Applied Energy 256 (2019) 113842

Contents lists available at ScienceDirect

Applied Energy

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

Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation

Mucun Sun, Cong Feng, Jie Zhang*

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

HIGHLIGHTS

- Develop aggregated probabilistic wind forecasting with spatio-temporal correlation.
- Gaussian mixture model is used to fit wind power probability density functions.
- Wind farm clustering is performed to improve the forecasting accuracy.
- The proposed method has improved pinball loss by up to 54% compared to benchmarks.

ARTICLE INFO

Keywords:
Probabilistic wind power forecasting
Spatio-temporal correlation
Aggregated probabilistic forecasting
Clustering
Pinball loss

ABSTRACT

Aggregated probabilistic wind power forecasting is important for power system operations. In this paper, an improved aggregated probabilistic wind power forecasting framework based on spatio-temporal correlation is developed. A Q-learning enhanced deterministic wind power forecasting method is used to generate deterministic wind power forecasts for individual wind farms. The spatio-temporal correlation between the member wind farms and the aggregated wind power is modeled by a joint distribution model based on the copula theory. The marginal distributions of actual aggregated wind power and forecasted power of member wind farms are built with Gaussian mixture models. Then, a conditional distribution of the aggregated wind power is deduced through the Bayesian theory, which is used for aggregated probabilistic forecasts. The effectiveness of the proposed aggregated probabilistic wind power forecasting framework is validated by using the Wind Integration National Dataset Toolkit. Numerical results of case studies at nine locations show that the developed aggregated probabilistic forecasting methodology has improved the pinball loss metric score by up to 54% compared to three benchmark models.

1. Introduction

The uncertain and variable nature of wind makes it challenging to be integrated into power systems, particularly at ever-increasing level of wind penetration. Studies have shown that the integration of geographically dispersed wind farms could reduce extreme power output, which is referred to as smoothing effect [1]. In addition, power produced from one wind farm at different times is typically temporally correlated [2]. It would be interesting to explore the impacts of spatio-temporal correlation on the performance of aggregated wind power forecasting. The benefits of spatio-temporal modeling for wind power forecasting at aggregated levels have been briefly discussed in [3]. In addition to wind power, spatio-temporal correlation modeling has also been applied to wind speed forecasting [4], load forecasting [5,6], and solar power forecasting [7].

1.1. Literature review

A collection of wind power forecasting methods that consider spatio-temporal correlation have been developed in the literature to assist power system planning and operations. Methods of spatio-temporal based wind power forecasting proposed in the recent literature can be generally classified into three groups: (1) Physical models, which are usually developed based on sophisticated meteorological information. For example, Pelikan et al. [8] developed a robust wind power forecasting model based on a combination of sigmoidal power functions with wind speed forecasts obtained from a mesoscale spatio-temporal refined numerical weather prediction (NWP) model. Physical models usually require high computation cost, and have better performance than purely statistical time series approaches in longer prediction time horizon (e.g., day-ahead and week-ahead) [9]. (2) Statistical models,

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

https://doi.org/10.1016/j.apenergy.2019.113842
Received 25 February 2019; Received in revised form 28 May 2019; Accepted 2 September 2019
Available online 01 October 2019

which quantify the relationship between geographically dispersed wind power time series. For example, Xie et al. [10] developed a short-term wind power forecasting model by leveraging the spatio-temporal correlation in wind speed and direction among different wind farms in west Texas. Statistical models are cost-saving since they do not require any data beyond historical wind power generation. However, the accuracy of statistical models drops with the increment of prediction time horizon. (3) Hybrid models, which combine the advantages of physical and statistical approaches to obtain globally optimal forecasts. For example, Qin et al. [11] proposed a hybrid wind power forecasting model based on long short-term memory network and deep learning neural network. Convolutional network and long short-term memory are used to exploit the spatio-temporal properties of wind farms. More research about spatio–temporal model based wind power forecasting has been studied in [12–14].

The spatio-temporal models discussed above only provide deterministic forecasts. To better account for the wind power uncertainty and variability, probabilistic wind power forecasts are needed. 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 [15]. Nonparametric approaches are distribution free, and their predictive distributions are estimated through observations. Quantile regression [16,17] and kernel density estimation (KDE) [18,19] are two of the most popular used methods for nonparametric probabilistic forecasting. Parametric approaches generally require low computational cost since a prior assumption of the predictive distribution shape is made before the parameter estimation. Parametric probabilistic wind power forecasting approaches have been widely used based on single distribution models [20,21] and mixture distribution models [22]. In contrast to traditional probabilistic wind power forecasting models, probabilistic wind forecasting technologies based on spatio-temporal effects have also been developed in the literature [23]. For example, Zhang et al. [24] used off-site information of geographically dispersed wind farms to capture spatio-temporal correlation and generated quantile forecasts. Then an Alternating Direction Method of Multipliers (ADMM)-based method was used to generate distributed probabilistic forecasts. Dowell et al. [25] proposed a probabilistic wind power forecasting method and the spatio-temporal correlation was captured through Sparse Vector Autoregression. A study conducted by Tang et al. [26] focused on multiple plants' probabilistic power forecasting, and a simple aggregation strategy was used. The marginal distribution of wind power in [26] is modeled by Gaussian distribution. Another study conducted by Li et al. [27] modeled the wind power uncertainties by using a particle filter algorithm. The aggregated wind power predictive distribution is obtained by mesoscale numerical weather prediction model and particle filter. Both [26,27] have aggregated all the wind farms in one step without clustering the farms. However, it is important to note that both [26,27] have found that considering spatio-temporal correlation among wind farms could improve the performance of probabilistic forecasts. Nevertheless, several challenges present in existing methods: (i) high dimensional matrices are involved in the spatio-temporal modeling, which adds additional computational burden; (ii) communication channels are needed for information transmission in [24], which might not be widely applicable at various spatial and temporal scales; (iii) the single logit-normal and Gaussian distribution in [25,26] may not be reliable for wind farm data with varying characteristics.

One of the most intuitive ways of modeling spatio-temporal correlation is to use a multivariate joint distribution. For example, Pinson et al. [28] has used a multivariate Gaussian distribution to describe the spatio-temporal relationship between forecasts of different wind farms. To address the challenge of modeling high dimensional multivariate non-Gaussian distributions, the Copula theory can be used. Based on the Sklar's theorem, the joint distribution can be modeled through univariate marginal-distribution functions and a Copula [29]. Copula theory has been widely used to characterize the dependence between different variables. For example, Tang et al. [26] used Copula to model correlation between wind farms and solar farms. Cui et al. [30] utilized high-dimensional Copula to couple the wind power ramp features among wind farms.

In the literature, the marginal distribution of wind power is commonly modeled by unimodal distributions such as Beta and Gamma [31] or nonparametric distributions such as KDE [32]. However, the unimodal distributions may not accurately quantify the variability of wind power and the nonparametric distributions are challenging to be solved analytically [30].

1.2. Research objective

To address the aforementioned limitations, in this paper we seek to develop an aggregated probabilistic wind power forecasting method by considering spatio-temporal correlation among wind farms with improved marginal distribution modeling of wind power and clustering. First, deterministic wind power forecasts are generated for individual wind farms (with selected deterministic forecasting methods). Second, the marginal distributions of the forecasted wind power of each wind farm and historical aggregated actual power are modeled through a Gaussian mixture model (GMM). In the third step, clustering is performed to divide all the wind farms into multiple clusters, with the aims of reducing computational cost and improving the overall aggregation accuracy. In the fourth step, Copula is adopted to build a spatio-temporal correlated joint model between the aggregated wind power and forecasted wind power of each wind farm. The conditional distribution of aggregated wind power is deduced through the Bayesian formula, which is then used (in conjunction with deterministic forecasts) to generate aggregated probabilistic wind power forecasts. The main contributions of this paper are as follows:

1. The proposed method uses Copula to model the spatio-temporal correlation among wind farms to improve the performance of aggregated probabilistic wind power forecasting;
2. A state-of-the-art deterministic forecasting method, i.e., Q-learning enhanced deterministic forecast, is adopted;
3. Wind farm clustering is performed to reduce the computational cost and also to improve the forecasting accuracy by taking advantage of the strong intra-cluster correlation;
4. A Gaussian mixture model is used to accurately fit the marginal distribution of wind power;
5. The cross-correlation among wind farm clusters is also modeled in the aggregation.

The rest of the paper is organized as follows. Section 2 describes the proposed aggregated probabilistic forecasting method, which consists of a deterministic forecasting method, marginal probability distribution modeling, clustering, and a spatio-temporal correlation based Copula model. Section 3 applies the developed spatio-temporal correlation based aggregated probabilistic forecasting method to nine wind farms and compares the proposed method with three benchmark aggregated probabilistic forecasting models. Concluding remarks and future work are discussed in Section 4.

2. Aggregated probabilistic forecasting framework

The overall framework of the developed conditional probabilistic aggregated wind power forecasting (cp-AWPF) framework is illustrated in Fig. 1. The five major steps are briefly described as follows:

1. Step 1: A Q-learning based ensemble deterministic forecasting method is adopted to select the best forecasting model from a pool of state-of-the-art machine learning based forecasting models at each time step, thus generating reinforced deterministic wind power

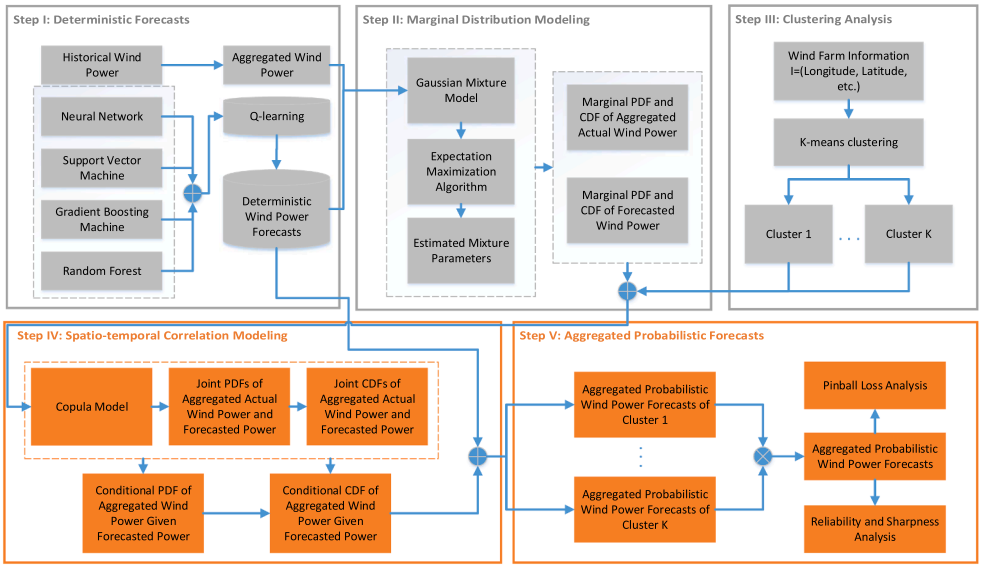


Fig. 1. Overall framework of the conditional probabilistic aggregated wind power forecasting.

forecasts for individual wind farms [33].

2. Step 2: GMM is used to fit the probability density functions (PDFs) of the historical aggregated actual wind power and the wind power forecasts at each wind farm.
3. Step 3: A prototype-based K-means clustering method is used to cluster wind farms into multiple non-overlapping clusters based on their spatial similarity.
4. Step 4: For each cluster, the joint distribution of historical aggregated actual wind power and wind power forecasts at each member wind farm is constructed based on the Copula theory.
5. Step 5: Aggregated probabilistic wind power forecasts are generated through the conditional distribution.

2.1. Q-learning enhanced deterministic forecasting

The developed spatio-temporal correlation based cp-AWPF framework is constructed based on deterministic forecasts. A large collection of methods have been developed in the literature to effectively generate deterministic wind forecasts. However, most of the existing deterministic methods are either selected based on the overall forecasting performance or ensembled by multiple models. Selecting a model based on the overall forecasting performance generally neglects the local performance of the selected model [33].

In this paper, a Q-learning enhanced deterministic forecasting method, developed in our previous work [34], is adopted. This method can choose the best forecasting model from a pool of state-of-the-art machine learning based forecasting models (i.e., artificial neural network, support vector machine, gradient boosting machine, and random forest) at each time step. To be more specific, the developed method trains Q-learning agents based on the rewards of transferring from the current model to the next model. For example, a Q-learning agent will receive a reward by transferring from the current forecasting model M_i to the next forecasting model M_j in each training step, from which the Q-learning agent will learn the optimal policy of the model selection. Then, this optimal policy will be applied to select the best model for forecasting in the next step based on the current model in the forecasting stage. The dynamic model selection process is expressed as [34]:

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

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

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

$$Q^{(e+1)}\left(s^e, a^e\right) = \left(1-\alpha\right)Q^e\left(s^e, a^e\right) + \alpha\left[R^e\left(s^e, a^e\right) + \gamma \max_{a\in\boldsymbol{A}} Q^e\left(s^{(e+1)}, a\right)\right] \tag{4}$$

where $\boldsymbol{S}$, $\boldsymbol{A}$, R, and Q are state space, action space, reward function, and Q-table in the dynamic model selection Markov Decision Process, respectively. s and a are possible state and action, respectively. I is the number of models (M) in the model pool. e is the episode index with the maximum of 100. $\alpha = 0.1$ is the learning rate that controls the aggressiveness of learning. $\gamma = 0.8$ is a discount factor that weights the future reward. 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. [34].

2.2. Wind power distribution modeling

The large variability in wind power poses challenges to accurately model the wind power through a unimodal distribution. Mixture distributions have been widely utilized in statistics to approximate multimodal distributions. To accurately characterize the variability of wind power, GMM is adopted in this paper to model the aggregated actual wind power and forecasted wind power. The probabilistic density function (PDF) of GMM is formulated as follows:

$$f_G(x|N_G, \omega_i, \mu_i, \sigma_i) = \sum_{i=1}^{N_G} \omega_i g_i(x|\mu_i, \sigma_i), \quad \mu_i \in U, \quad \sigma_i \in \Sigma, \quad \omega_i \in \Omega \tag{5}$$

where N_G is the number of mixture components, $U(\mu_i \in U)$ is the expected value vector, $\Sigma(\sigma_i \in \Sigma)$ is the standard deviation vector, and the $\Omega(\omega_i \in \Omega)$ is the weight vector. Each component $g(x;\mu_i, \sigma_i)$ follows a normal distribution, which can be expressed as:

$$g\left(x\middle|\mu, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{6}$$

The GMM distribution has two constraints: (1) the integral of Eq. (5) equals unity, and (2) the summation of weight parameters equals unity as well, which are expressed as follows:

$$\int_{-\infty}^{+\infty} f_G\left(x \middle| N_G, \omega_i, \mu_i, \sigma_i\right)dx = \int_{-\infty}^{+\infty} \sum_{i=1}^{N_G} \omega_i g_i\left(x \middle| \mu_i, \sigma_i\right)dx = 1 \tag{7}$$

$$\sum_{i=1}^{N_G} \omega_i = 1 \tag{8}$$

The parameters of GMM are estimated through the expectation maximization (EM) algorithm. The goal of EM is to maximize the likelihood function with respect to the parameters. More details about EM can be found in [35]. The CDF ($\widehat{F}$) corresponding to the estimated PDF is expressed as:

$$F_G\left(x \middle| N_G, \omega_i, \mu_i, \sigma_i\right) = \sum_{i=1}^{N_G} \left[\frac{\sqrt{\pi}}{2}\omega_i\sigma_i erf\left(\frac{\mu_i - x}{\sigma_i}\right)\right] + C \tag{9}$$

where C is an integral constant, and $erf(\cdot)$ is the Gaussian error function which is expressed as:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{10}$$

2.3. Clustering analysis

Clustering unlabelled wind farms is an unsupervised problem, which distinguishes and labels the type of training data based on the characteristics of the data itself. In this paper, we aim to reduce the dimensionality due to the large number of wind farms by dividing wind farms into multiple clusters, thereby reducing the computational cost. In addition, wind farm clustering could also strengthen the intra-cluster correlation, i.e., wind farms in the same cluster are more strongly correlated, which could potentially increase the accuracy of the aggregated wind power forecasting for each cluster and thus the overall forecasting accuracy.

2.3.1. Unsupervised clustering based on spatial correlation

In this paper, K-means is adopted to divide the member wind farms into multiple non-overlapping clusters based on wind farm characteristics. K-means is a widely used unsupervised clustering algorithm. Given a dataset $\psi = x_1, x_2...,x_n$ with n instances, the K-means algorithm divides the dataset into K disjoint clusters as $\psi = C_1, C_2, ..., C_k$. The objective function of K-means is:

$$min \sum_{k=1}^{K} \sum_{x \in C_k} d\left(x, m_k\right) \tag{11}$$

where m_k is the centroid of cluster C_k, $d(\cdot)$ measures the distance between x and the centroid of cluster C_k. In this paper, the Pearson's correlation coefficient is used as a distance metric to characterize the spatio-temporal correlation of wind farms.

2.3.2. Clustering assessment metric

It is challenging to validate unsupervised clustering results since the objects are unlabeled. Unsupervised clustering can be measured quantitatively based on cohesion and separation [36]. Cohesion measures how closely related the objects are in a cluster, while Separation measures how distinct or well-separated a cluster (or centroid) is from other clusters. In this paper, the Silhouette width (SW) is used to evaluate the clustering performance, which quantifies both cohesion and separation. It can be expressed as:

$$SW = \frac{1}{N}\sum_{i=1}^{N} \frac{\eta_b(i) - \eta_a(i)}{max(\eta_b(i) - \eta_a(i))} \tag{12}$$

where $\eta_a(i)$ is the average distance between object i and all other data in the same cluster, and $\eta_b(i)$ is the smallest average distance between object i to other objects in the neighbour cluster. The SW value ranges from -1 to $+1$, where $SW = +1$ indicates desired clustering, while $SW = -1$ indicates undesired clustering.

2.4. Wind farm spatio-temporal correlation modeling

Once the marginal wind power distribution of a single wind farm is defined, the spatio-temporal correlation among wind farms in each cluster can be modeled by Copula. Copula is one of the most widely used methods for modeling the dependency among random variables. Given a cluster, the aggregated actual wind power p^{Σ} is expressed as

$$p^{\Sigma} = \sum_{i=1}^{N} p_i \tag{13}$$

where p_i is the actual wind power of the ith wind farm, and N is the total number of wind farms to be aggregated in the cluster.

2.4.1. Spatial correlation

Suppose $\hat{p}_i^t$ is the t-hour-ahead (tHA) forecasted wind power at the ith wind farm, $f(\hat{p}_i^t)$ and $F(\hat{p}_i^t)$ denote the corresponding marginal PDF and marginal cumulative distribution function (CDF), respectively. The spatial correlation Θ_S among the N wind farms is modeled as:

$$F(\hat{P}^t) = C(F(\hat{p}_1^t), ...,F(\hat{p}_N^t)) \tag{14}$$

where $\hat{P}^t = \{\hat{p}_1^t, \hat{p}_1^t, ...,\hat{p}_N^t\}$ is a N-dimension vector denoting the tHA wind power forecasts of the N wind farms, and $C(\cdot)$ is the Copula function.

2.4.2. Temporal correlation

In addition to the spatial correlation among wind farms, the forecasted wind power with different look ahead times are typically temporally correlated [37]. Such temporal correlation Θ_T is modeled as:

$$F(\hat{P}^t, \hat{P}^{t+1}) = C(F(\hat{p}_1^t), ...,F(\hat{p}_N^t), F(\hat{p}_1^{t+1}), ...,F(\hat{p}_N^{t+1})) \tag{15}$$

where, $\hat{P}^{t+1} = \{\hat{p}_1^{t+1}, \hat{p}_1^{t+1}, ...,\hat{p}_N^{t+1}\}$ is a N-dimension vector denoting the $(t+1)$HA wind power forecasts of the N wind farms.

2.4.3. Spatio-temporal correlation

Based on Eqs. (13)–(15), the joint CDF of the forecasted wind power at individual wind farms and the aggregated actual wind power (of all farms), $F(p^{\Sigma}, \hat{P}^t, \hat{P}^{t+1})$ can be modeled through their marginal CDFs and the Copula function with a spatio-temporal joint structure Θ_{S-T}, which is expressed as:

$$\begin{aligned} F(p^{\Sigma}, \hat{P}^t, \hat{P}^{t+1}) &= C(p^{\Sigma}, F(\hat{p}_1^t),...,F(\hat{p}_N^t), F(\hat{p}_1^{t+1}),...,F(\hat{p}_N^{t+1})) \\ &= C(p^{\Sigma}, J, K) \end{aligned} \tag{16}$$

where,

$$J = F(\hat{p}_1^t), ...,F(\hat{p}_N^t) \tag{17}$$

$$K = F(\hat{p}_1^{t+1}), ...,F(\hat{p}_N^{t+1}) \tag{18}$$

Similarly, the joint PDF of the forecasted wind power at individual member farms and the aggregated actual wind power (of all farms) is expressed as:

$$f\left(p^{\Sigma}, \hat{P}^t, \hat{P}^{t+1}\right) = c\left(F(p^{\Sigma}), J, K\right)\cdot f(p^{\Sigma})\cdot \prod_{i=1}^{N} f(\hat{p}_i^t) \tag{19}$$

where the marginal PDF is modeled by using the aforementioned GMM distribution based on historical actual and forecasting data. Then, the conditional joint PDF of the aggregated wind power given the power forecasts of all the member farms is deduced from the Bayesian formula, given by:

$$f\left(p^{\Sigma} \middle| \hat{P}^t, \hat{P}^{t+1}\right) = \frac{c(F(p^{\Sigma}), J, K)}{c(J, K)}\cdot f(p^{\Sigma}) \tag{20}$$

The conditional distribution of the aggregated wind power given all the member farms forecasts can be trained by using historical actual and forecasting data. With any given deterministic forecasts of individual

Table 1
Data summary of the selected 9 WIND Toolkit sites.

Site Name	Site ID	Lat.	Long.	Capacity (MW)	State
S1	4816	29.38	-100.37	16	TX
S2	8979	31.53	-95.62	16	TX
S3	10069	32.31	-98.26	16	TX
S4	10526	32.44	-100.55	16	TX
S5	1342	27.12	-97.86	16	TX
S6	2061	27.95	-97.40	14	TX
S7	9572	31.99	-100.118	16	TX
S8	10527	32.45	−100.41	2	TX
S9	11038	32.72	-100.92	6	TX

wind farms, we can use the copula model and the trained conditional PDF in Eq. (20) to calculate the conditional CDF. Then a large number of scenarios of the aggregated wind power of each cluster can be generated by sampling from the conditional CDF. The PDF of the aggregated wind power in each cluster can be deduced based on the scenarios. The final PDF of all aggregated wind farms is calculated through convolution. One prerequisite of convolution is that the variables being convoluted should be independent. If the clusters are cross-correlated with each other, a copula based dependent convolution (such as the one proposed by Zhang et al. [38]) could be used to characterize the cross-correlation among clusters. The probabilistic forecasts could also be represented in the form of quantiles and confidence intervals based on the PDF. The pseudocode of quantile forecasts based on inverse transform is illustrated in Algorithm 1.

Algorithm 1. Generate quantile forecasts based on inverse transform

Data: Deterministic wind power forecasts
Result: Quantile forecasts
1 Initialization: Obtain PDF of each single wind farm and aggregated wind power through GMM
2 Model spatio-temporal correlation among wind farms through a joint distribution by Copula
3 Calculate conditional PDF and CDF of the aggregated wind power
4 Sample from conditional CDF through inverse transform to generate a large number of aggregated wind power scenarios
5 Generate quantile forecasts based on the distribution of generated scenarios

3. Case studies and results

3.1. Data summary

The developed cp-AWPF framework was evaluated at 9 wind farms in Texas that were selected from the Wind Integration National Dataset (WIND) Toolkit [39]. 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 day-ahead (DA) forecasts from WIND Toolkit, 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 9 wind farms spans two years from January 1st 2011 to December 31st 2012. The data information at the selected 9 wind farms is briefly summarized in Table 1. For all the 9 locations, the first 3/4 of the data is used as training data. The number of scenarios generated from the conditional distribution is set as $N_s = 5{,}000$. The accuracy of the forecasts is evaluated by the remaining 1/4 of data. Though the developed cp-AWPF method is capable of generating forecasts at multiple forecasting horizons, only 1HA-6HA and day-ahead wind power forecasts are generated in this study.

3.2. Deterministic forecasting results

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

$$NMAE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{x}_t - x_t}{C_n}\right| \times 100\% \tag{21}$$

$$NRMSE = \frac{1}{C_n}\sqrt{\frac{\sum_{t=1}^{T}(\hat{x}_t - x_t)^2}{T}} \times 100\% \tag{22}$$

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_n is the capacity of the nth wind farm.

A smaller NRMSE or NMAE indicates better forecasting performance. The forecasting errors by using the Q-learning based deterministic forecasting 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. Day-ahead deterministic forecasts are provided in the WIND Toolkit dataset, which are generated from numerical weather prediction (NWP) models [40]. It is shown that the day-ahead NMAE and NRMSE are in the range of 11–14% and 15–19%, respectively. Overall, the accuracies of the Q-learning based 1HA to 6HA deterministic forecasts and the NWP-based day-ahead deterministic forecasts are reasonable.

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

Model	LAT	Metric	Site								
			S1	S2	S3	S4	S5	S6	S7	S8	S9
Q-learning	1HA	NMAE(%)	6.63	6.85	6.70	6.74	6.67	5.38	7.76	6.85	7.40
		NRMSE(%)	10.55	10.93	11.04	10.66	9.93	8.37	11.93	10.72	11.51
	2HA	NMAE(%)	10.72	11.08	11.20	11.28	10.36	8.34	10.90	11.15	11.92
		NRMSE(%)	15.86	16.33	16.93	16.64	14.60	12.31	17.14	16.31	17.32
	3HA	NMAE(%)	13.95	14.22	14.65	14.88	12.44	10.49	14.56	14.27	14.82
		NRMSE(%)	19.43	19.86	20.92	20.71	16.94	14.93	20.12	19.75	20.50
	4HA	NMAE(%)	16.43	16.76	17.69	18.00	13.84	12.17	16.28	17.16	16.87
		NRMSE(%)	21.98	22.22	23.92	23.83	18.56	16.85	21.79	22.68	22.41
	5HA	NMAE(%)	18.19	18.37	20.36	20.38	15.35	13.53	17.87	19.33	18.43
		NRMSE(%)	23.65	23.71	26.51	26.01	20.19	18.40	23.15	24.66	24.03
	6HA	NMAE(%)	19.85	19.60	21.94	22.63	16.51	14.70	18.41	20.93	19.60
		NRMSE(%)	25.09	24.85	27.82	27.82	21.31	19.47	23.55	26.03	24.78
NWP	DA	NMAE(%)	12.70	12.59	13.21	13.97	13.97	11.63	13.37	13.44	13.45
		NRMSE(%)	16.85	17.44	18.07	18.70	18.43	15.41	18.17	18.43	18.21

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

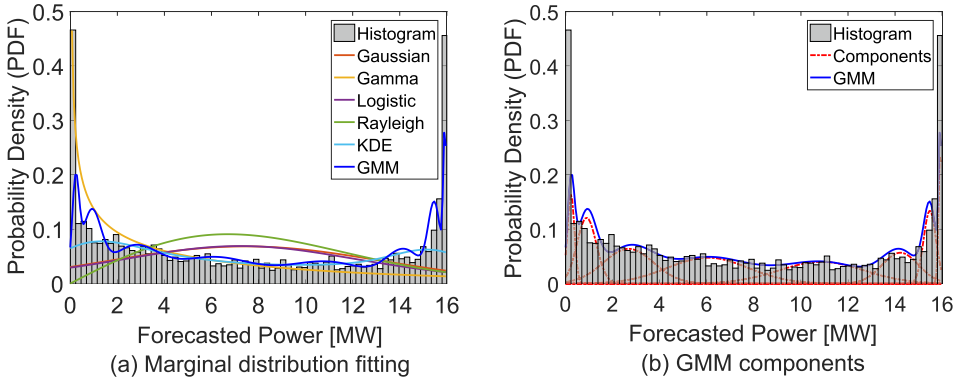


Fig. 2. Distributions of wind power forecasts at the S4 site.

Table 3
Information criteria of the estimated distribution.

Site	Metric	Distribution type					
		Gaussian	Gamma	Logistic	Rayleigh	KDE	GMM
S1	AIC	23760	23910	24070	23810	21367	**18512**
	BIC	23768	23918	24078	23818	21374	**18548**
S2	AIC	23840	22980	24080	24210	20081	**16802**
	BIC	23848	22988	24088	24218	20105	**16846**
S3	AIC	27450	26010	27960	29910	21900	**17664**
	BIC	27458	26018	27968	29918	21908	**17708**
S4	AIC	27140	25940	27620	29130	22180	**17026**
	BIC	27148	25948	27628	29138	22196	**17074**
S5	AIC	23870	23776	24160	23871	20124	**18442**
	BIC	23878	23784	24168	23819	20136	**18498**
S6	AIC	22360	21090	21610	22000	18900	**16430**
	BIC	22368	21098	21618	22008	18928	**16496**
S7	AIC	23610	23540	23920	24200	21103	**18900**
	BIC	23618	23548	23928	24208	21119	**18948**
S8	AIC	7338	6987	7699	7307	6692	**5483**
	BIC	7346	6995	7677	7315	6708	**5514**
S9	AIC	15660	15280	15970	15660	13214	**12090**
	BIC	15668	15288	15978	15668	13228	**12136**

Note: The best information criterion at each location is in boldface.

3.3. Performance of marginal wind power distribution models

Fig. 2(a) shows the marginal probability distributions of wind power forecasts from six distribution types at the S4 site (i.e., Gaussian, Gamma, Logistic, Rayleigh, KDE, and GMM distributions). For the Gaussian, Gamma, Logistic, and Rayleigh distributions, the parameters are estimated through the maximum likelihood (ML) method. The parameters of the KDE and GMM distributions are estimated by using the expectation maximization (EM) algorithm. Fig. 2(b) illustrates the mixture components of the GMM distribution at the S4 site. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are used to evaluate the estimated wind power distribution accuracy. For a given model with parameters θ, the AIC and BIC are defined as:

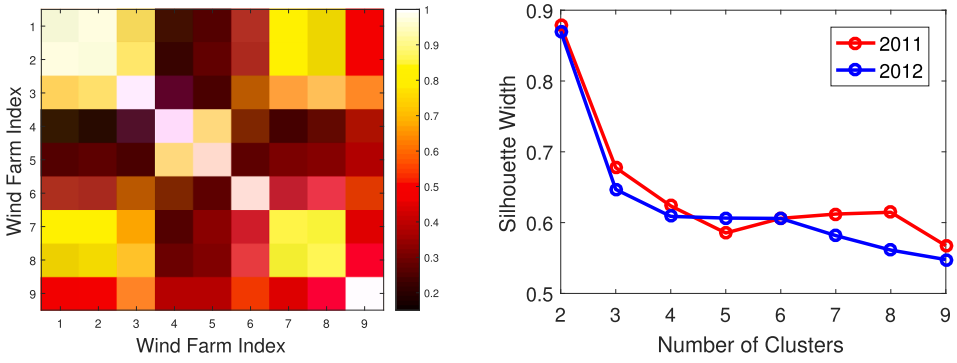
$$AIC = -2L(\hat{\theta}|x) + 2k \tag{23}$$

$$BIC = -2L(\hat{\theta}|x) + k\log n \tag{24}$$

where $\hat{\theta}$ is the maximum likelihood estimation of the distribution parameter, $L(\hat{\theta}|x)$ is the log-likehood function of $\hat{\theta}$, k is the number of parameters, and n is the length of observed data. The preferred model is the one with the lowest AIC and BIC [41]. The AIC and BIC values of the nine wind farms with different distribution types are summarized in Table 3. Results show that the GMM distribution outperforms other single distributions and KDE for modeling wind power forecasts. Therefore, GMM is adopted to fit the marginal wind power distribution within the Copula model.

3.4. Wind farm clustering results

The correlation among the selected 9 wind farms are visualized in Fig. 3(a). It is seen that the selected wind farms are strongly correlated. Fig. 3(b) shows the SW among all the wind farms in year 2011 and year 2012. It is seen that $K = 2$ always results in the highest SW, which indicates desired clustering. To better visualize the two clusters, the wind farms are plotted on a Texas map according to their geographical dispersion in Fig. 4. The two wind farm clusters are differentiated by red and blue colors. The red cluster is mainly located in north Texas and the blue cluster is located in south Texas. The mean correlation coefficient in the red cluster and the blue cluster are 0.65 and 0.58, respectively. In addition, the mean correlation coefficient without



(a) Heatmap (b) Sihoucoefficient

Fig. 3. Clustering analysis.

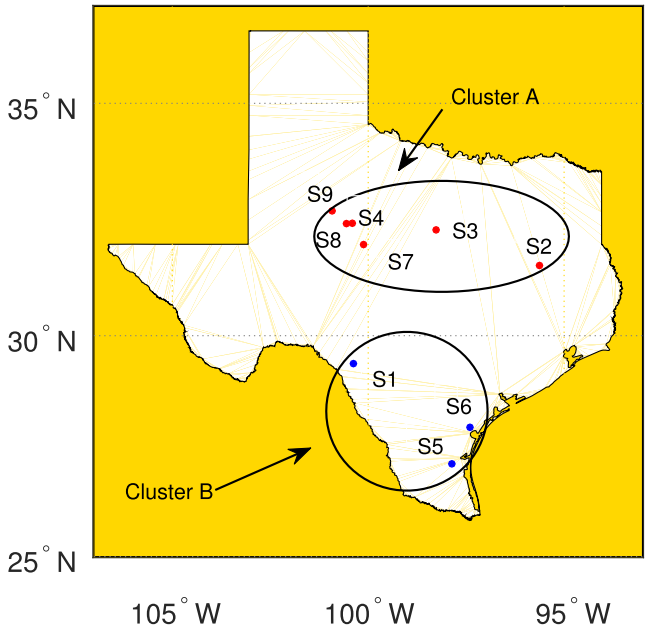


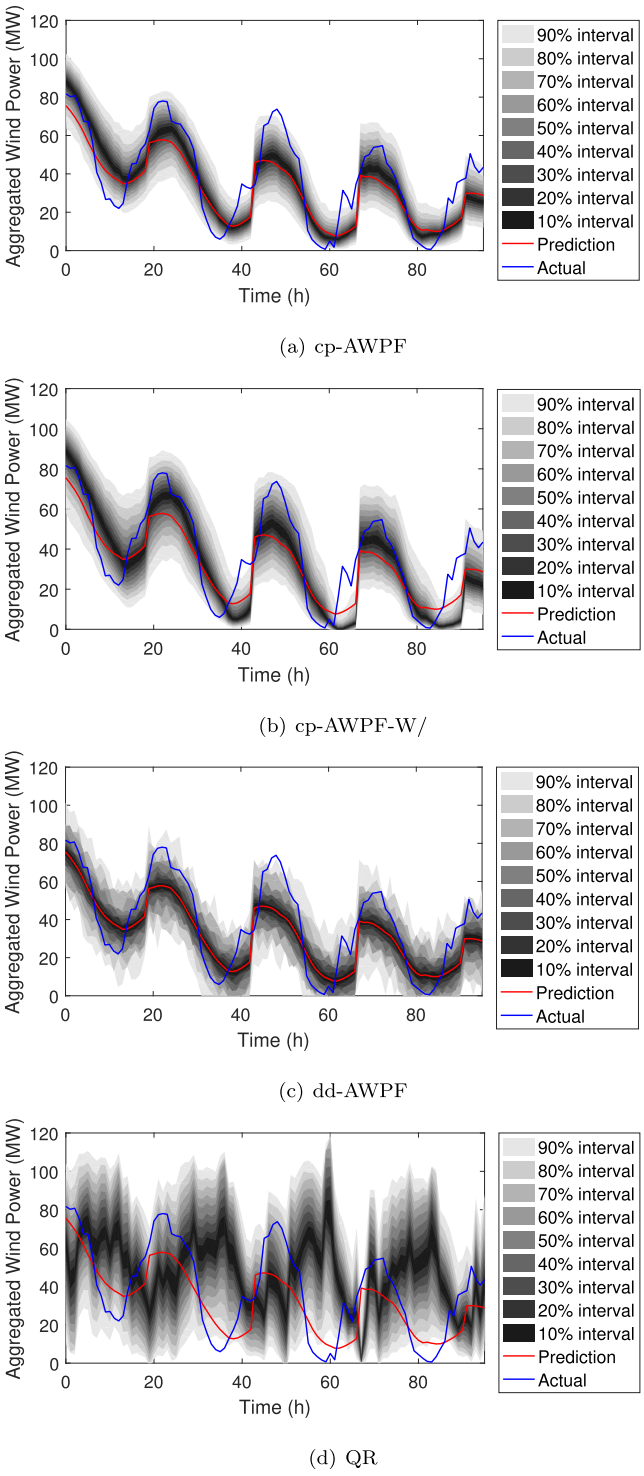
Fig. 4. Locations of clustered wind farms.

clustering is 0.54, which is lower than both of the intra-cluster correlation, further showing the effectiveness of clustering.

3.5. Probabilistic forecasting results

Within each cluster, a joint distribution of the aggregated wind power and the forecasted power of member wind farms is determined. Then the conditional distribution of the aggregated wind power can be calculated, which is used to generate aggregated wind power forecasting scenarios (e.g., 5,000). The quantiles of the aggregated wind power are calculated based on the empirical distribution of the generated scenarios. To evaluate the performance of cp-AWPF, three baseline models are selected for comparison, which are: quantile regression (QR), cp-AWPF without clustering (cp-AWPF-W/), and a data-driven two-step aggregated probabilistic wind power forecasting method with clustering (dd-AWPF). The reasons for choosing these baseline models are: (i) QR is a widely used method in probabilistic forecasting, which allows us to explore the forecasting enhancement by considering spatio-temporal correlation; (ii) since a clustering model is included in cp-AWPF, it is important to compare the accuracy of the proposed cp-AWPF method with the method without clustering; (iii) for the baseline model of dd-AWPF, the forecasting error distribution could be fitted through GMM based on historical deterministic aggregated wind power forecasts. Then, Monte Carlo sampling is used to generate a large number of forecasting error scenarios. These aggregated error scenarios could be used together with deterministic aggregated wind power forecasts to generate probabilistic aggregated wind power forecasts. Details of the GMM fitting and scenario generation for dd-AWPF could be found in [42].

With the estimated empirical predictive PDF of the aggregated wind power, the quantiles q_1, q_2, …, q_{99} can be calculated. To better visualize probabilistic forecasts, the 99 quantiles are converted into nine predictive intervals (PIs) I_β ($\beta = 10$, …, 90) in a 10% increment. Fig. 5(a) shows the day-ahead aggregated probabilistic wind power forecasts of all the 9 wind farms from 2012-07-05 to 2012-07-08, which are generated from the proposed cp-AWPF model. It is observed that in the entire representative period, the aggregated wind power reasonably lies within the PIs. Fig. 5(b), (c), and (d) show the aggregated probabilistic forecasts generated from the baseline cp-AWPF-W/, dd-AWPF, and QR methods, respectively. It is seen that the PIs of the probabilistic forecasting with clustering in Fig. 5(a) are narrower than the PIs without



(a) cp-AWPF

(b) cp-AWPF-W/

(c) dd-AWPF

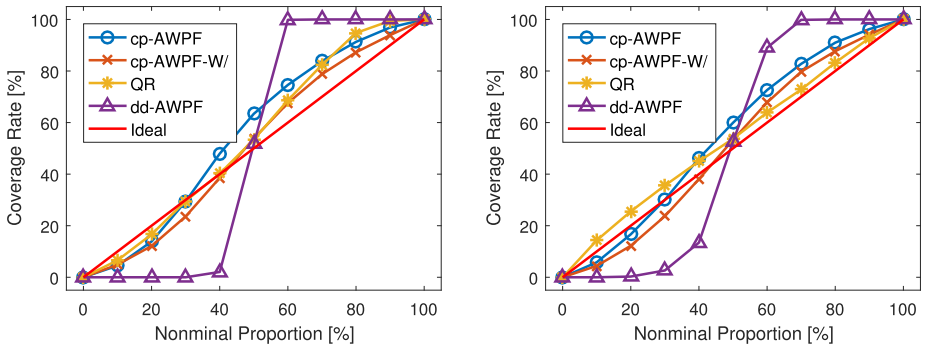
(d) QR

Fig. 5. Day-ahead aggregated probabilistic wind power forecasts.

Table 4
Normalized pinball loss of different models with different look-ahead hours.

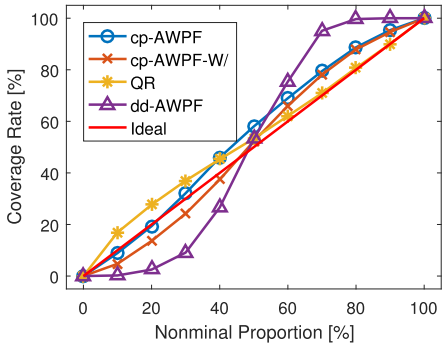
Model	Look-ahead Times						
	1HA	2HA	3HA	4HA	5HA	6HA	DA
cp-AWPF	**1.91**	**2.03**	**2.31**	**2.81**	**3.46**	**4.10**	**3.28**
cp-AWPF-W/	2.92	2.93	3.06	3.40	3.94	4.63	4.26
dd-AWPF	2.20	2.38	2.71	3.19	3.83	4.55	3.47
QR	2.98	3.34	4.76	5.80	6.53	7.02	7.27

Note: The smallest normalized pinball loss value is in boldface.

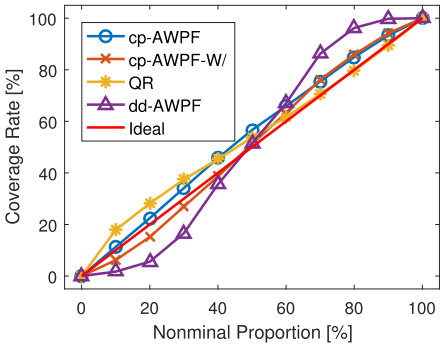


(a) 1HA

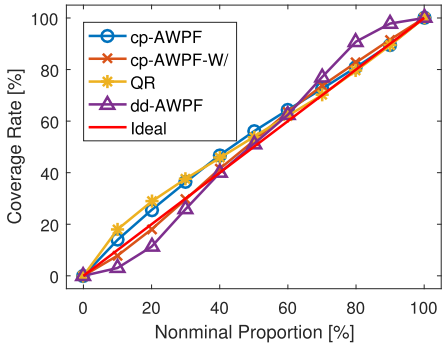
(b) 2HA



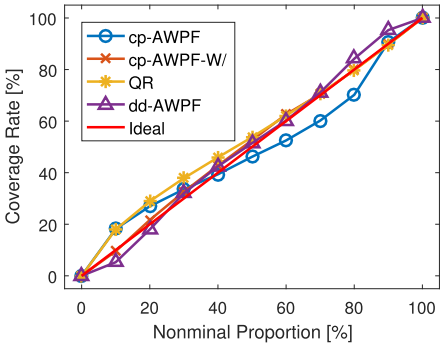
(c) 3HA



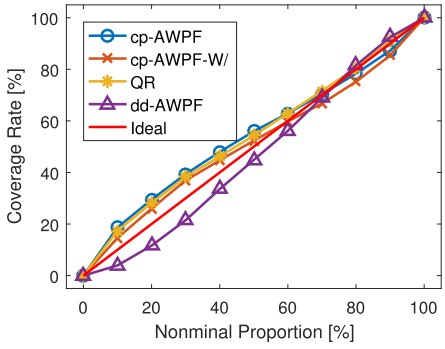
(d) 4HA



(e) 5HA



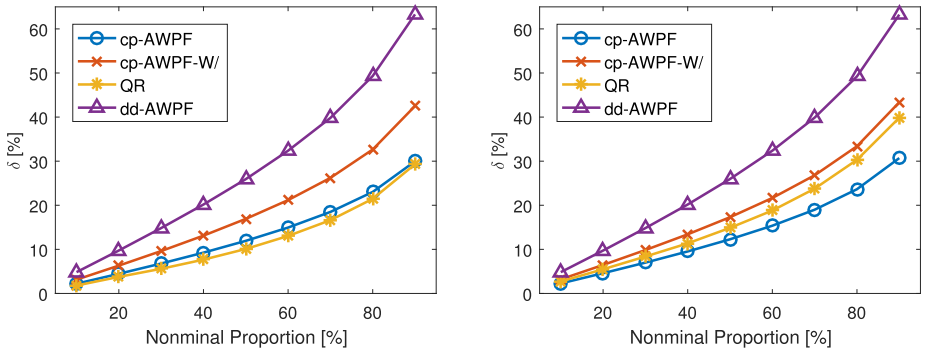
(f) 6HA



(g) DA

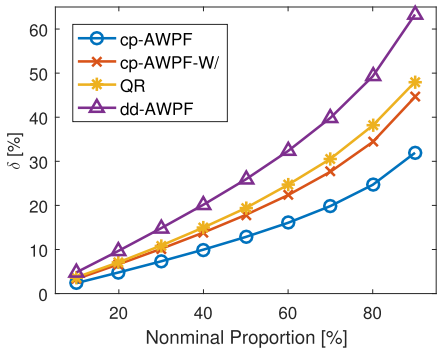
Fig. 6. Reliability comparison of different models with different look-ahead times.

clustering in Fig. 5(b). This is due to that clustering has enhanced the intra-cluster correlation among wind farms. In addition, the PIs of both cp-AWPF and cp-AWPF-W/ are significantly narrower than those of QR in Fig. 5(d) which is due to the consideration of spatio-temporal correlation. The dd-AWPF has a similar width of PIs as cp-AWPF-W/. Nevertheless, the PIs of cp-AWPF-W/ are smoother than those of dd-AWPF, which indicates a stable and reliable probabilistic forecast. It is also observed that the width of the PIs varies with the variability of the

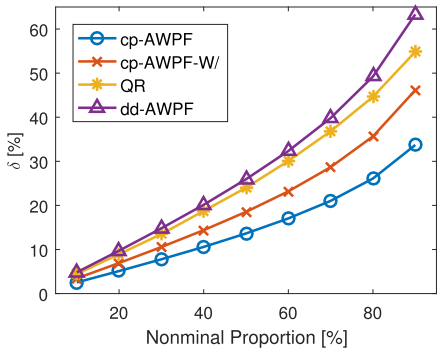


(a) 1HA

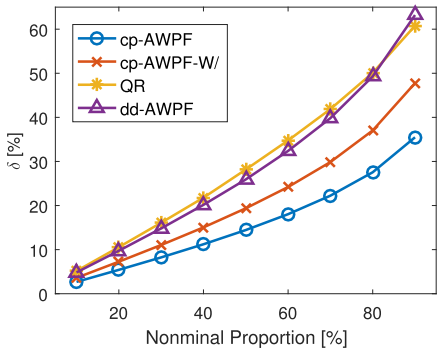
(b) 2HA



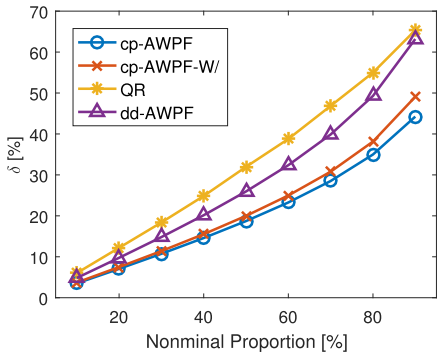
(c) 3HA



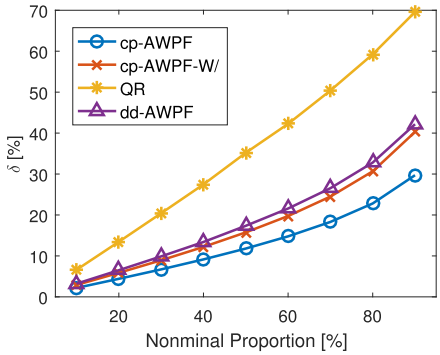
(d) 4HA



(e) 5HA



(f) 6HA



(g) DA

Fig. 7. Sharpness comparison of different models with different look-ahead times.

aggregated wind power. For example, when the wind power fluctuates more frequently, the PI tends to be wider, and thereby the uncertainty in wind power forecasts is relatively higher.

3.5.1. *Pinball loss*

Pinball loss is a widely used metric to evaluate the overall performance of probabilistic forecasts, which is defined by:

$$L_{m,t}\left(q_{m,t}, p_t\right) = \begin{cases} (1 - \frac{m}{100}) \times (q_{m,t} - p_t), & p_t < q_{m,t} \\ \frac{m}{100} \times (p_t - q_{m,t}), & p_t \geqslant q_{m,t} \end{cases} \quad (25)$$

where $q_{m,t}$ represents the mth quantile at time t. To show the effectiveness of the developed spatio-temporal based probabilistic forecasting framework, the normalized pinball loss values of different models with different look-ahead time are compared in Table 4. The sum of pinball loss is averaged over all quantiles from 1% to 99% and normalized by the aggregated wind farm capacity. A lower pinball loss score indicates a better probabilistic forecast. Results show that the proposed cp-AWPF has improved the pinball loss by up to 54% compared to the three benchmark models, which validates the effectiveness of spatio-temporal correlation and clustering. Note that the cp-AWPF method has shown a better accuracy than baseline models for both 1HA-6HA and day-ahead forecasts, which validates the robustness of the methodology. Furthermore, cp-AWPF without clustering also has better pinball loss than QR, which shows the improvement from spatio-temporal correlation modeling. The dd-AWPF model outperforms both QR and cp-AWPF-W/, which indicates the enhancement resulted from clustering. In addition to pinball loss, two more standard metrics, i.e., reliability and sharpness, are also calculated to assess the performance of aggregated probabilistic forecasting.

3.5.2. Reliability

Reliability (RE) stands for the correctness of a probabilistic forecast that matches the observation frequencies [15]:

$$RE = \left[\frac{\xi^{(1-\alpha)}}{N} - \left(1 - \alpha\right)\right] \times 100\% \quad (26)$$

where N is the number of test samples, and $\xi^{(1-\alpha)}$ is the number of times that the actual test samples lie within the αth prediction interval. 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 100% with a 10% increment. Fig. 6 shows the 1HA to 6HA and DA reliability plots of the aggregated probabilistic forecasts with different forecasting models. A forecast presents better reliability when the curve is closer to the diagonal. It is seen from Fig. 6 that overall QR has better reliability performance. It is mainly because the PIs of QR are much wider than those of the proposed cp-AWPF method. A wider PI indicates that the result takes more errors into consideration; however, note that the reliability over the 90th confidence interval is similar between cp-AWPF and QR, which is generally more important in probabilistic forecasting applications. In addition, the proposed cp-AWPF model has shown better reliability than dd-AWPF, indicating the enhancement resulted from spatio-temporal correlation modeling.

3.5.3. Sharpness

Sharpness indicates the capacity of a forecasting system to forecast extreme probabilities [43]. 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 sharpness is measured by the average size of the predictive intervals. The sharpness plots of cp-AWPF and baseline models (i.e., cp-AWPF-W/, dd-AWPF, and QR) with different look-ahead hours are compared in Fig. 7. The expected interval size increases with increasing the nominal coverage rate, and the sharpness of the proposed cp-AWPF model is significantly better than that of the baseline models (i.e., cp-AWPF-W/ and QR) except at the 1HA horizon. It is mainly because the reliability and sharpness are two complementary metrics, and the better reliability of cp-AWPF at 1HA sacrifices the sharpness to some extent. Note that the cp-AWPF model has shown better sharpness than dd-AWPF, indicating the enhancement resulted from spatio-temporal correlation modeling. Overall, the interval size of the proposed cp-AWPF model ranges from 2% to 45%, which indicates low sharpness. In addition, cp-AWPF has significantly better sharpness than cp-AWPF-W/ at all look-ahead times, which validates the effectiveness of clustering.

4. Conclusion

In this paper, an aggregated conditional probabilistic wind power forecasting (cp-AWPF) framework was developed by considering spatio-temporal correlation and wind farm clustering. The K-means clustering approach was applied to cluster 9 wind farms into two clusters. In each cluster, GMM was adopted to accurately model the marginal wind power distribution. Then, the spatio-temporal correlation between the member wind farms and the aggregated wind power was modeled through a high-dimensional joint distribution based on Copula theory. Inverse sampling was applied on the conditional CDF of the joint distribution to generate aggregated probabilistic forecasts. Results at 9 selected wind farms showed that:

1. cp-AWPF could reduce the pinball loss score by up to 54% compared to three benchmark models.
2. The GMM model has shown better goodness-of-fit to wind power distribution than single-distribution models and KDE.
3. Clustering could enhance intra-cluster correlation among member wind farms, thus providing better probabilistic forecasting accuracy.
4. The developed cp-AWPF framework is robust at different forecasting time horizons and locations.
5. cp-AWPF has shown better sharpness than models without considering spatio-temporal correlation and clustering. The reliability of cp-AWPF is close to the ideal diagonal, which indicates reasonable reliability.

Potential future work will (i) utilize the aggregated probabilistic wind power forecasting in stochastic power system operations, and (ii) explore the influence of meteorological and geographical conditions on the aggregated probabilistic forecasts.

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).

References

[1] Tastu J, Pinson P, Kotwa E, Madsen H, Nielsen HA. Spatio-temporal analysis and modeling of short-term wind power forecast errors. Wind Energy 2011;14(1):43–60.
[2] Malvaldi A, Weiss S, Infield D, Browell J, Leahy P, Foley AM. A spatial and temporal correlation analysis of aggregate wind power in an ideally interconnected Europe. Wind Energy 2017;20(8):1315–29.
[3] Lenzi A, Steinsland I, Pinson P. Benefits of spatiotemporal modeling for short-term wind power forecasting at both individual and aggregated levels. Environmetrics 2018;29(3). e2493.
[4] Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 2018:1. https://doi.org/10.1109/TSTE.2018.2844102.
[5] Zhao T, Zhang Y, Chen H. Spatio-temporal load forecasting considering aggregation features of electricity cells and uncertainties in input variables. J Electr Eng Technol 2018;13(1):38–50.
[6] Mohan N, Soman K, Kumar SS. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Appl Energy 2018;232:229–44.
[7] Agoua XG, Girard R, Kariniotakis G. Probabilistic model for spatio-temporal photovoltaic power forecasting. IEEE Trans Sustain Energy 2018:1. https://doi.org/10.1109/TSTE.2018.2847558.
[8] Pelikan E, Eben K, Resler J, Juruš P, Krč P, Brabec M, et al. Wind power forecasting by an empirical model using nwp outputs. Environment and electrical engineering (EEEIC), 2010 9th international conference on, IEEE. 2010. p. 45–8.
[9] Feng C, Cui M, Hodge B-M, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy

2017;190:1245–57.
[10] Xie L, Gu Y, Zhu X, Genton MG. Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. IEEE Trans Smart Grid 2014;5(1):511–20.
[11] Qin Y, Li K, Liang Z, Lee B, Zhang F, Gu Y, et al. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. Appl Energy 2019;236:262–72.
[12] Zhao Y, Ye L, Pinson P, Tang Y, Lu P. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. IEEE Trans Power Syst 2018;33(5):5029–40.
[13] Baxevani A, Lenzi A. Very short-term spatio-temporal wind power prediction using a censored gaussian field. Stoch Environ Res Risk Assess 2018;32(4):931–48.
[14] Fang X, Hodge B-M, Du E, Zhang N, Li F. Modelling wind power spatial-temporal correlation in multi-interval optimal power flow: a sparse correlation matrix approach. Appl Energy 2018;230:531–9.
[15] Sun M, Feng C, Chartan EK, Hodge B-M, Zhang J. A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization. Appl Energy 2019;238:1497–505.
[16] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy: An Int J Progr Appl Wind Power Convers Technol 2006;9(1–2):95–108.
[17] Wang Y, Zhang N, Tan Y, Hong T, Kirschen DS, Kang C. Combining probabilistic load forecasts. IEEE Trans Smart Grid 2018:1. https://doi.org/10.1109/TSG.2018.2833869.
[18] Bessa RJ, Miranda V, Botterud A, Zhou Z, Wang J. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renew Energy 2012;40(1):29–39.
[19] Juban J, Siebert N, Kariniotakis GN. Probabilistic short-term wind power forecasting for the optimal management of wind generation. Power Tech, 2007 IEEE Lausanne. IEEE; 2007. p. 683–8.
[20] Lange M. On the uncertainty of wind power predictions analysis of the forecast accuracy and statistical distribution of errors. J Sol Energy Eng 2005;127(2):177–84.
[21] Pinson P. Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. J Roy Stat Soc: Ser C (Appl Stat) 2012;61(4):555–76.
[22] Gómez-Lázaro E, Bueso M, Kessler M, Martín-Martínez S, Zhang J, Hodge B-M, et al. Probability density function characterization for aggregated large-scale wind power based on weibull mixtures. Energies 2016;9(2):91.
[23] Zhang N, Kang C, Xu Q, Jiang C, Chen Z, Liu J. Modelling and simulating the spatio-temporal correlations of clustered wind power using copula. J Electr Eng Technol 2013;8(6):1615–25.
[24] Zhang Y, Wang J. A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information. IEEE Trans Power Syst 2018;33(5):5714–26.
[25] Dowell J, Pinson P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. IEEE Trans Smart Grid 2016;7(2):763–70.
[26] Tang C, Wang Y, Xu J, Sun Y, Zhang B. Efficient scenario generation of multiple renewable power plants considering spatial and temporal correlations. Appl Energy 2018;221:348–57.
[27] Li P, Guan X, Wu J. Aggregated wind power generation probabilistic forecasting based on particle filter. Energy Convers Manage 2015;96:579–87.
[28] Pinson P, Madsen H, Nielsen HA, Papaefthymiou G, Klöckl B. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy: An Int J Progr Appl Wind Power Convers Technol 2009;12(1):51–62.
[29] Rüschendorf L. On the distributional transform, sklar's theorem, and the empirical copula process. J Stat Plann Inference 2009;139(11):3921–7.
[30] Cui M, Krishnan V, Hodge B, Zhang J. 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.
[31] Louie H. Characterizing and modeling aggregate wind plant power output in large systems. Power and energy society general meeting, 2010 IEEE. IEEE; 2010. p. 1–8.
[32] Wang Z, Wang W, Liu C, Wang Z, Hou Y. Probabilistic forecast for multiple wind farms based on regular vine copulas. IEEE Trans Power Syst 2018;33(1):578–89.
[33] Feng C, Sun M, Zhang J. Reinforced Deterministic and Probabilistic Load Forecasting via Q-Learning Dynamic Model Selection. IEEE Trans Smart Grid 2019.
[34] Feng C, Zhang J. Reinforcement learning based dynamic model selection for short-term load forecasting. 2019 IEEE international conference on innovative smart grid technologies (ISGT). IEEE; 2019. p. 1–6.
[35] Hartley HO. Maximum likelihood estimation from incomplete data. Biometrics 1958;14(2):174–94.
[36] Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I. An extensive comparative study of cluster validity indices. Pattern Recogn 2013;46(1):243–56.
[37] Wang C, Liang Z, Liang J, Teng Q, Dong X, Wang Z. Modeling the temporal correlation of hourly day-ahead short-term wind power forecast error for optimal sizing energy storage system. Int J Electr Power Energy Syst 2018;98:373–81.
[38] Zhang N, Kang C, Singh C, Xia Q. Copula based dependent discrete convolution for power system uncertainty analysis. IEEE Trans Power Syst 2016;31(6):5204–5.
[39] Draxl C, Clifton A, Hodge B-M, McCaa J. The wind integration national dataset (wind) toolkit. Appl Energy 2015;151:355–66.
[40] Zhang J, Draxl C, Hopson T, Delle Monache L, Vanvyve E, Hodge B-M. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. Appl Energy 2015;156:528–41.
[41] Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media; 2003.
[42] Cui M, Feng C, Wang Z, Zhang J, Wang Q, Florita A, et al. Probabilistic wind power ramp forecasting based on a scenario generation method. IEEE Power energy society general meeting, Chicago, IL. 2017.
[43] Sun M, Feng C, Zhang J, Chartan EK, Hodge B-M. Probabilistic short-term wind forecasting based on pinball loss optimization. 2018 IEEE international conference on probabilistic methods applied to power systems (PMAPS). IEEE; 2018. p. 1–6.