adaptive quantile-copula estimator for kernel density forecast and discussed how to select the adequate kernels for modeling different variables. Other nonparametric methods include machine learning [21,22] and deep learning based methods [23,24]. Parametric approaches assume a pre-defined shape of predictive distribution with an analytical expression, and the unknown distribution parameters are estimated based on historical data. A number of distribution types,

* Corresponding author.
E-mail address: jiezhang@utdallas.edu (J. Zhang).

https://doi.org/10.1016/j.renene.2019.11.145

such as Gaussian distribution [25], Beta distribution [26], Logit-Normal distribution [27], and mixed models [28,29], have been used as predictive distributions in the literature for parametric probabilistic wind power forecasting. *Most of the probabilistic forecasting methods mentioned above are developed based on a single model, especially for parametric approaches which require a distribution assumption. However, it is challenging to accurately characterize the associated uncertainty by using a single distribution type, due to the daily and seasonal variability in wind power.* Ensemble methods have been shown to present better forecasting performance for deterministic wind power forecasting. However, the performance of applying ensemble methods to probabilistic wind forecasting is still not well studied.

1.1.1. Ensemble methods in wind forecasting

Methods of constructing ensemble forecasts can be mainly classified into two groups: competitive ensemble methods and cooperative ensemble methods [30]. *Competitive ensemble methods* use induction algorithms with different parameters or initial conditions to build individual forecasting models. The final refined ensemble prediction is obtained from pruning and aggregating individual forecasts. *Competitive ensemble methods* generally require a large diversity of data and parameters to obtain different forecasts from individual predictors [31]. Therefore, competitive ensemble methods usually require expensive computation cost, and they are usually used in medium-term and long-term forecasting. Bagging [32] and boosting [33] are two commonly used competitive ensemble methods. To better account for the performance of weak models, ensemble forecasting approaches based on adaptive boosting (i.e., assign large weights to the models with larger errors) are used in Refs. [34–36]. Feng et al. [37,38] proposed a machine learning based multi-model forecasting framework consisting of an ensemble of four single machine learning algorithms with various kernels to generate deterministic forecasts. For *cooperative ensemble methods*, the dataset is divided into several sub-datasets and each sub-dataset is forecasted separately, and the final forecasts are obtained by aggregating all the sub-forecasts. Cooperative ensemble methods usually have lower computation burden due to less parameter tuning work, which are normally used in very short-term or short-term forecasting [31]. Artificial neural networks based autoregressive integrated moving average [39] and generalized autoregressive conditional heteroskedasticity based autoregressive integrated moving average [40] are two commonly used cooperative ensemble methods that combine suitable models for linear and non-linear time series. Ren et al. [41] developed an AdaBoost based empirical mode decomposition (EMD) artificial neural network to generate ensemble wind speed forecasts. Wang et al. [42] also developed a cooperative ensemble method to generate wind speed forecasts, which improved EMD through decomposing the original data into more stationary signals with different frequencies. These stationary signals are considered as different inputs to a genetic algorithm and backpropagation based neural network, and the final forecast is the aggregation of each single prediction. Gensler et al. [43] used both cooperative and competitive ensemble methods to generate wind forecasts, where the weight of each input model is adaptively modified based on its performance.

The ensemble forecasting methods discussed above are mainly used for deterministic wind speed/power forecasting. Only a few recent studies have been done in the literature on ensemble probabilistic wind power forecasting. In Ref. [44], Zhang et al. proposed an ensemble probabilistic wind power forecasting approach based on empirical mode decomposition, sample entropy techniques and extreme learning machine. Lin et al. [13] combined multiple probabilistic forecasting models based on sparse Bayesian learning, kernel density estimation, and beta distribution estimation. The weight parameters of the multi-model ensemble are solved by an expectation maximizing algorithm and continuous ranked probability score optimization. Kim et al. [45] developed an enhanced ensemble method for probabilistic wind power forecasting. The wind speed spatial ensemble was built by using correlation based weight and kriging models, and the temporal ensemble was built through an average ensemble of three models (i.e., an exogenous variable model, a polynomial regression model, and an analog ensemble model). Wang et al. [23] used wavelet transform to distinguish the non-linear series, and an ensemble technique was used to cancel out the diverse errors of point forecasts. Probabilistic wind power forecasts were generated by using a convolutional neural network. Nevertheless, several challenges present in existing methods of ensemble probabilistic wind power forecasting: (i) high dimensional matrices are involved in spatio-temporal based model in Ref. [45], which adds additional computational burden; (ii) parametric and nonparametric probabilistic forecasting models are combined in Ref. [13], which may encounter unbalanced issue with limited number of observations; (iii) little work has been done on the ensemble and optimization of predictive distributions.

1.2. Research objective

To integrate the advantages of different types of predictive distributions, this paper develops an ensemble probabilistic wind power forecasting framework that combines multiple predictive distribution types. First, quantile forecasts (or quantile functions) based on each single predictive distribution type (e.g., Gaussian, gamma, and laplace) are generated through a pinball loss optimization based framework. Second, a set of weight parameters are assigned to the quantile values or the quantile functions of individual probabilistic wind power forecasts. The weight parameters are determined through solving an optimization problem that minimizes a loss function. Note that the optimal weight parameters of each individual model are adaptively and dynamically updated at each time step. *The main contributions of this paper include: (i) developing an ensemble probabilistic wind power forecasting framework based on different predictive distributions; (ii) comparing the proposed ensemble forecasting method with different benchmark ensemble strategies at both day-ahead and very-short-term forecasting horizons; and (iii) exploring both competitive and cooperative ensemble strategies with the developed forecasting framework.*

The rest of this paper is organized as follows. Section 2 introduces the **competitive** and **cooperative** ensemble strategies for probabilistic forecasting. Section 3 describes the proposed ensemble probabilistic forecasting framework, which is composed of a deterministic forecasting method, a multi-distribution database, a parameter optimization process, and a surrogate model selection process. Section 4 compares the proposed method with multiple benchmark ensemble models at seven locations. Concluding remarks and future work are discussed in Section 5.

2. Ensemble probabilistic forecasting framework

The overall framework of the proposed ensemble probabilistic forecasting method, named Multi-Distribution Ensemble (MDE), is a two-step forecasting method, which consists of deterministic forecasts generation and ensemble probabilistic forecasts generation. The proposed MDE framework is able to integrate with both competitive and cooperative ensemble strategies. The main differences between competitive and cooperative ensemble models are listed as follows.

1. The competitive ensemble method combines different individual predictive distribution models together to an ensemble distribution model first. Then, probabilistic forecasts are obtained from the ensemble distribution model.
2. The cooperative ensemble method generates different individual probabilistic forecasts first based on each predictive distribution type. Then, ensemble probabilistic forecasts are generated through combining individual probabilistic forecasts to ensemble probabilistic forecasts.

2.1. Competitive MDE method

The overall framework of the developed competitive MDE method is illustrated in Fig. 1. In the first step, a Q-learning-based forecasting framework is adopted to generate short-term deterministic wind forecasts (i.e., 1-6 h-ahead (HA)), and a numerical weather prediction (NWP) model is used to generate day-ahead deterministic forecasts. The generated deterministic forecasts are considered as mean values of predictive distributions. In the second step, a set of unknown parameters (i.e., weight parameters and standard deviations) of the ensemble model are determined by minimizing the pinball loss based on training data. Then a surrogate model is constructed to represent each optimal parameter as a function of the deterministic forecast. During online forecasting, a set of pseudo-optimal parameters of the ensemble model are estimated by the surrogate model and deterministic forecasts. Finally, probabilistic forecasts are generated with the distribution means (i.e., deterministic forecasts) and pseudo-optimal parameters.

2.2. Cooperative MDE method

The overall framework of the cooperative MDE method is illustrated in Fig. 2. The major differences between the cooperative MDE and competitive MDE methods are highlighted in the **blue** dashed square. For the cooperative MDE method, the first step is also to generate deterministic forecasts. In the second step, individual probabilistic forecasts are generated by using each single predictive distribution based on the training dataset [46]. The unknown parameters (i.e., standard deviations) of each predictive distribution are optimized. A weight parameter is assigned to quantile forecasts from each individual model, and these weight parameters are optimized again by minimizing the pinball loss. Then a surrogate model is developed to represent each optimal weight as a function of the deterministic forecast. Similar to the competitive MDE method, during the online forecasting, pseudo-optimal weights are estimated and used to generate ensemble probabilistic forecasts. Finally, the method with the minimum pinball loss is chosen to produce the final ensemble probabilistic forecasts. Overall, the competitive MDE method optimizes the unknown weight parameters and standard deviations simultaneously; the cooperative MDE method optimizes the standard deviations when generating the individual model forecasts, and optimize the weight parameters at the ensemble stage.

3. Modeling and formulation of MDE

3.1. Q-learning enhanced deterministic forecasting

A large number of models have been developed in the literature for deterministic wind forecasting. Most of existing deterministic forecasting methods are either selected based on the overall performance or ensembled by multiple models. Ensembling multiple deterministic models enhances the robustness of the forecasting by reducing the risk of unsatisfactory models, but does not guarantee the best accuracy [47]. Selecting a model based on the overall

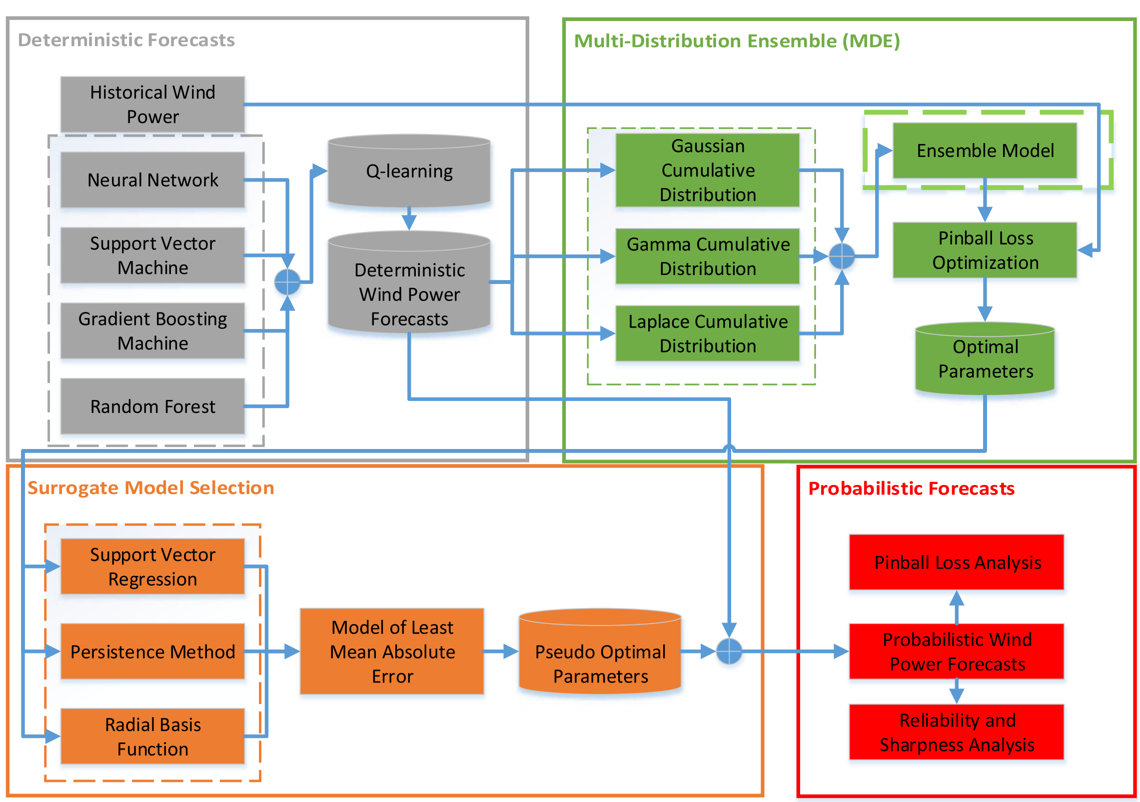


Fig. 1. Overall framework of the MDE competitive probabilistic wind power forecasting method.

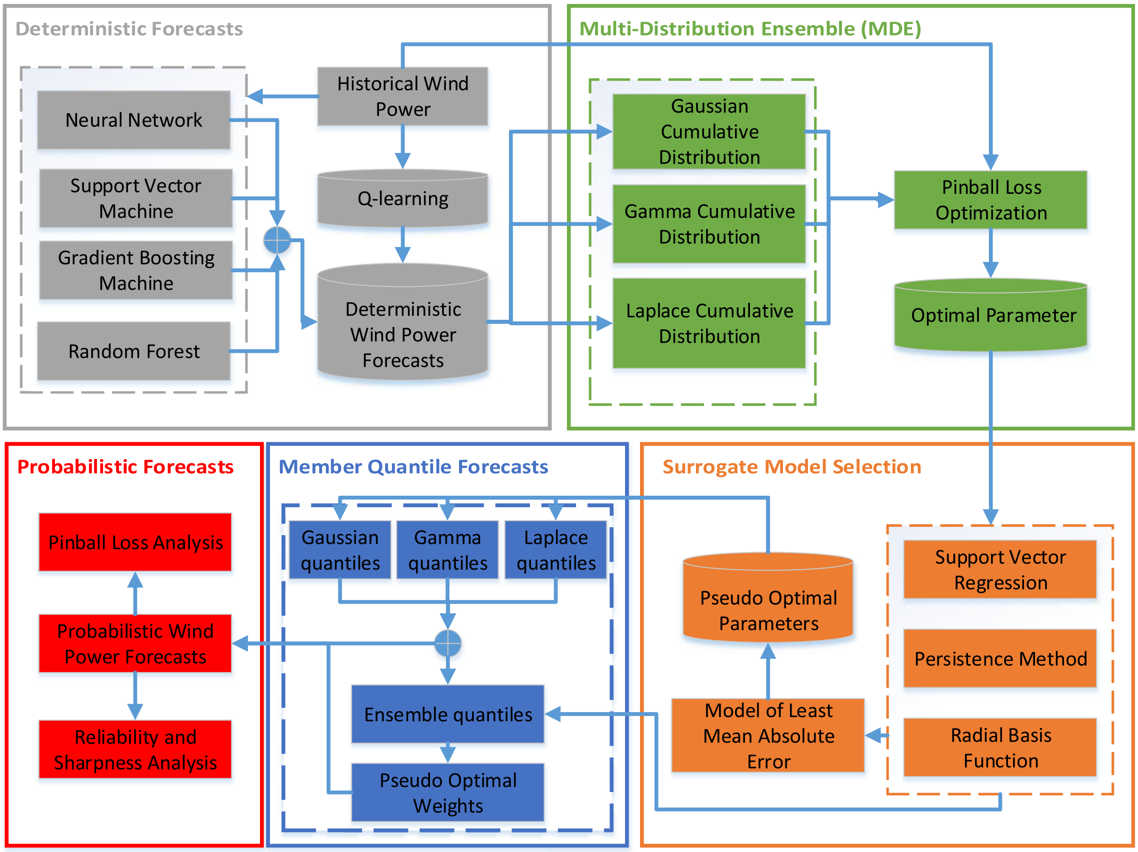


Fig. 2. Overall framework of the MDE cooperative probabilistic wind power forecasting method.

forecasting performance, generally neglects the local performance of the selected model.

In this paper, a Q-learning enhanced deterministic forecasting method is adopted for short-term forecasting (1HA to 6HA), which seeks to choose the best forecasting model from a pool of state-of-the-art machine learning based forecasting models at each forecasting time step [48]. The developed method trains Q-learning agents based on the rewards of transferring from the current model to the next model. The Q-learning agents converge to the optimal dynamic model selection policy, which will be applied to select the best model for forecasting in the next step based on the current model. The dynamic model selection process is expressed as:

$$\boldsymbol{S} = \{\boldsymbol{s}\} = \{s_1, s_2, \ldots, s_I\} \tag{1}$$

$$\boldsymbol{A} = \{\boldsymbol{a}\} = \{a_1, a_2, \ldots, a_I\} \tag{2}$$

$$R^t(s_i, a_j) = ranking(M_i) - ranking(M_j) \tag{3}$$

where $\boldsymbol{S}$, $\boldsymbol{A}$, and R are state space, action space, and reward function in the dynamic model selection Markov Decision Process, respectively. The parameters s and a are possible state and action, respectively. I is the number of models (M) in the model pool. The reward function is defined as the model performance improvement, which ensures the effective and efficient convergence of Q-learning. More details about the Q-learning enhanced deterministic forecasting can be found in Ref. [48].

3.2. Probabilistic wind power forecasting model

In this section, multiple predictive distribution types are used to generate different individual probabilistic wind power forecasting models. Probabilistic forecasts usually take the form of probability density functions (PDFs) or quantiles. Parametric approaches have been widely used in the literature to estimate the density or distribution [49]. For a certain form of predictive distribution, the PDF can generally be characterized by a mean value μ and a standard deviation value σ as $f(x|\mu,\sigma)$, and the corresponding cumulative distribution function (CDF) can be deduced and denoted as $F(x|\mu,\sigma)$. The quantile function is one way of prescribing a probability distribution, which is the inverse function of its corresponding CDF. Therefore, the quantile function of a certain predictive distribution can be denoted as $q_i = F^{-1}\left(\frac{i}{100}\right)$, where i stands for the ith quantile. The quantile functions $q_{i,t}(.)$ are distinct for different predictive distributions, such as Gaussian, gamma, and laplace. Gaussian distribution is one of the most commonly used predictive distributions in parametric approaches [25]. The PDF of Gaussian distribution $f(x|\mu,\sigma)$ is expressed as

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

where μ and σ are the mean and standard deviation, respectively. The laplace distribution has also been applied to describe wind power generation [16]. The PDF of laplace distribution $f(x|\mu,b)$ is formulated as

$$f(x|\mu,b)=\frac{1}{2b}e^{-\frac{|x-\mu|}{b}} \tag{5}$$

where μ and b are the location and scale parameters, respectively. The relationship between the standard deviation σ and the shape parameter of laplace distribution can be expressed as

$$b=\frac{\sigma}{\sqrt{2}} \tag{6}$$

Similarly, for gamma distribution, the PDF is expressed as

$$f(x|k,\theta)=\frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^{k}\Gamma(k)} \tag{7}$$

where k and θ are the shape parameter and scale parameter, respectively. The relationship between the mean value μ, standard deviation σ, k, and θ can be expressed as

$$\mu=k\theta \tag{8}$$

$$\sigma=\sqrt{k}\theta \tag{9}$$

The location parameter and shape parameter from all of the above distributions can be represented by the mean and standard deviation. Therefore, the PDF can be obtained by estimating the mean and standard deviation in the training or forecasting stages. The pseudocode of generating the MDE forecasting member models is illustrated in Algorithm 1.

Algorithm 1. Generate quantile forecasts through each single distribution.

```
Data: Deterministic wind power forecasts
Result: Quantile forecasts
Initialization: Obtain PDF of a single model and represent it in the form
  of mean μ and standard deviation σ as f(x|μ,σ);
Calculate CDF F(x|μ,σ) of the predictive distribution;
if CDF is invertible then
  | Calculate quantile function through q_i(F^{-1}(x|μ,σ, i/100))
else
  | Use Newton-Raphson method to calculate quantile q_i
end
Calculate optimal σ's through pinball loss optimization;
Build a surrogate model between deterministic forecasts and optimal σs;
Estimate pseudo-optimal σ based on deterministic forecasts;
Generate quantile forecasts;
```

3.3. MDE model formulation

3.3.1. Competitive MDE method

For the competitive MDE method, to combine the advantages of different predictive distributions, a general combined quantile function with N member models is formulated as follows:

$$q_{i,t}^{p}(\mu_t,\omega_t^{p},\sigma_t)=\sum_{n=1}^{N}\omega_{t,n}^{p}q_{i,t,n}(\mu_t,\sigma_{t,n}) \tag{10}$$

where $q_{i,t}^{p}$ is the combined ith quantile at the tth forecasting step, $q_{i,t,n}(\mu_t,\sigma_{t,n})$ is the ith member quantile at the tth forecasting step, and $\omega_{t,n}^{p}$ is the weight of the nth model at time t satisfying that

$$0\leq\omega_{t,n}^{p}\leq 1 \tag{11}$$

$$\sum_{n=1}^{N}\omega_{t,n}^{p}=1 \tag{12}$$

The unknown parameters in the ensemble quantile function can be solved by minimizing the pinball loss at each time step. Pinball loss is a widely used metric to evaluate probabilistic forecasts, which is defined by

$$L_{i,t}\left(q_{i,t},x_t\right)=\begin{cases}\left(1-\frac{i}{100}\right)\times\left(q_{i,t}-x_t\right), & x_i<q_{i,t}\\ \frac{i}{100}\times\left(x_t-q_{i,t}\right), & x_t\geq q_{i,t}\end{cases} \tag{13}$$

where x_t represents the observation at time t. For a given i percentage, the quantile $q_{i,t}$ represents the value of a random variable whose CDF is i percentage at time t. The pinball loss optimization problem for the competitive MDE model can be formulated as:

$$\begin{aligned}&\min_{\omega_t^{p},\sigma_t}\sum_{i=1}^{99}L_{i,t}\left(\sum_{n=1}^{N}\omega_{t,n}^{p}q_{i,t,n}(\mu_t,\sigma_{t,n}),x_t\right)\\ &\text{s.t.}\sum_{n=1}^{N}\omega_{t,n}^{p}=1\\ &0\leq\omega_{t,n}^{p}\leq 1,\ \forall n\in N\\ &\sigma_l\leq\sigma\leq\sigma_\mu.\end{aligned} \tag{14}$$

where σ_l and σ_u represent the lower and upper bounds of the unknown standard deviation, respectively, which are selected based on the forecasting target [50].

3.3.2. Cooperative MDE method

For the cooperative MDE method, with the calculated ith quantile of the nth model, a general combined quantile with N individual models is formulated as:

$$q_{i,t}^{o}(\mu_t,\omega_t^{o})=\sum_{n=1}^{N}\omega_{t,n}^{o}q_{i,t,n}(\mu_t) \tag{15}$$

$$0\leq\omega_{t,n}^{o}\leq 1 \tag{16}$$

$$\sum_{n=1}^{N}\omega_{t,n}^{o}=1 \tag{17}$$

where $q_{i,t}^{o}$ is the combined ith quantile forecasts at the tth forecasting step, $q_{i,t,n}(\mu_t)$ is the ith member quantile forecasts at the tth forecasting step from model n, and $\omega_{t,n}^{o}$ is the weight of the quantile forecasts from the nth model. Different from the competitive MDE method, weight parameters are the only unknown parameters in the cooperative MDE method.

It is seen from Eqs. (10) and (15) that the weights of individual models at each time step are different, and the location and shape parameters at each time step are also varying. Therefore, we need to adaptively estimate the unknown parameters at each time step. By contrast, for the cooperative MDE method, the weight parameters can be estimated by solving the following optimization problem:

$$\begin{aligned}&\min_{\omega_t^o}\sum_{i=1}^{99}L_{i,t}\Big(\sum_{n=1}^{N}\omega_{t,n}^o q_{i,t,n}(\mu_t),x_t\Big)\\&\text{s.t.}\sum_{n=1}^{N}\omega_{t,n}^o=1\\&0\le\omega_{t,n}^o\le1,\ \forall n\in N.\end{aligned}\tag{18}$$

The optimization problems in Eqs. (14) and (18) are strictly constrained by three constraints: (i) all the weights of ensemble member models or member quantile forecasts must be nonnegative; (ii) the sum of all weights equals one; and (iii) the standard deviation of each component model must be in the variable range of the corresponding distribution. In this paper, the genetic algorithm (GA) [51] is adopted to solve the two optimization problems. In this study, the population size is set to be 50, the mutation percentage is 0.1, the crossover percentage is 0.8, and the maximum number of iterations is 200. The optimization stops when the improvement is less than 0.01%.

3.4. Surrogate model

The optimization models in Eqs. (14) and (18) calculate the optimal parameters of predictive distributions with observations. However, when we generate probabilistic forecasts, we do not know the observations and can only assume the deterministic forecasts as mean values of the predictive distribution. The optimal standard deviations and optimal weights at each forecasting time step are needed to generate probabilistic forecasts. To this end, a surrogate model is built to represent the optimal parameters as functions of the deterministic forecasts in the training stage, which can be expressed as:

$$\widehat{\sigma}_{t,n}=f(\widehat{x}_t)\tag{19}$$

$$\widehat{\omega}_{t,n}=g(\widehat{x}_t)\tag{20}$$

where $\widehat{x}_t$ is the deterministic forecast at time t, $f(\widehat{\sigma})$ and $g(\widehat{\omega})$ are the surrogate models of the optimal standard deviation and weight parameter of the predictive distribution, respectively. The surrogate models are used to estimate the pseudo-optimal standard deviation and pseudo-optimal weight parameters at the online forecasting stage.

Several possible surrogate models can be used, such as support vector regression (SVR), radial basis function (RBF), and persistence model (PS). In this paper, a surrogate model selection framework is developed to choose the most suitable surrogate model for estimating weights and standard deviations of each ensemble member models. Fig. 3 shows the procedure of selecting the optimal surrogate model. At the forecasting training stage, a small portion (i.e., 8% of the training data in this paper) of the training data is used to build a surrogate model. The optimal training parameters are fed into a surrogate model pool (that consist of SVR, RBF, and PS) to find the best fitted surrogate model between the point forecast value and each optimal parameters. The model with the minimum mean absolute error (MAE) is selected as the surrogate model for that parameter to be used at the online forecasting stage. In the forecasting stage, the optimal unknown parameters of the predictive distribution are estimated through the trained surrogate model, and thus the estimated parameters are referred to as pseudo-optimal parameters.

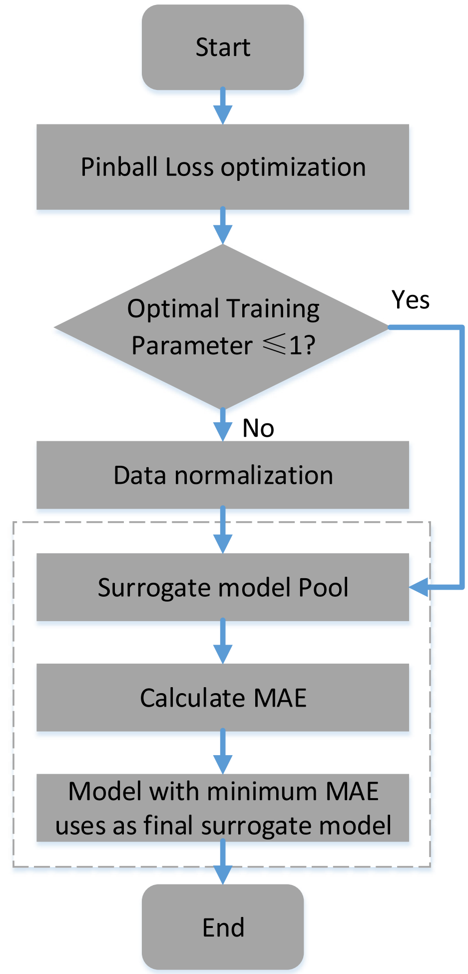


Fig. 3. Surrogate model selection.

4. Case studies and results

4.1. Data summary

The developed MDE probabilistic forecasting framework was evaluated at 7 locations selected from the Wind Integration National Dataset (WIND) Toolkit [52]. The WIND Toolkit includes meteorological information (e.g., wind direction, wind speed, air temperature, surface air pressure, density at hub height), synthetic actual wind power, and wind power forecasts generated by the Weather Research and Forecasting (WRF) model. It covers over 126,000 locations in the United States. In addition to the day-ahead forecasting, very-short-term forecasts (i.e., 1HA to 6HA) are generated by using the Q-learning enhanced deterministic forecasting method. In this study, the duration of the collected data at the selected 7 locations spans two years from January 1st 2011 to December 31st 2012. The data information at the selected 7 locations is briefly summarized in Table 1. For all the locations, the first 3/4 of the data is used as training data, in which the first 11/12 is used to train the deterministic forecast models and the remaining 1/12 of the training data is used to build the surrogate models of the optimal standard deviations and weight parameters. The effectiveness of the forecasts is validated by the remaining 1/4 of the data. The developed MDE method is able to generate probabilistic forecasts at multiple forecasting horizons, and 1HA-6HA and day-ahead wind power forecasts are explored in this study. The Q-

Table 1
Data summary of the selected 7 WIND Toolkit sites.

Case No.	Site ID	Lat.	Long.	Capacity (MW)	State
C1	4816	29.38	−100.37	16	TX
C2	8979	31.53	−95.62	16	TX
C3	10069	32.31	−98.26	16	TX
C4	10526	32.44	−100.55	16	TX
C5	1342	27.12	−97.86	16	TX
C6	2061	27.95	−97.40	14	TX
C7	9572	31.99	−100.118	16	TX

learning enhanced deterministic forecasts are implemented using the gbm package [53], randomForest package [54], e1071 package [55], and nnet package [56] in R version 3.4.2. The optimization model for the ensemble is implemented using the GA package [51] in R version 3.4.2.

4.2. Benchmarks and comparison settings

In the paper, two different weight averaging methods for model ensemble and three single predictive distribution models are used as baselines in the case studies. The two benchmark weight averaging based ensemble methods are Arithmetic Averaging (AA) and Weighted Averaging (WA).

1. AA: The arithmetic weights apply equally to different models:

$$\omega_{t,n}^{AA} = \frac{1}{N}. \tag{21}$$

Then, the combined quantile of AA can be calculated through:

$$q_{i,t}^{AA}\left(\mu_t, \omega_t^{AA}\right) = \sum_{n=1}^{N} \omega_{t,n}^{AA} q_{i,t,n}(\mu_t) \tag{22}$$

2. WA: Each quantity to be averaged is assigned a weight that represents the relative importance of that quantity. The model with a higher accuracy is assigned a larger weight:

$$\omega_{t,n}^{WA} = \frac{1 \Big/ \sum\limits_{i=1}^{99} L_{i,t,n}\left(q_{i,t,n}(\mu_t), x_t\right)}{\sum\limits_{n=1}^{N} 1 \Big/ \sum\limits_{i=1}^{99} L_{i,t,n}\left(q_{i,t,n}(\mu_t), x_t\right)} \tag{23}$$

Similarly, the combined quantile of WA can be calculated through:

$$q_{i,t}^{WA}\left(\mu_t, \omega_t^{WA}\right) = \sum_{n=1}^{N} \omega_{t,n}^{WA} q_{i,t,n}(\mu_t) \tag{24}$$

The three single predictive distribution based optimization models are Q-learning with Gaussian distribution (Q-Gaussian), Q-learning with Gamma distribution (Q-Gamma), and Q-learning with Laplace distribution (Q-Laplace). The details of the pinball loss optimization based model can be found in Ref. [16].

4.3. Deterministic forecasting results

Normalized indices of standard metrics root mean squared error and mean absolute error, *i.e.*, NMAE and NRMSE, are adopted to evaluate the performance of deterministic forecasts. They are defined by:

$$NMAE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\hat{x}_t - x_t}{C} \right| \times 100\% \tag{25}$$

$$NRMSE = \frac{1}{C} \sqrt{\frac{\sum_{t=1}^{T} (\hat{x}_t - x_t)^2}{T}} \times 100\% \tag{26}$$

where $\hat{x}_t$ is the forecasted wind power, x_t is the actual wind power, x_{max} is the maximum actual wind power, T is the sample size, and C is the capacity of the wind farm.

A smaller NRMSE or NMAE indicates a better forecasting performance. The forecasting errors by using the Q-learning based deterministic model at the selected locations are summarized in Table 2. It is shown that the 1HA NMAE and NRMSE are in the ranges of 5%–8% and 8%–12%, respectively. The numerical weather prediction (NWP) method is used to produce day-ahead deterministic forecasts. It is shown that the day-ahead NMAE and NRMSE are in the range of 11%–14% and 15%–19%, respectively. An example of the day-ahead forecasts at the C3 site from 2012-08-09 to 2012-08-19 is illustrated in Fig. 4. Overall, the accuracies of the Q-learning based 1HA to 6HA deterministic forecasts and the NWP-based day-ahead deterministic forecasts are reasonable.

4.4. Surrogate model accuracy

Three types of surrogate models, i.e., SVR, RBF, and persistence model are adopted in this paper to build the surrogate model of the optimal parameter (i.e., standard deviation and weight). For all the surrogate models in this study, the NMAE and NRMSE are in the range of 9%–17% and 13%–24%, respectively. Overall, the accuracies of the surrogate models are reasonable.

4.5. Ensemble probabilistic forecasting results

To show the robustness of the developed MDE probabilistic forecasting framework, the normalized pinball loss values at different look-ahead time of the 7 wind farms are illustrated in Fig. 5. Fig. 6 shows how the normalized pinball loss varies with the look-ahead time at the C7 site. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the wind farm capacity at each site. A lower pinball loss score indicates a better probabilistic forecast. Results show that the MDE-competitive model has improved the pinball loss by up to 20.5%, and the MDE-cooperative model has improved the pinball loss by up to 8.5% compared to the three individual member models (i.e., Q-Gaussian, Q-Gamma, and Q-Laplace) and two baseline weight averaging ensemble models (i.e., AA and WA). It is seen from Fig. 5 that the MDE-competitive model has the smallest pinball loss value at all locations for different look-ahead times except for 1HA forecasts. For 1HA forecasts, the MDE-cooperative method has the smallest pinball loss value at all locations. Note that for all the look-ahead hours, the MDE-cooperative model has also shown a better accuracy than individual member models, which validates the effectiveness of the MDE ensemble framework. The MDE-competitive model has shown a better accuracy than individual member models for 2HA-6HA and day-ahead forecasts.

The reason why the MDE-competitive method has a worse accuracy than the MDE-cooperative model and some single-distribution methods in 1HA forecasts is that, the accuracy of 1HA deterministic forecasts is significantly better than that of other look-ahead times. Thus there is less improvement space for the ensemble model to further improve the performance by optimizing the standard deviation and weight parameters simultaneously.

Table 2
Deterministic wind power forecasting results by using Q-learning and NWP.

Model	LAT	Metric	Site						
			C1	C2	C3	C4	C5	C6	C7
Q-learning	1HA	NMAE(%)	6.63	6.85	6.70	6.74	6.67	5.38	7.76
		NRMSE(%)	10.55	10.93	11.04	10.66	9.93	8.37	11.93
	2HA	NMAE(%)	10.72	11.08	11.20	11.28	10.36	8.34	10.90
		NRMSE(%)	15.86	16.33	16.93	16.64	14.60	12.31	17.14
	3HA	NMAE(%)	13.95	14.22	14.65	14.88	12.44	10.49	14.56
		NRMSE(%)	19.43	19.86	20.92	20.71	16.94	14.93	20.12
	4HA	NMAE(%)	16.43	16.76	17.69	18.00	13.84	12.17	16.28
		NRMSE(%)	21.98	22.22	23.92	23.83	18.56	16.85	21.79
	5HA	NMAE(%)	18.19	18.37	20.36	20.38	15.35	13.53	17.87
		NRMSE(%)	23.65	23.71	26.51	26.01	20.19	18.40	23.15
	6HA	NMAE(%)	19.85	19.60	21.94	22.63	16.51	14.70	18.41
		NRMSE(%)	25.09	24.85	27.82	27.82	21.31	19.47	23.55
NWP	DA	NMAE(%)	12.70	12.59	13.21	13.97	13.97	11.63	13.37
		NRMSE(%)	16.85	17.44	18.07	18.70	18.43	15.41	18.17

Note: LAT is the abbreviation for look-ahead time. DA: day-ahead.

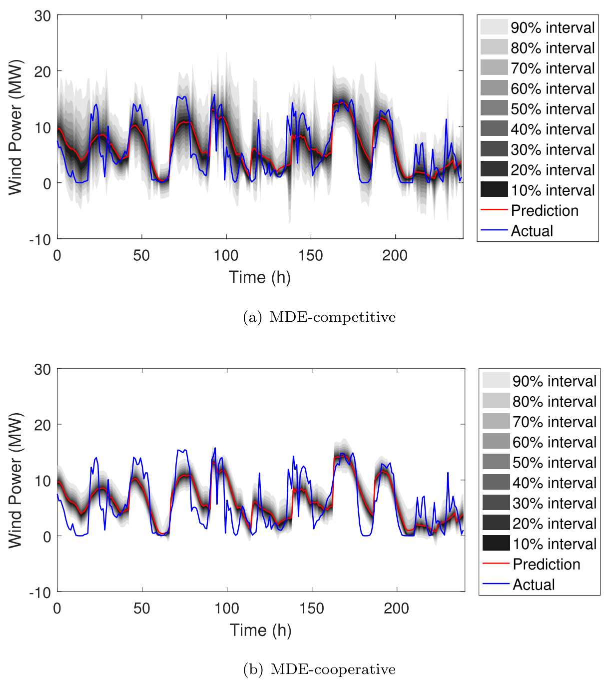


Fig. 4. Day-ahead MDE-competitive and MDE-cooperative forecasts at the C3 site.

Instead, the cooperative model that refines the 1HA forecasts from different single models performs better. By contrast, the competitive method performs better when the deterministic forecasting accuracy is relatively worse. The accuracy of deterministic forecasts is a major factor that affects the performance (i.e., pinball loss) of the final probabilistic forecasts.

In addition, the one-step optimization in MDE-competitive and two-step optimization in MDE-cooperative also affects the probabilistic forecasting performance. The MDE-competitive method first ensembles the distribution members together, and then uses this combined distribution model to generate probabilistic forecasts, which optimizes the standard deviations and weight parameters simultaneously. The MDE-cooperative method first generates different single quantile forecasts, and then ensembles the single quantile forecasts to refined quantile forecasts. This two-step optimization (i.e., determine standard deviations first and then weights) may introduce more uncertainty compared to the one-step optimization. Similarly, the surrogate model accuracy and forecasting look-ahead times may also affect the final probabilistic forecasting results.

With estimated parameters through pinball loss optimization and surrogate modeling, the quantiles $q_{t,1}$, $q_{t,2}$, …, $q_{t,99}$ can be calculated. To better visualize probabilistic forecasts, the 99 quantiles are converted into nine prediction intervals (PIs) I_t^{β} ($\beta = 10, \ldots, 90$) in a 10% increment. I_t^{β} stands for the PIs of wind power with $1 - \beta$ nominal coverage rate, which can be expressed through the lower bound $q_t^{\alpha_L}$ and the upper bound $q_t^{\alpha_U}$ as $[q_t^{\alpha_L}, q_t^{\alpha_U}]$, where α_L and α_U are lower and upper nominal proportions, respectively, which equal to $\beta/2$ and $(1 - \beta/2)$, correspondingly. Fig. 4(a) and (b) show the MDE-competitive and MDE-cooperative based day-ahead probabilistic wind power forecasts at the C3 site from 2012-08-09 to 2012-08-19. It is seen from Fig. 4 that the PIs of the MDE-competitive method cover the actual and forecasted wind power better than the MDE-cooperative method. For the MDE-cooperative method, even it has narrower PIs, some part of the PIs could not cover the observations. Therefore, the MDE-competitive method is more reliable than the MDE-cooperative method in day-ahead forecasts. The width of the prediction interval varies with the wind power variability. When the wind power fluctuates more frequently, the PIs tend to be wider, and thereby the uncertainty in wind power forecasts is relatively higher.

In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the probabilistic forecasting accuracy.

4.5.1. *Reliability*

Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies [18]:

$$RE = \left[\frac{\xi^{(1-\beta)}}{T} - (1 - \beta)\right] \times 100\% \tag{27}$$

where T is the number of test samples, and $\xi^{(1-\beta)}$ is the number of times that the actual test samples lie within the βth prediction interval. With measured empirical coverages, a reliability diagram can be plotted to describe the quantile forecast series with different nominal proportions. A reliability plot shows whether a given method tends to systematically underestimate or overestimate the uncertainty. In this study, the nominal coverage rate ranges from 10% to 90% with a 10% increment. The blue curves in Fig. 7 show the 1- to 6-HA and 24-HA reliability plots of the probabilistic forecasts at the selected C7 site with different forecasting models. A forecast

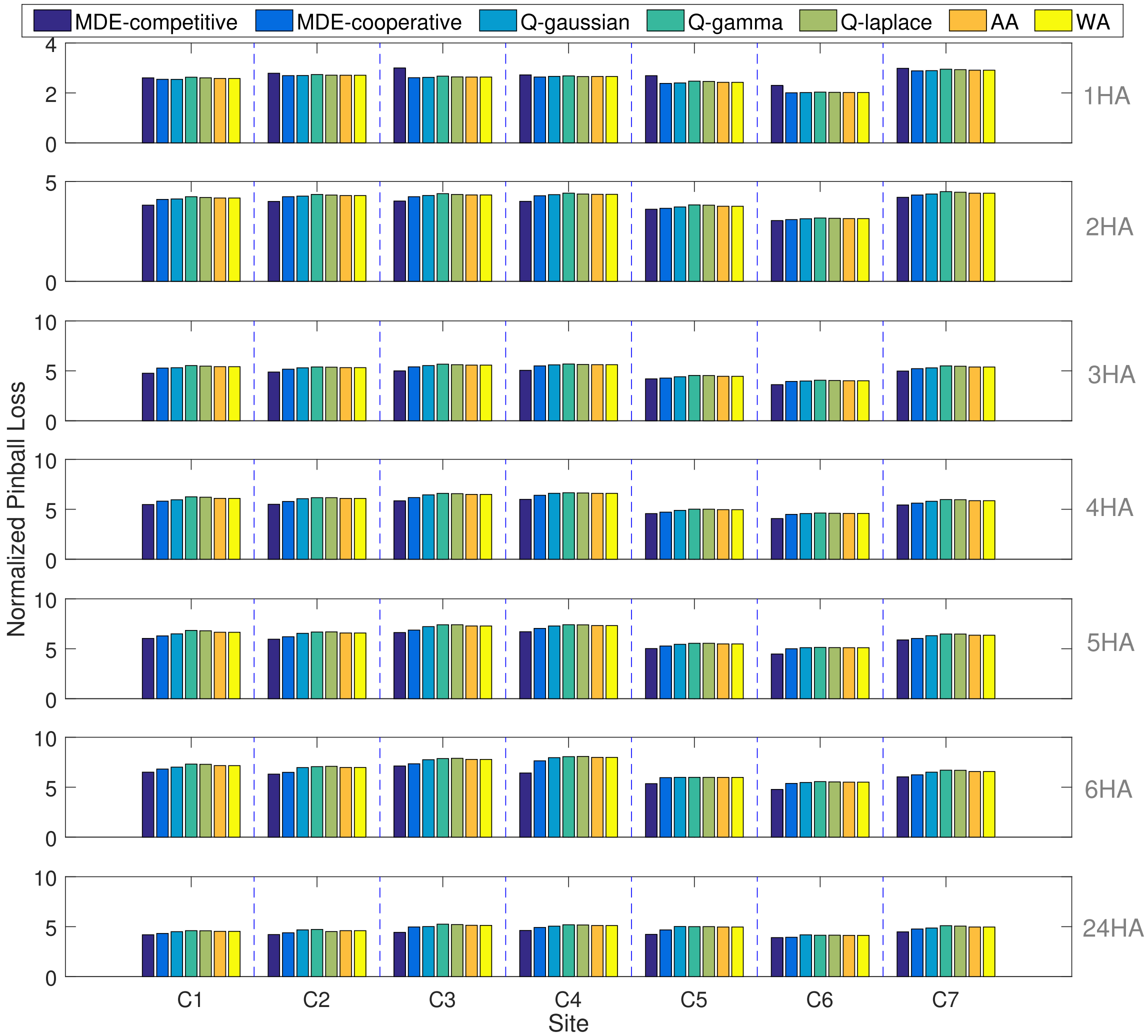


Fig. 5. Normalized pinball loss of different ensemble methods and single methods.

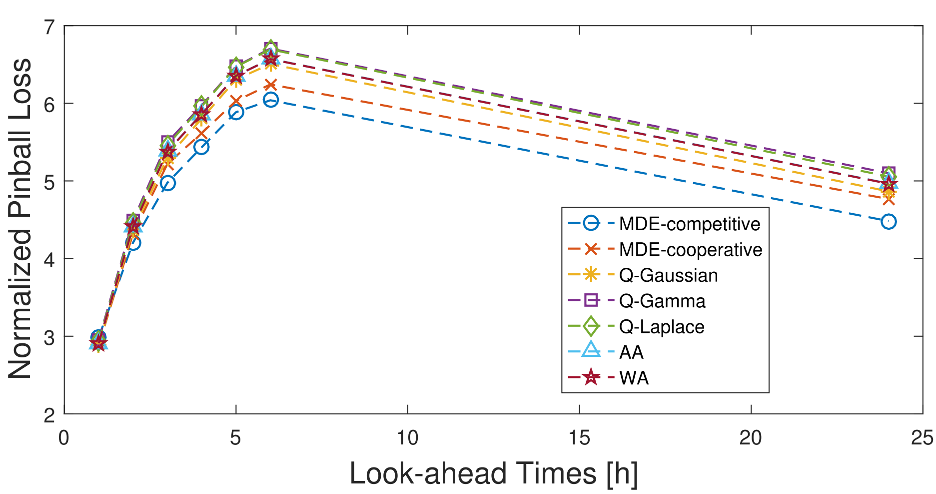
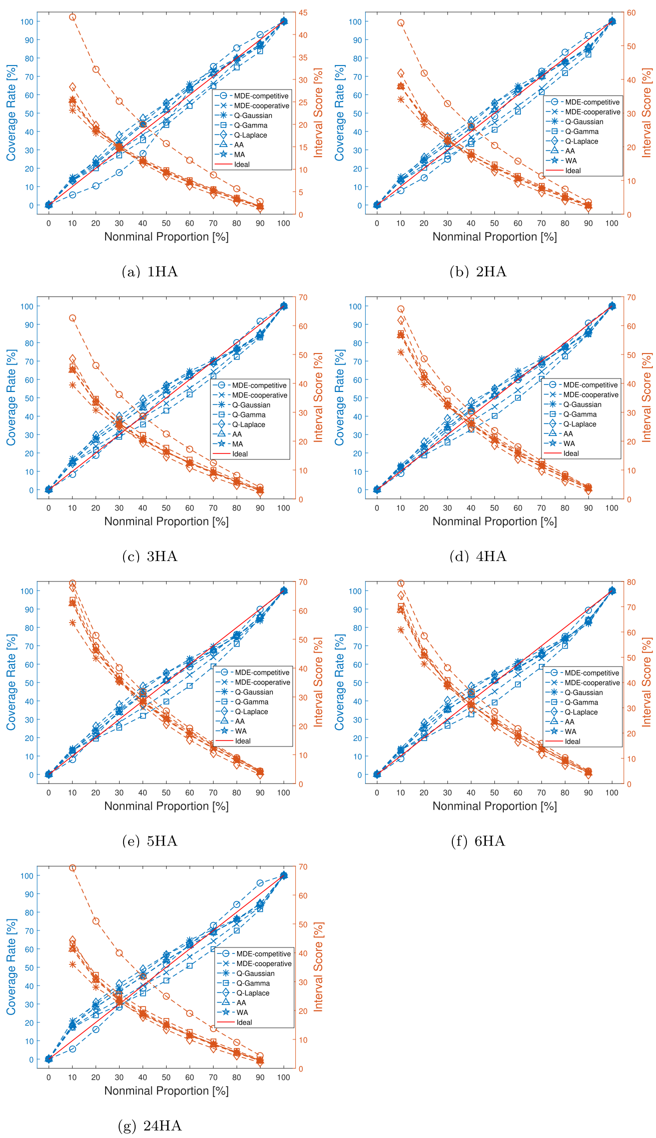


Fig. 6. Normalized pinball loss of different models with different look-ahead hours at the C7 site.

presents better reliability when the curve is closer to the diagonal. Overall it is seen from Fig. 7 that the proposed MDE ensemble framework (both competitive and cooperative) has better reliability performance than the baseline methods. Note that for 1-HA forecasts, the MDE-cooperative method has the best reliability performance and the MDE-competitive method has the best reliability performance for 2HA-6HA and 24HA forecasts. It is also seen that the two baseline ensemble methods (AA and MA) have better



(a) 1HA (b) 2HA (c) 3HA (d) 4HA (e) 5HA (f) 6HA (g) 24HA

Fig. 7. Reliability and sharpness comparison of different models at the C7 site with different look-ahead times (blue lines: reliability; orange lines: sharpness). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

performance than the single-distribution models. Overall the results show the effectiveness of ensemble modeling for enhancing probabilistic forecasts. The reliability plots of all the case study sites are provided in Appendix A.

4.5.2. Sharpness

Sharpness indicates the capacity of a forecasting system to forecast wind power with extreme probability [57]. This criterion evaluates the predictions independently of the observations, which gives an indication of the level of usefulness of the predictions. For example, a system that provides only uniformly distributed predictions is less useful for decision-making under uncertainty. Predictions with perfect sharpness are discrete predictions with a probability of one (i.e., deterministic predictions). The averaged PIs size δ^{β} at nominal coverage rate (1-β) can be expresses as:

$$\delta^{\beta} = \frac{1}{T}\sum_{t=1}^{T}\left(q_t^{\alpha_U} - q_t^{\alpha_L}\right) \times 100\% \tag{28}$$

An interval score $S^{\beta}(\hat{x}_t)$ is defined to reward narrow PIs and penalize the targets which are out of PIs range as follows [58]:

$$S^{\beta}(\hat{x}_t) = \begin{cases} 2\beta\delta^{\beta}(\hat{x}_t) + 4\left(q_t^{\alpha_L} - \hat{x}_t\right), & \hat{x}_t < q_t^{\alpha_L} \\ 2\beta\delta^{\beta}(\hat{x}_t), & \hat{x}_t \in I_t^{\beta} \\ 2\beta\delta^{\beta}(\hat{x}_t) + 4\left(\hat{x}_t - q_t^{\alpha_U}\right), & \hat{x}_t > q_t^{\alpha_U} \end{cases} \tag{29}$$

The sharpness plots of the proposed ensemble system (i.e., MDE-competitive and MDE-cooperative); and baseline models (Q-Laplace, Q-Gaussian, Q-Gaussian, AA, and WA ensemble methods) at the C7 site are compared in Fig. 7. The expected interval score decreases with increasing nominal coverage rate, and the sharpness of the MDE-competitive and MDE-cooperative models are slightly worse than some of the baseline models (e.g., Q-laplace and Q-gamma). It is mainly because the reliability and sharpness are two complementary metrics, and the improvement of reliability will sacrifice the sharpness to some extent. Overall, the interval size of the MDE forecasts ranges from 10% to 120%, which indicates reasonable sharpness. The sharpness plots of all the case study sites are provided in Appendix A.

5. Conclusion

In this paper, a multi-distribution ensemble (MDE) probabilistic wind forecasting framework was developed, with both competitive and cooperative ensemble strategies. Three types of predictive distributions (i.e., Gaussian, Gamma, and Laplace) were adopted as ensemble members. The optimal ensemble weight parameters and standard deviations of different predictive distributions were determined by minimizing the sum of pinball loss in the training stage. A set of surrogate models of the optimal parameters were constructed to estimate the pseudo-optimal parameters in the forecasting stage. Case studies at 7 selected sites show that:

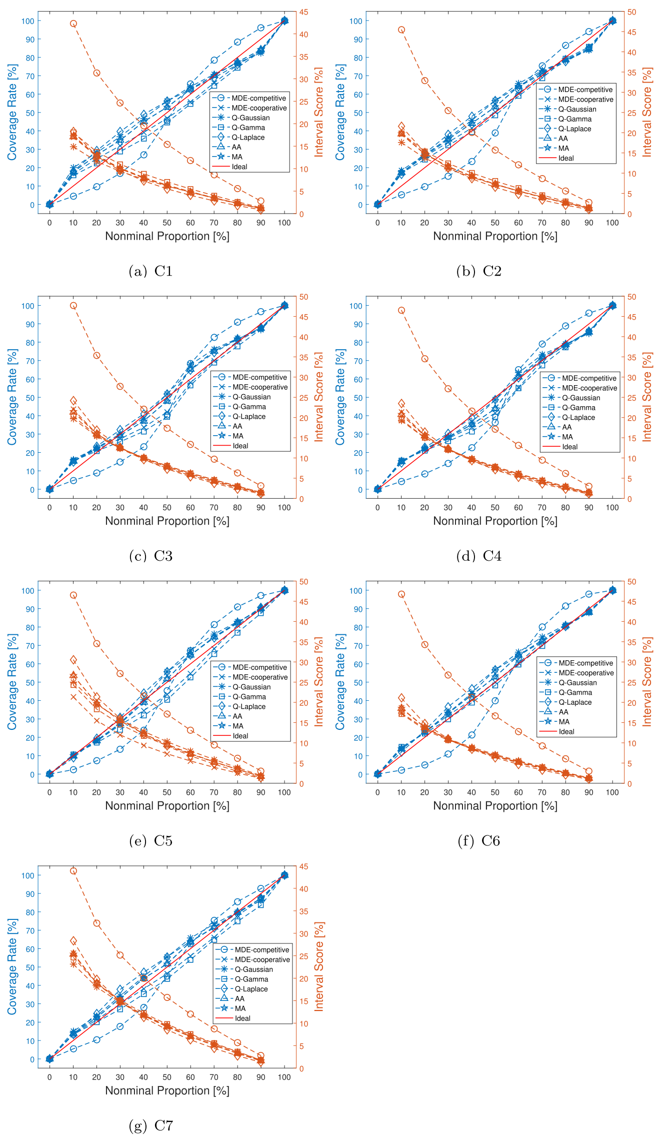
1. The developed MDE probabilistic forecasting framework could reduce the pinball loss score by up to 20.5% compared to benchmark models.
2. The MDE framework is robust under different forecasting time horizons at different locations.
3. The MDE-competitive method performs better in 2HA-6HA and 24HA forecasts and the MDE-cooperative method performs better in 1HA forecasts.
4. Both competitive and cooperative MDE methods have shown better reliability than single-distribution models and benchmark ensemble models. The sharpness intervals size of MDE forecasts ranges from 10% to 120%, which indicates reasonable sharpness.

Potential future work will: (i) integrate more effective distribution types into the MDE framework, and (ii) explore how the number and type of ensemble members affect the probabilistic forecasting results.

Acknowledgement

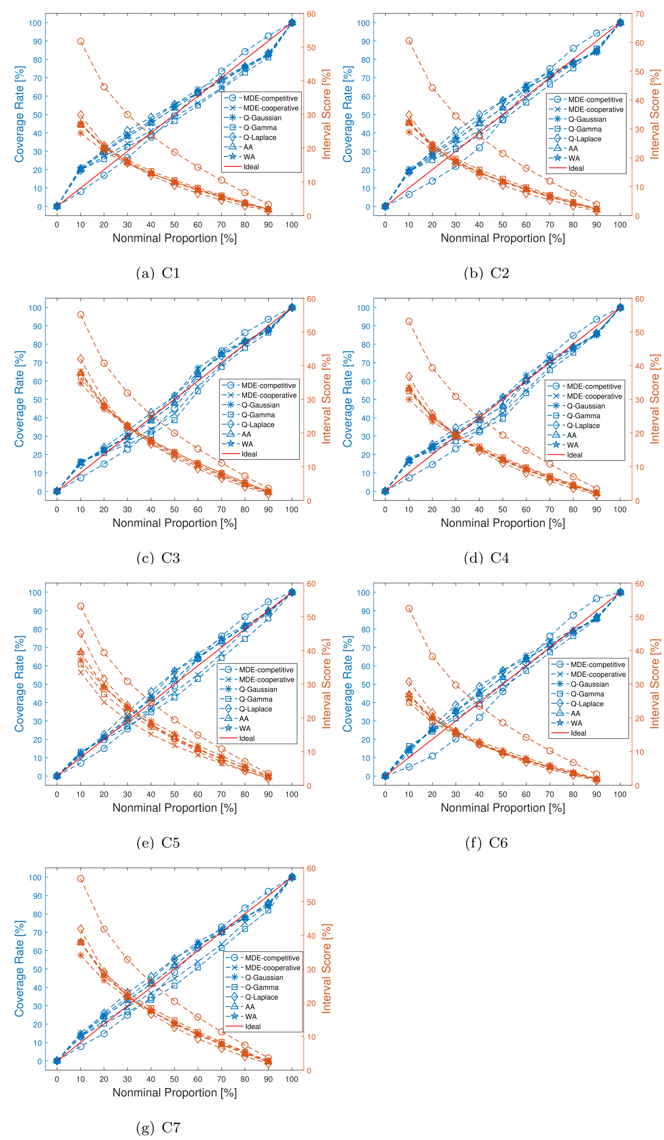
This work was supported by the National Renewable Energy Laboratory under Subcontract No. XAT-8-82151-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308).

Appendix A. Reliability and sharpness plots of case study sites at different look-ahead times



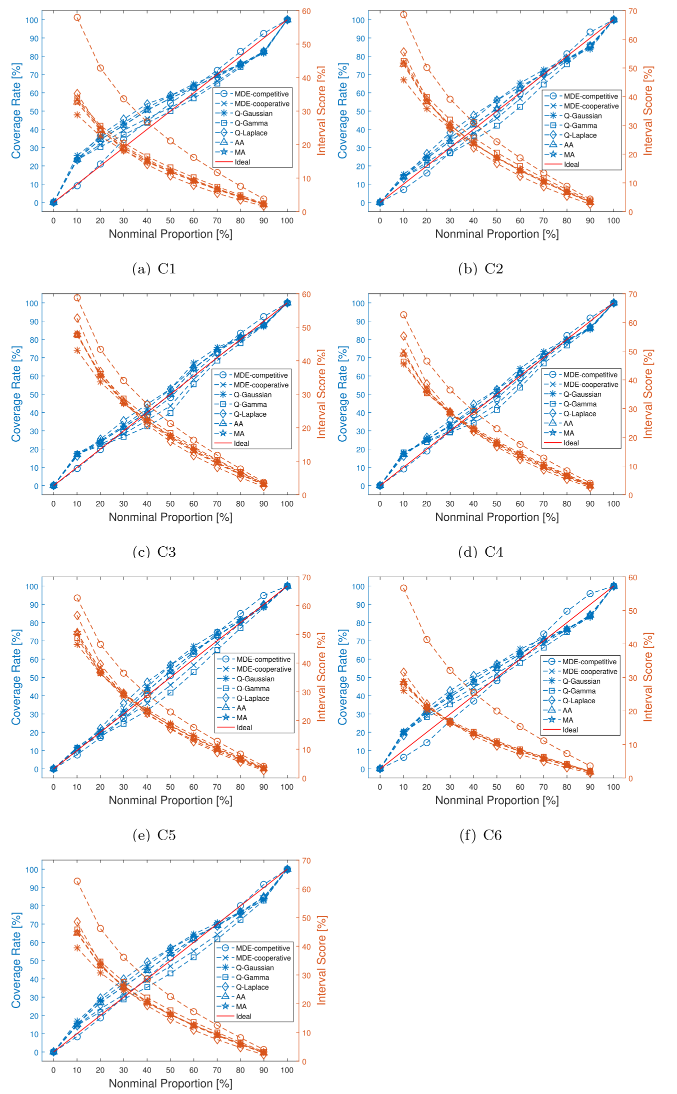
(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.8. 1HA reliability and sharpness comparison of different models at the 7 sites.



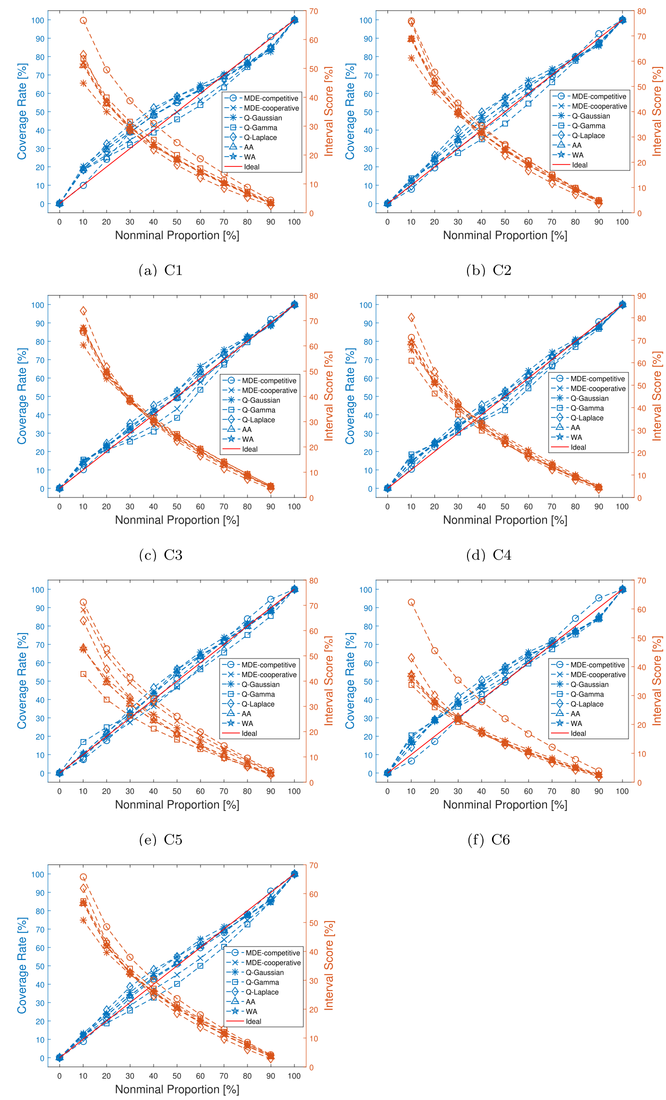
(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.9. 2HA reliability and sharpness comparison of different models at the 7 sites.



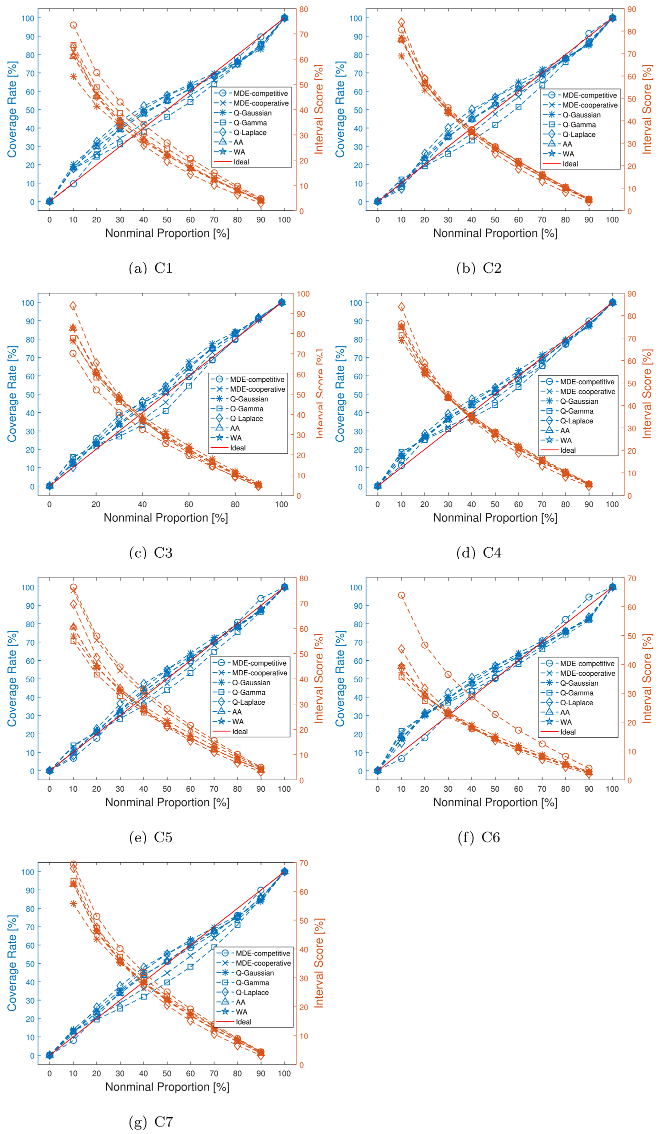
(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.10. 3HA reliability and sharpness comparison of different models at the 7 sites.



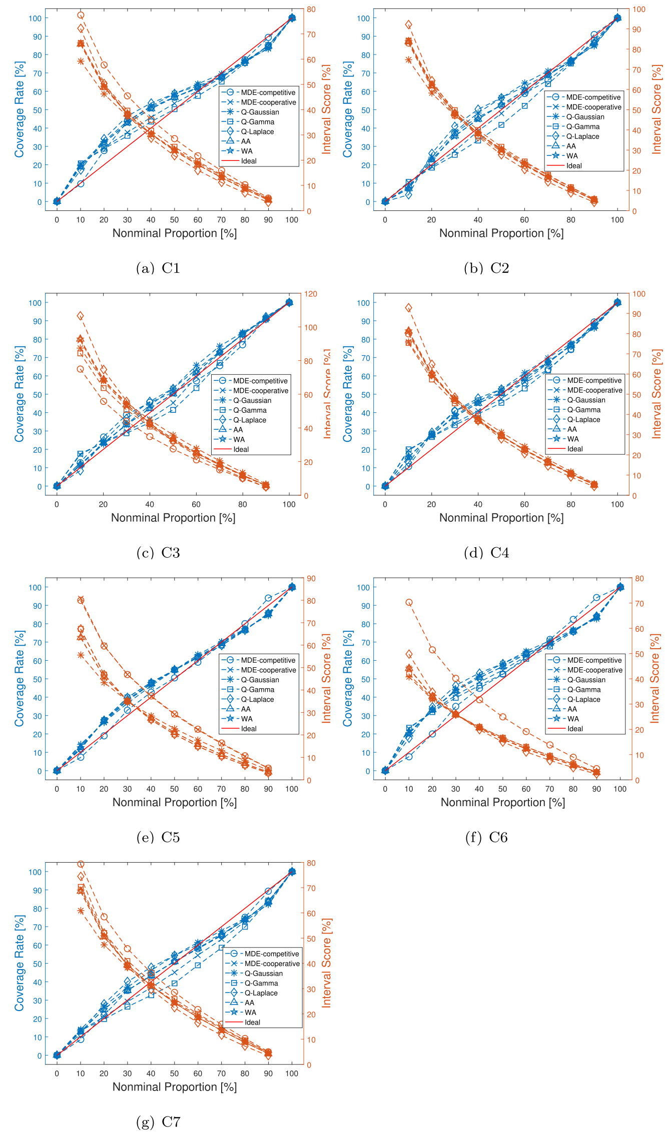
(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.11. 4HA reliability and sharpness comparison of different models at the 7 sites.



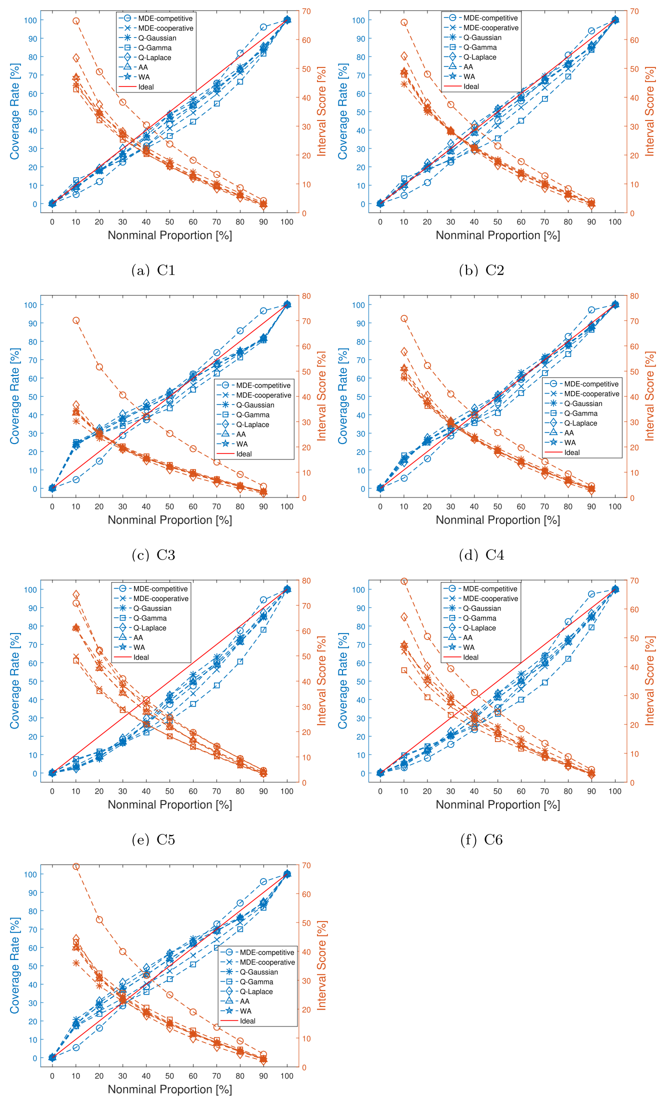
(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.12. 5HA reliability and sharpness comparison of different models at the 7 sites.



(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.13. 6HA reliability and sharpness comparison of different models at the 7 sites.



(a) C1 (b) C2 (c) C3 (d) C4 (e) C5 (f) C6 (g) C7

Fig. A.14. 24HA reliability and sharpness comparison of different models at the 7 sites.

References

[1] A. Bracale, P. De Falco, An advanced bayesian method for short-term probabilistic forecasting of the generation of wind power, Energies 8 (9) (2015) 10293–10314.

[2] J.M. Freedman, J. Manobianco, J. Schroeder, B. Ancell, K. Brewster, S. Basu, V. Banunarayanan, B.-M. Hodge, I. Flores, The Wind Forecast Improvement Project (Wfip): A Public/private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations. The Southern Study Area, Final Report, Tech. Rep, AWS Truepower, LLC, Albany, NY (United States), 2014.

[3] J.U. Jørgensen, C. Möhrlen, Aeso wind power forecasting pilot project, 2008. Tech. rep., Technical report, Ebberup, Denmark.

[4] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, The state-of-the-art in short-term prediction of wind power: a literature overview, Tech. rep., ANEMOS. plus, 2011.

[5] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R.J. Bessa, V. Miranda, Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: a case study of Illinois, IEEE Trans. Sustain. Energy 4 (1) (2012) 250–261.

[6] D. Sari, S. M. Ryan, Reliability of wind power scenarios and stochastic unit commitment cost, Industrial and Manufacturing Systems Engineering Technical Reports and White Papers 4.

[7] X. Xia, Y. Lin, J. Fang, W. Jiang, Y. He, N. Shang, Y. Ding, Optimal dispatch model considering the bidding uncertainty of wind power in deregulated power system, J. Eng. (13) (2017) 2389–2394, 2017.

[8] P. Pinson, C. Chevallier, G.N. Kariniotakis, Trading wind generation from short-term probabilistic forecasts of wind power, IEEE Trans. Power Syst. 22 (3) (2007) 1148–1156.

[9] D. Lee, H. Shin, R. Baldick, Bivariate probabilistic wind power and real-time price forecasting and their applications to wind power bidding strategy development, IEEE Trans. Power Syst. 33 (6) (2018) 6087–6097, https://doi.org/10.1109/TPWRS.2018.2830785.

[10] Y. Wang, Z. Zhou, A. Botterud, K. Zhang, Optimal wind power uncertainty intervals for electricity market operation, IEEE Trans. Sustain. Energy 9 (1) (2017) 199–210.

[11] M.A. Matos, R.J. Bessa, Setting the operating reserve using probabilistic wind power forecasts, IEEE Trans. Power Syst. 26 (2) (2011) 594–603.

[12] E. Du, N. Zhang, C. Kang, B. Kroposki, H. Huang, M. Miao, Q. Xia, Managing wind power uncertainty through strategic reserve purchasing, IEEE Trans. Power Syst. 32 (4) (2017) 2547–2559.

[13] Y. Lin, M. Yang, C. Wan, J. Wang, Y. Song, A multi-model combination approach for probabilistic wind power forecasting, IEEE Trans. Sustain. Energy 10 (1) (2018) 226–237.

[14] J. Shi, G. Zhang, W.-J. Lee, Evaluating the impact of wind power probabilistic forecasting on very-short-term generation scheduling for wind-storage combined generation system, in: Industrial and Commercial Power Systems Technical Conference (I&CPS), 2018 IEEE/IAS 54th, IEEE, 2018, pp. 1–7.

[15] I.N. Moghaddam, B.H. Chowdhury, M. Doostan, Optimal sizing and operation of battery energy storage systems connected to wind farms participating in electricity markets, IEEE Trans. Sustain. Energy (2018) 1, https://doi.org/10.1109/TSTE.2018.2863272, 1.

[16] M. Sun, C. Feng, J. Zhang, E.K. Chartan, B.-M. Hodge, Probabilistic short-term wind forecasting based on pinball loss optimization, in: 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, 2018, pp. 1–6.

[17] H.A. Nielsen, H. Madsen, T.S. Nielsen, Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts, Wind Energy: Int. J. Prog. Appl. Wind Power Convers. Technol. 9 (1–2) (2006) 95–108.

[18] J. Juban, N. Siebert, G.N. Kariniotakis, Probabilistic short-term wind power forecasting for the optimal management of wind generation, in: Power Tech, 2007 IEEE Lausanne, IEEE, 2007, pp. 683–688.

[19] C. Wan, J. Lin, J. Wang, Y. Song, Z.Y. Dong, Direct quantile regression for nonparametric probabilistic forecasting of wind power generation, IEEE Trans. Power Syst. 32 (4) (2017) 2767–2778.

[20] R.J. Bessa, V. Miranda, A. Botterud, Z. Zhou, J. Wang, Time-adaptive quantile-copula for wind power probabilistic forecasting, Renew. Energy 40 (1) (2012) 29–39.

[21] G. Cocchi, L. Galli, G. Galvan, M. Sciandrone, M. Cantù, G. Tomaselli, Machine learning methods for short-term bid forecasting in the renewable energy market: a case study in Italy, Wind Energy 21 (5) (2018) 357–371.

[22] C. Wan, Z. Xu, P. Pinson, Direct interval forecasting of wind power, IEEE Trans. Power Syst. 28 (4) (2013) 4877–4878.

[23] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, Y.-t. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, Appl. Energy 188 (2017) 56–70.

[24] Y. Xu, L. Shi, Y. Ni, Deep-learning-based scenario generation strategy considering correlation between multiple wind farms, J. Eng. (13) (2017) 2207–2210, 2017.

[25] M. Lange, On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors, J. Sol. Energy Eng. 127 (2) (2005) 177–184.

[26] H. Bludszuweit, J.A. Domínguez-Navarro, A. Llombart, Statistical analysis of wind power forecast error, IEEE Trans. Power Syst. 23 (3) (2008) 983–991.

[27] P. Pinson, Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions, J. R. Stat. Soc.: Series C (Applied Statistics) 61 (4) (2012) 555–576.

[28] D. Ke, C. Chung, Y. Sun, A novel probabilistic optimal power flow model with uncertain wind power generation described by customized Gaussian mixture model, IEEE Trans. Sustain. Energy 7 (1) (2016) 200–212.

[29] Y. Wang, Q. Hu, D. Meng, P. Zhu, Deterministic and probabilistic wind power forecasting using a variational bayesian-based adaptive robust multi-kernel regression model, Appl. Energy 208 (2017) 1097–1112.

[30] D. Opitz, R. Maclin, Popular ensemble methods: an empirical study, J. Artif. Intell. Res. 11 (1999) 169–198.

[31] Y. Ren, P. Suganthan, N. Srikanth, Ensemble methods for wind and solar power forecasting—a state-of-the-art review, Renew. Sustain. Energy Rev. 50 (2015) 82–91.

[32] L. Breiman, Bagging predictors machine learning, 24, 123–140.

[33] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227.

[34] D.L. Shrestha, D.P. Solomatine, Experiments with adaboost. rt, an improved boosting scheme for regression, Neural Comput. 18 (7) (2006) 1678–1710.

[35] Y. Ren, X. Qiu, P.N. Suganthan, Empirical mode decomposition based adaboost-backpropagation neural network method for wind speed forecasting, in: Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on, IEEE, 2014, pp. 1–6.

[36] J. Wu, B. Zhang, K. Wang, Application of adaboost-based bp neural network for short-term wind speed forecast, Power Syst. Technol. 36 (9) (2012) 221–225.

[37] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model

methodology with deep feature selection for short-term wind forecasting, Appl. Energy 190 (2017) 1245–1257.
[38] C. Feng, M. Sun, M. Cui, E.K. Chartan, B.-M. Hodge, J. Zhang, Characterizing forecastability of wind sites in the United States, Renew. Energy 133 (2019) 1352–1365.
[39] H. Liu, H.-q. Tian, Y.-f. Li, Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction, Appl. Energy 98 (2012) 415–424.
[40] A. Lojowska, D. Kurowicka, G. Papaefthymiou, L. van der Sluis, Advantages of arma-garch wind speed time series modeling, in: Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on, IEEE, 2010, pp. 83–88.
[41] R. Ye, Ensemble Time Series Forecasting with Applications in Renewable Energy, Ph.D. thesis, 2016.
[42] S. Wang, N. Zhang, L. Wu, Y. Wang, Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method, Renew. Energy 94 (2016) 629–636.
[43] A. Gensler, B. Sick, Forecasting wind power - an ensemble technique with gradual competitive weighting based on weather situation, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 4976–4984.
[44] G. Zhang, Y. Wu, K.P. Wong, Z. Xu, Z.Y. Dong, H.H.-C. Iu, An advanced approach for construction of optimal wind power prediction intervals, IEEE Trans. Power Syst. 30 (5) (2014) 2706–2715.
[45] D. Kim, J. Hur, Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method, Energy 157 (2018) 211–226.
[46] M. Sun, C. Feng, E.K. Chartan, B.-M. Hodge, J. Zhang, A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization, Appl. Energy 238 (2019) 1497–1505.
[47] C. Feng, J. Zhang, Short-term load forecasting with different aggregation strategies, in: ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2018.
[48] C. Feng, M. Sun, J. Zhang, Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection, IEEE Trans. Smart Grid (2019).
[49] S.A. Fatemi, A. Kuh, M. Fripp, Parametric methods for probabilistic forecasting of solar irradiance, Renew. Energy 129 (2018) 666–676, https://doi.org/10.1016/j.renene.2018.06.022. http://www.sciencedirect.com/science/article/pii/S0960148118306530.
[50] C.H. Hansen, C.J. Doolan, K.L. Hansen, Wind Farm Noise: Measurement, Assessment, and Control, John Wiley & Sons, 2017.
[51] L. Scrucca, et al., Ga: a package for genetic algorithms in r, J. Stat. Softw. 53 (4) (2013) 1–37.
[52] C. Draxl, A. Clifton, B.-M. Hodge, J. McCaa, The wind integration national dataset (wind) toolkit, Appl. Energy 151 (2015) 355–366.
[53] G. Ridgeway, M.H. Southworth, S. RUnit, Package 'gbm', Viitattu, 10, 2013, p. 40, 2013.
[54] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R. News 2 (3) (2002) 18–22.
[55] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, A. Weingessel, Misc functions of the department of statistics (e1071), tu wien, R package 1 (2008) 5–24.
[56] B. Ripley, W. Venables, M.B. Ripley, Package 'nnet', R Package Version 7, 2016, pp. 3–12.
[57] C. Gallego-Castillo, R. Bessa, L. Cavalcante, O. Lopez-Garcia, On-line quantile regression in the rkhs (reproducing kernel hilbert space) for operational probabilistic forecasting of wind power, Energy 113 (2016) 355–365.
[58] M. Cui, V. Krishnan, B. Hodge, J. Zhang, A copula-based conditional probabilistic forecast model for wind power ramps, IEEE Trans. Smart Grid (2018) 1, https://doi.org/10.1109/TSG.2018.2841932, 1.