

Very Short-term Probabilistic Prediction for Regional Wind Power Generation Based on OPNPIs

Yan Zhou, Yonghui Sun, Sen Wang, Rabea Jamil Mahfoud, Dongchen Hou, and Jianxi Wang

Abstract—Due to the uncertainty and fluctuation of wind power generation, probabilistic prediction for regional wind power generation is critical to accurately quantify the uncertainty of meaningful information to the dispatching departments of power grid. This paper proposes an approach of very short-term probabilistic prediction for regional wind power generation based on optimal performance-based nonparametric prediction intervals (OPNPIs). First, the deterministic prediction for regional wind power generation considering the division of wind farms based on the detrending-based partial cross-correlation analysis (DPCCA) is studied. Based on the deterministic prediction and its prediction errors, the OPNPIs are proposed considering the reliability and overall performance for the uncertainty analysis. Furthermore, a regulating coefficient is studied to further enhance the performance of PIs. Effectiveness of the proposed method is verified through multistep PIs of 15-minute based on the real wind power generation data.

Index Terms—Detrending-based partial cross-correlation analysis, Huber-based approach, nonparametric prediction intervals, overall performance, regional wind power generation.

I. INTRODUCTION

NOWADAYS, due to the high penetration levels and the inherent variability of wind energy [1]–[3], wind power prediction is essential for the dispatching department to adjust schedules, so that the system reserves and operational cost are effectively reduced [4]–[6], and also applied for electricity market trading [7], [8].

In recent years, lots of studies on parametric or nonparametric prediction intervals (PIs) for wind power generation have been developed [9]–[17]. The parametric PIs were quantified based on the deterministic prediction and the error estimation based on parametric error distribution. The deterministic predictions based on extreme learning machine (ELM) were

efficient and effective, without the need for parameters' optimization [18], [19]. To further improve the forecasting performance, the approach of multiple-ELM was proposed in [20]. In [18] and [21], based on the deterministic predictions, considering normal distribution and accuracy of PIs coverage, the advanced parametric PIs for prediction errors were performed. Although the performance of parametric PIs can be enhanced by advanced deterministic predictions, it is still affected by the assumption accuracy of error distribution. Hence, to further improve PIs performance, an effective solution is to study nonparametric uncertainty analysis. Efficient nonparametric PIs including nonlinear quantile regression (NQR) could be quantified, without prior assumption of error distribution, used for probabilistic predictions of wind power and photovoltaic (PV) power generation [22]–[24], as well as electricity price and load [25], [26]. They were mainly quantified based on the reliability evaluated by the deviation between prediction interval coverage probability (PICP) and prediction interval nominal confidence (PINC). Considering both reliability and sharpness, conventional nonparametric PIs could be further enhanced [27], [28]. For improved forecasting performance, considering both overall performance and reliability [18], not only the accuracy of PIs coverage and sharpness, but also the offsets of points outside PIs should be considered. Inspired by this, it is necessary to perform an uncertainty analysis of prediction errors for optimal forecasting performance considering both overall performance and reliability. Also, compared with the conventional quantification of PIs that only considered the prediction errors, analyzing the relationship between wind power time series as input and prediction errors can further improve the uncertainty analysis [18].

Compared with the prediction for outputs of an individual wind farm, the prediction for regional outputs makes more sense to the dispatching department. The upscaling-based approach studies the correlations between outputs and numerical weather prediction (NWP) data. Then the regional outputs can be quantified by upscaling the representative wind farms [29], [30]. However, it is difficult to obtain representative wind farms and quantification of regional outputs accurately. Meanwhile, the sum of wind power generations with similar characteristics may cause a sudden change of the cluster outputs which results in low accuracy of prediction. Considering the smoothing effect of regional outputs [31], [32], it is effective to directly study the regional outputs, rather than cluster outputs of the wind farms with similar characteristics. In [33], based on the spatio-temporal correlation between historical wind speeds and wind speeds from NWP data as the weights, regional

Manuscript received June 15, 2022; revised November 3, 2022; accepted January 31, 2023. Date of online publication February 14, 2024; date of current version October 26, 2024. This work was supported by the National Natural Science Foundation of China (62073121); National Key R&D Program of China Technology and application of wind power/photovoltaic power prediction for promoting renewable energy consumption (2018YFB0904200) and eponymous Complement S&T Program of State Grid Corporation of China (SGLNDK00KJJS1800266).

Y. Zhou, Y. H. Sun (corresponding author, email: sunyonghui168@gmail.com), S. Wang, D. C. Hou, and J. X. Wang are with the College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China.

R. J. Mahfoud is with the College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China.

DOI: 10.17775/CSEEJPES.2022.02790

outputs were quantified. The approach which directly studies regional wind power outputs can reduce the influence of the fluctuation and sudden change of individual wind farm outputs. For the very short-term probabilistic prediction, due to the insufficient accuracy of NWP, it is reasonable to only utilize the historical wind power observations as the input of the prediction model [18], [21]. However, in fact, due to the high synchronicity between wind speed and wind generation, NWP is also helpful to obtain the trend of wind power outputs within a few hours [34], [35]. Thus, in addition to considering power observations as input variables, whether to consider wind speed data as input variables or not should be analyzed for each wind farm. Considering the smoothing effect, it is necessary to divide regional wind farms into separated wind farm clusters considering NWP data as input variables or not. However, the conventional Pearson correlation [36] cannot effectively quantify the common trend caused by external factors which can also affect accuracy of the correlations [37]. Due to the slow periodical fluctuations caused by the seasonal or other factors [38], which affect the correlation analysis of wind power and wind speed, in this paper, a detrending-based partial cross-correlation analysis (DPCCA) is studied for the accurate utilization of input variables, considering a reasonable division of wind farms with high correlation or low correlation.

In this paper, a novel approach of nonparametric PIs for regional wind power outputs is proposed. The main contributions of this paper are as follows.

1) In conventional very short-term prediction methods for wind power, considering only the historical power observations as model input makes it impossible to accurately predict the future output trend. To optimize the utilization of input data for enhanced deterministic prediction, the proposed DPCCA is employed to divide the regional wind farms into wind farm clusters with high correlation or low correlation between the wind power observations and wind speed data.

2) Optimal performance-based nonparametric prediction intervals (OPNPIs) for uncertainty analysis of prediction errors considering reliability and overall performance are proposed to enhance the PIs performance. Different from conventional uncertainty quantification that was only based on reliability which results in insufficient overall performance. Furthermore, the regulating coefficient is supplemented in the cost function of OPNPIs, which makes the proposed PIs more flexible to regulate the nominal proportion of boundary quantiles for the optimized forecasting performance, compared with conventional nonparametric PIs.

The rest of the paper is organized as follows: In Section II, methodologies of the proposed method are described. The quantification process of the proposed PIs is given in Section III. Case studies are carried out to verify the performance of the proposed approach with numerical comparisons in Section IV. Finally, conclusions are given in Section V.

II. METHODOLOGIES

In this section, first, the division of regional wind farms based on DPCCA and Huber-based multiple-pair extreme learning machines (HMELM) for deterministic prediction are,

respectively, demonstrated. Second, for the uncertainty analysis, an OPNPIs-based approach is proposed. Moreover, it is further enhanced by the optimal regulating coefficient.

A. Division of Regional Wind Farms

In this paper, the high or low correlation based on DPCCA is utilized to portray the correlation between wind speed data and wind power observations for each wind farm. Then, the wind farms are clustered with high correlation or low correlation for accurate utilization of NWP data. The proposed DPCCA is given as follows:

1) Wind power observations or wind speed data for each wind farm are reconstructed as a new vector, defined by:

$$X_j(i) = \sum_{s=1}^i x_j(s), \quad i = 1, 2, 3, \dots, S, \quad j = 1, 2 \quad (1)$$

where $x_j(s)$ denotes the s th prediction value of wind speed or observation of wind power, S represents size of the training data set, and j denotes the selection of vector for wind power or wind speed.

2) The quantification of the moving average is used to reduce impact of the overall trend, given by:

$$\bar{X}_j(i) = \begin{cases} \frac{1}{ww} \sum_{k=0}^{ww-1} X_j(i-k), & \text{if } i > 1 \\ X_j(i), & \text{if } i = 1 \end{cases} \quad (2)$$

where $\bar{X}_j(i)$ is i th value of moving average, ww is the width of the computational window, as detrending scale.

3) The covariance of detrending between j_1 th and j_2 th vectors is given by:

$$V^2(j_1, j_2) = \frac{\sum_{i=ww}^S (X_{j_1}(i) - \bar{X}_{j_1}(i))(X_{j_2}(i) - \bar{X}_{j_2}(i))}{S - ww + 1} \quad (3)$$

Therefore, a matrix of covariance between the j_1 th and j_2 th vectors is formulated as:

$$\mathbf{V}^2 = \begin{bmatrix} V^2(j_1, j_1) & V^2(j_1, j_2) \\ V^2(j_1, j_2) & V^2(j_2, j_2) \end{bmatrix} \quad (4)$$

4) Cross-correlation can be quantified by:

$$\rho(j_1, j_2) = \frac{V^2(j_1, j_2)}{\sqrt{V^2(j_1, j_1) \cdot V^2(j_2, j_2)}} \quad (5)$$

Then, the matrix of cross-correlation can be presented by:

$$\boldsymbol{\rho} = \begin{bmatrix} \rho(j_1, j_1) & \rho(j_1, j_2) \\ \rho(j_1, j_2) & \rho(j_2, j_2) \end{bmatrix} \quad (6)$$

5) After the detrending-based cross-correlation analyzed above, the quantification of partial cross-correlation analysis is utilized to improve accuracy of correlation quantification, which is defined by:

$$\mathbf{C} = \boldsymbol{\rho}^{-1} = \begin{bmatrix} C(1, 1) & C(1, 2) \\ C(2, 1) & C(2, 2) \end{bmatrix} \quad (7)$$

Thus, the correlation of DPCCA is given by:

$$\rho_p = \frac{-C(1, 2)}{\sqrt{C(1, 1) \cdot C(2, 2)}} \quad (8)$$

Based on the sort of DPCCA correlation values for wind farm output and wind speed from lowest to highest, and the corresponding deterministic prediction accuracies in training, the optimal division of regional wind farms with high correlation or low correlation can be obtained. After that, to effectively study wind speed data for the deterministic prediction of cluster outputs with high correlation, composite wind speed of wind farms with high correlations in the cluster is proposed, quantified by:

$$S_c = \frac{\sum_{i=1}^{N_w} C_i S_i}{\sum_{i=1}^{N_w} C_i} \quad (9)$$

where N_w is the number of wind farms in the cluster, C_i and S_i represent the capacity and wind speed of i th wind farm, respectively. Capacities of wind farms are utilized as weights to consider the impact of weather conditions on wind farms [33].

B. Huber-based Multiple-pair Extreme Learning Machines

In this paper, the deterministic prediction based on HMELM for the wind farm cluster outputs is utilized. Its prediction output is quantified by:

$$f(\mathbf{x}_i) = \frac{\sum_{j=1}^{N_c} h_j f^{(j)}(\mathbf{x}_i)}{\sum_{j=1}^{N_c} h_j} \quad (10)$$

where $\mathbf{x}_i$ denotes the i th input, $f^{(j)}(\mathbf{x}_i)$ and h_j are the output and its weight of j th ELM, respectively, and N_c denotes the total number of ELMs. The Huber-based cost function is formulated to combine multiple ELMs, one has:

$$\text{Huber}(t_i, f(\mathbf{x}_i)) = \begin{cases} \frac{1}{2}(t_i - f(\mathbf{x}_i)), & |t_i - f(\mathbf{x}_i)| \leq \delta \\ \delta|t_i - f(\mathbf{x}_i)| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases} \quad (11)$$

where t_i and $f(\mathbf{x}_i)$ denote the i th actual observation and output of HMELM, respectively, and δ is the threshold to select L1-norm or L2-norm. Here, h_j and δ are optimized based on prediction accuracy. After the quantification of wind farm cluster outputs, the deterministic prediction of regional outputs can be easily obtained [39].

C. Optimal Performance-based Nonparametric PIs

The reliability and overall performance considering sharpness are set as the evaluation criteria of wind power generation PIs [18], [21], [27], [40], which are given by

1) Reliability: The absolute value of average coverage error (ACE) between PICP and PINC is utilized to assess the reliability of PIs, quantified by:

$$|\text{ACE}| = |\text{PICP} - \text{PINC}| \quad (12)$$

where $|\text{PICP} - \text{PINC}|$ denotes the coverage deviation of PIs with a certain confidence. The smaller the value of $|\text{ACE}|$, the better the reliability of PIs.

2) Sharpness: The average width of PIs is used to measure the sharpness of PIs, given by:

$$\delta_t^\alpha(\mathbf{x}_i) = U_t^\alpha(\mathbf{x}_i) - L_t^\alpha(\mathbf{x}_i) \quad (13)$$

3) Overall performance: The interval score is used to measure overall performance of PIs, which is quantified by:

$$Sc_t^\alpha(\mathbf{x}_i) = \begin{cases} -2(1-\alpha)\delta_t^\alpha(\mathbf{x}_i) - 4[L_t^\alpha(\mathbf{x}_i) - t_i], & \text{if } t_i < L_t^\alpha(\mathbf{x}_i) \\ -2(1-\alpha)\delta_t^\alpha(\mathbf{x}_i), & \text{if } t_i \in I_t^\alpha(\mathbf{x}_i) \\ -2(1-\alpha)\delta_t^\alpha(\mathbf{x}_i) - 4[t_i - U_t^\alpha(\mathbf{x}_i)], & \text{if } t_i > U_t^\alpha(\mathbf{x}_i) \end{cases} \quad (14)$$

$$Sc_t^\alpha = \frac{1}{T_p} \sum_{i=1}^{T_p} Sc_t^\alpha(\mathbf{x}_i) \quad (15)$$

where α is PINC, T_p is the size of testing samples, $\delta_t^\alpha(\mathbf{x}_i)$ represents width of the i th interval, $U_t^\alpha(\mathbf{x}_i)$ and $L_t^\alpha(\mathbf{x}_i)$ are the upper bound and lower bound of the PI as the prediction results, respectively. $I_t^\alpha(\mathbf{x}_i)$ denotes PI, $Sc_t^\alpha(\mathbf{x}_i)$ is the score of PI, and Sc_t^α is the interval score.

After deterministic prediction of regional outputs, for the nonparametric uncertainty quantification of the prediction errors, training data consist of the regional wind power generation observations as inputs and prediction errors as outputs, given by:

$$\text{Sample}_i = \{(\mathbf{x}_i', E_i)\}_{i=1}^T \quad (16)$$

where T denotes the size of training samples, $\mathbf{x}_i'$ denotes the input of i th training sample, E_i denotes the prediction error of i th training sample, which is defined by:

$$E_i = t_i - y_i \quad (17)$$

where y_i denotes the deterministic prediction for regional wind power generation.

The proposed nonparametric PIs for prediction errors are mainly based on the criterion of interval score based on (14) and (15), which is quantified by:

$$|\text{Score}| = \frac{1}{T} \sum_{i=1}^T [-2\alpha \text{Width}(\mathbf{x}_i') + D_i] \quad (18)$$

$$D_i = 2(|U^\alpha(\mathbf{x}_i') - E_i| + |L^\alpha(\mathbf{x}_i') - E_i|) \quad (19)$$

where $|\text{Score}|$ denotes the absolute value of interval score, $\text{Width}(\mathbf{x}_i')$ denotes width of i th PI, $U^\alpha(\mathbf{x}_i')$ and $L^\alpha(\mathbf{x}_i')$ denote the upper and lower bounds, respectively. D_i denotes twice the sum of distances between the bounds and actual observations. There are two cases for distance, that is, observations within or outside the intervals. Thus, D_i can be further expressed by:

$$D_i = \begin{cases} 4d_i + 2\text{Width}(\mathbf{x}_i'), & t_i \notin I_t^\alpha(\mathbf{x}_i') \\ 2\text{Width}(\mathbf{x}_i'), & t_i \in I_t^\alpha(\mathbf{x}_i') \end{cases} \quad (20)$$

where d_i denotes the distance between the i th actual observation outside PI and the nearer bound. According to (18)–(20), it is obvious interval score considers not only the reliability and sharpness but also the offsets of points outside PIs. To observe performance of PIs, the average offset (AO) is used as an auxiliary criterion, which is expressed by:

$$\text{offset}_i = \begin{cases} t_i - U_t^\alpha(\mathbf{x}_i'), & t_i > U_t^\alpha(\mathbf{x}_i') \\ L_t^\alpha(\mathbf{x}_i') - t_i, & t_i < L_t^\alpha(\mathbf{x}_i') \end{cases},$$

$$i = 1, 2, 3, \dots, N_o \quad (21)$$

$$AO = \frac{1}{N_o} \sum_{i=1}^{N_o} \text{offset}_i \quad (22)$$

where N_o denotes number of observations outside PIs.

According to (18), the cost function of the proposed OPNPIs is formulated as:

$$\min_{\substack{\xi_i, \bar{\xi}_i, \\ \beta_U, \beta_L}} \sum_{i=1}^T [-2\alpha(f(\mathbf{x}'_i, \beta_U) - f(\mathbf{x}'_i, \beta_L)) + 2\bar{\xi}_i + 2\underline{\xi}_i] \quad (23)$$

subject to

$$f(\mathbf{x}'_i, \beta_U) - f(\mathbf{x}'_i, \beta_L) \geq 0 \quad (24)$$

$$0 \leq f(\mathbf{x}'_i, \beta_U), f(\mathbf{x}'_i, \beta_L) \leq 1 \quad (25)$$

$$-\bar{\xi}_i \leq E_i - f(\mathbf{x}'_i, \beta_U) \leq \bar{\xi}_i \quad (26)$$

$$-\underline{\xi}_i \leq E_i - f(\mathbf{x}'_i, \beta_L) \leq \underline{\xi}_i \quad (27)$$

$$\bar{\xi}_i, \underline{\xi}_i \geq 0 \quad (28)$$

$$\bar{\xi}_i = |f(\mathbf{x}'_i, \beta_U) - E_i| \quad (29)$$

$$\underline{\xi}_i = |f(\mathbf{x}'_i, \beta_L) - E_i| \quad (30)$$

where $f(\mathbf{x}'_i, \beta_U)$ and $f(\mathbf{x}'_i, \beta_L)$ denote linear function outputs of ELMs for upper and lower boundary quantiles as the bounds of PIs, respectively, while β_U and β_L denote the output weights of ELMs for the upper and lower bounds of PIs, respectively. $\bar{\xi}_i$ and $\underline{\xi}_i$ are the auxiliary variables. As analyzed in [22], based on the output weights of ELMs, the upper and lower quantiles as the bounds of PIs are obtained effectively. Construction of PIs is mainly formulated as a linear optimization problem and the optimal output weights of ELMs for the upper and lower boundary quantiles can be efficiently quantified by the linear programming (LP) algorithm. Different from QR approach based on the reliability of the cost function given in [22], the cost function of the proposed OPNPIs for PIs quantification is mainly based on the interval score for optimized overall performance of probabilistic prediction, to implement the uncertainty analysis of wind power output based on LP.

As analyzed above, the cost function given in (23) mainly focuses on the overall performance. However, the decisive criterion of PIs performance is interval score, meanwhile considering reasonable reliability [22], [27]. Utilizing (23) as the cost function, PIs will have an excellent interval score, but probably low reliability. Hence, the regulation for the improvement of reliability should be supplemented in the cost function. Thus, a regulating coefficient is needed to obtain better reliability, by optimizing the boundary quantiles. To obtain optimal performance considering both reliability and overall performance, the cost function (23) is modified as:

$$\min_{\substack{\xi_i, \bar{\xi}_i, \\ \beta_U, \beta_L}} \sum_{i=1}^T [-2(\alpha + K)(f(\mathbf{x}'_i, \beta_U) - f(\mathbf{x}'_i, \beta_L)) + 2\bar{\xi}_i + 2\underline{\xi}_i] \quad (31)$$

subject to

$$(24)-(30) \quad (32)$$

where K denotes the coefficient to regulate the nominal proportion of boundary quantiles for optimal PIs performance, and is optimized according to PIs performance. When the value of PICP is too small, the value of α which represents PINC in (18) should be increased. On the contrary, when the value of PICP is too large, the value of α should be decreased. That is, $\alpha + K$ in (31) can be used as a new value of PINC, which helps to quantify the optimal PIs by regulating the bounds of PIs.

III. PROPOSED MODEL CONSTRUCTION

In this section, steps and related demonstration of the proposed method are given. First, regional wind farms are divided into two clusters based on DPCCA. According to the prediction accuracy, the optimal division of wind farms is obtained. Then, the deterministic prediction and its prediction errors are quantified. Uncertainty analysis for the prediction errors of regional outputs can be performed, based on the proposed OPNPIs with the optimal regulating coefficient. The diagram of the proposed PIs is shown in Fig. 1, and detailed steps of the proposed method are described as follows:

Step 1: Set PIs confidence and normalize the dataset.

Step 2: Quantify DPCCA correlations between wind power observations and wind speed data, and sort the wind farms according to the correlations from lowest to highest.

Step 3: FOR each number of wind farms with low correlation from 1 to $N_w - 1$:

- a) Divide regional wind farms into clusters with high correlation or low correlation according to the sort of DPCCA correlations and the number of wind farms with low correlation. Then, the regional outputs can be quantified based on deterministic prediction models for the cluster outputs.
- b) Record all the prediction accuracies of regional outputs with different numbers of wind farms with low correlations. Then the optimal division of wind farms can be obtained, according to prediction accuracies.

Step 4: For each resolution in the search space of particle swarm optimization (PSO) [41] for the optimal Huber thresholds and the weights of multiple prediction outputs:

- a) Based on the proposed deterministic prediction for the cluster outputs with DPCCA and HMELM, the root mean square errors (RMSEs) and mean absolute errors (MAEs) of prediction results are obtained.
- b) Record the optimal thresholds and the related weights.

Step 5: Collect the samples according to (16) for quantification of uncertainty, based on regional outputs and prediction errors.

Step 6: FOR each resolution in the search space of PSO for the optimal regulating coefficient K in (31):

- a) According to (31) and (32), quantify output weights β_U and β_L .
- b) Based on the prediction results and the proposed OPNPIs, quantify the interval score and reliability of PIs in training.

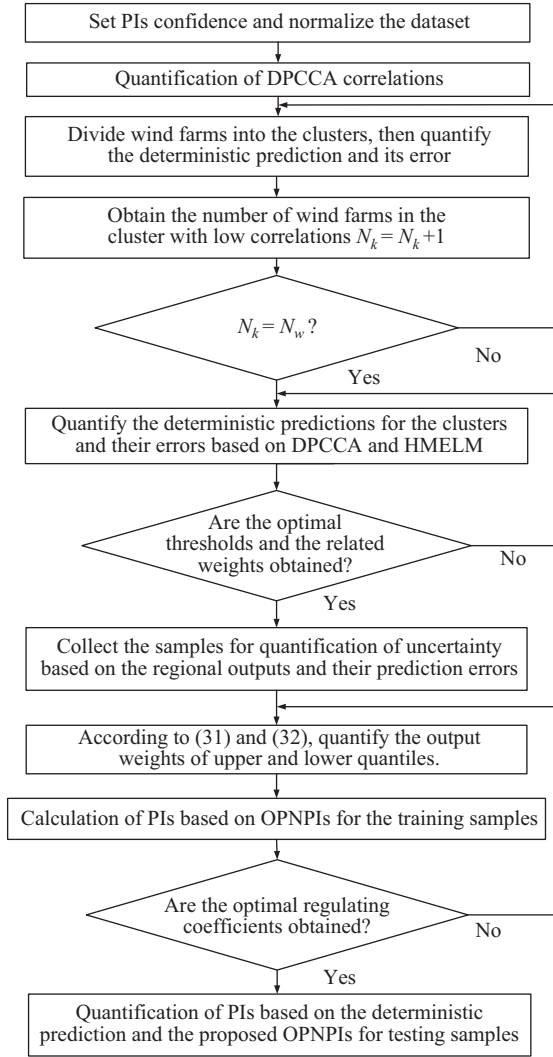


Fig. 1. Diagram of the proposed PIs.

Step 7: For testing samples, based on the optimal thresholds, regulating coefficient, and the related output weights, the PIs can be quantified.

Remark 1: With a look-ahead time within a few hours, the very short-term forecasts of wind generation which only use wind power observations as input perform well [18], [22], [27]. However, NWP is the only basis for determining the future trend of wind generation. The utilization of NWP should be reasonably considered, based on the comparison of prediction accuracies and the DPCCA between outputs and wind speeds, which helps to reduce the effect of seasonal trending. For the wind farms with high correlation, the generation of each wind farm has high synchronicity with wind speed. Thus, it is significant to consider the wind speed data while training. Besides, considering the smoothing effect [31], the observations of wind farm clusters are utilized for deterministic prediction to reduce the mutability of wind power. Meanwhile, the Huber-based cost function helps to obtain robust quantification [20], [39].

Remark 2: Based on the deterministic prediction and assumption of error distribution, the conventional parametric

PIs are calculated by considering the reliability [18], [21]. Approaches can achieve reliable PIs. However, the decisive criteria of PIs include not only reliability but also overall performance. Thus, the criterion of overall performance should be considered while formulating PIs. Meanwhile, the error assumption of PIs influences the forecasting performance [22]. The proposed OPNPIs consider both reliability and overall performance, and are nonparametric PIs without dependence on the parametric assumption of prediction errors. The uncertainty quantification is mainly based on (29), which is an enhanced cost function of interval score with the regulating coefficient by optimizing the nominal proportion of boundary quantiles.

IV. CASE STUDIES

A. Introduction of Dataset

In this part, the wind power generations and NWP of 167 wind farms with 15-minute resolution covering June to November in 2020 are utilized. The wind farms are located in the north of China. Wind power observations and wind speeds are utilized after normalization based on the capacity and maximum of wind speed, respectively. Each sample group consists of a 24-day dataset, a 20-day dataset of which is utilized for training, while the rest are for testing.

B. Numerical Analysis of Prediction Accuracy

In the proposed deterministic prediction, considering the correlations of DPCCA and the prediction accuracies of regional outputs, the wind farms are divided into two clusters. The width of computational window for DPCCA in (2) is set to 2. The datasets in June and November are utilized for numerical analysis. With 80% of the training samples used for training and the rest used for testing in the cross-validation, the prediction errors of regional outputs according to the number of wind farms with low correlation are depicted in Figs. 2 and 3. As introduced from Step 1 to Step 4, the prediction approaches for outputs of wind farm clusters with high correlation and low correlation are, respectively, studied, then the regional output prediction can be obtained. When the number of wind farms with low correlation is 0, all wind farms are considered as high correlation wind farms. When the number of wind farms with low correlation is 167, all wind farms are considered as low correlation wind farms. NWP data and historical power observations are considered as model input for output analysis of wind farms with high correlation, while only historical power observations are considered as model input for the output analysis of wind farms with low correlation. By comparing the prediction errors, the numerical comparison reveals the impact of different input data utilizations on prediction accuracy. In addition, considering the smoothing effect which can reduce the adverse effect caused by a sudden change in some wind farms' output [40], the wind farm cluster output is directly studied in the proposed prediction approach, rather than adding the power prediction results of each wind farm to calculate the cluster output. As shown in the figures, considering both MAEs and RMSEs for the deterministic predictions with 60-min look-ahead time [18], [21], the prediction models with 92 and 82 low correlation

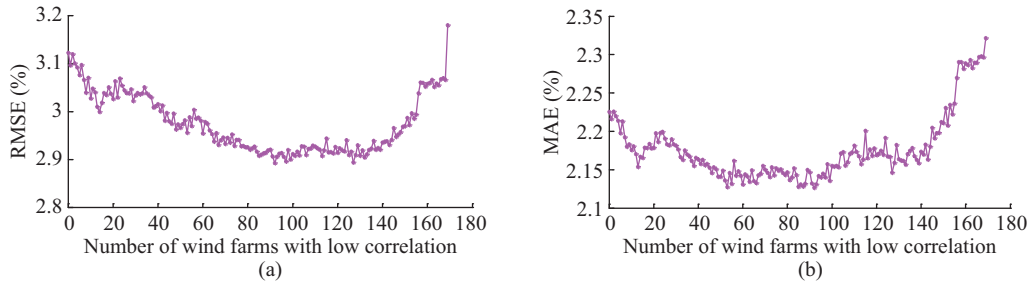


Fig. 2. The prediction errors of regional outputs in June. (a) RMSE. (b) MAE.

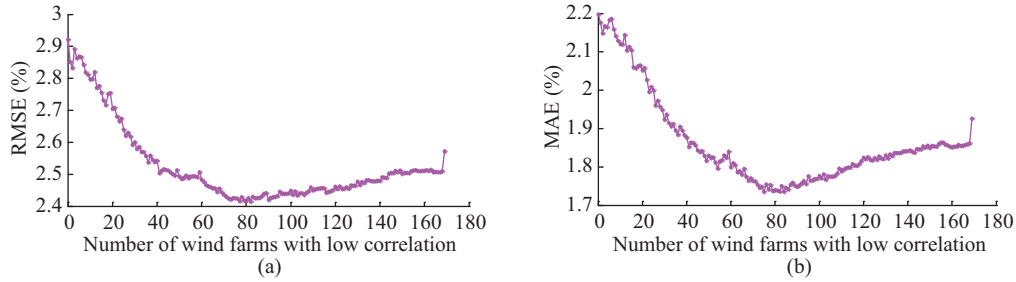


Fig. 3. The prediction errors of regional outputs in November. (a) RMSE. (b) MAE.

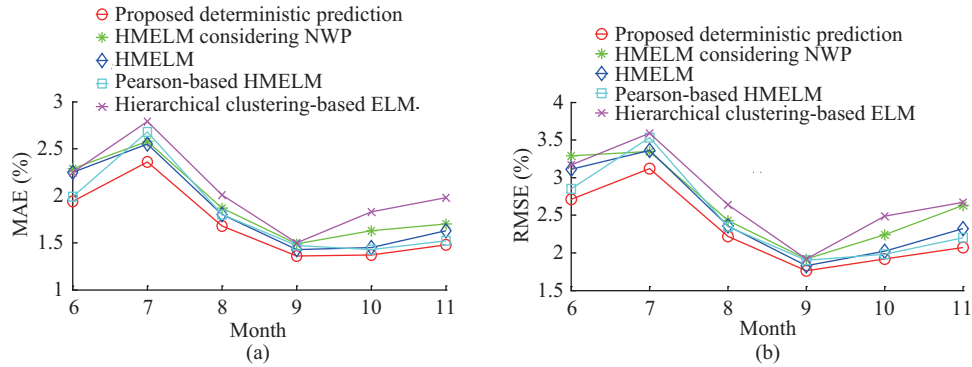


Fig. 4. The comparison of deterministic predictions. (a) MAE. (b) RMSE.

wind farms have the best accuracies in June and November, respectively. The numerical comparison reveals it is significant to analyze the utilization of NWP data before constructing the deterministic prediction model. Based on the numerical comparison of prediction accuracies, the optimal division of wind farms can be performed.

To verify the effectiveness of the proposed deterministic prediction based on DPCCA, several prediction approaches with a 60-min look-ahead time from June to November are utilized for numerical comparison, including the correlation-based HMELM which divides the wind farms by Pearson correlations, the HMELMs, respectively, considering or not considering NWP data as input variables, and the hierarchical clustering-based approach. Considering both the accuracy and computational efficiency, the number of pair ELMs for HELM is set to 50. MAEs and RMSEs of all the methods from June to November are shown in Fig. 4. From the numerical results, the proposed deterministic prediction has the best prediction accuracy, which reveals the effectiveness of DPCCA and the analysis of NWP utilization. Thus, it is obvious that advanced

analysis between wind power and wind speed for each wind farm is needed.

C. Numerical Analysis of Forecasting Performance

In the proposed approach, OPNPIs are studied to quantify PIs, by directly optimizing the overall performance with improved reliability based on the regulating coefficient. In Fig. 5, the forecasting performances of regional outputs based on different regulating coefficients with PINC of 90% and 60-min look-ahead time in June and November are illustrated in the training process. With a resolution of 0.001, considering both reliability and overall performance of PIs, the optimal regulating coefficients are -0.001 and -0.003 , respectively. Reliability is evaluated by the absolute value of ACE according to (12), while the overall performance is evaluated by interval score according to (14) and (15). In the proposed method, the range of regulating coefficient with optimal interval score is first determined focusing only on the two decimal places, and then the regulating coefficient value corresponding to the minimum coverage deviation is selected from the range. The results of the proposed methods (PMs) without the regulating

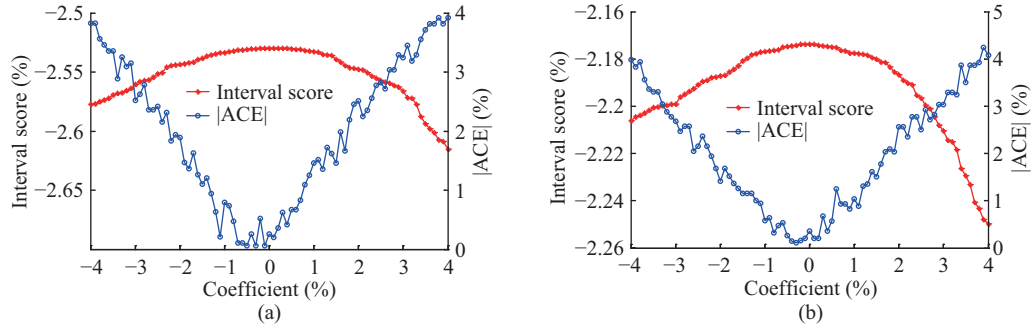


Fig. 5. PIs performances with regulating coefficients. (a) June. (b) November.

coefficient and with the optimal regulating coefficient are given in Table I. Furthermore, the conventional PIs approaches including direct quantile regression (DQR) [22], machine learning-based linear programming (MLLP) [27], bootstrap-based ELM (BELM) [21], and conditional probability-based PIs (CPPI) [18], meanwhile, the parametric PIs based on HMELM are utilized as benchmarks for numerical comparison. Among all the results, PM with an optimal regulating coefficient has the best forecasting performance. Even though both CPPI and BELM have accurate deterministic predictions and optimize the quantification of uncertainty, their performances are still affected by accuracy of the parametric error distribution assumption. Both DQR and MLLP are nonparametric probabilistic prediction approaches, which have good universality. MLLP, which considers the optimization of sharpness is an enhanced approach of DQR. According to the numerical comparison, MLLP and CPPI have better forecasting performances than DQR and BELM. As is shown in Table I, the forecasting performance of PM can be further enhanced by optimizing the regulating coefficient, based on the numerical comparison between PM ($K = 0$) and PM (Optimal K). Besides, to further verify the effectiveness of the proposed OPNPIS for error analysis, the parametric PIs based on the proposed deterministic prediction and normal distribution are utilized as a benchmark. The numerical comparison reveals the superiority of the nonparametric approach for uncertainty quantification.

TABLE I
RESULTS OF PIS PERFORMANCES

Method	Jun.		Nov.	
	PICP	Score	PICP	Score
DQR	88.54%	-0.0331	88.02%	-0.0186
MLLP	88.54%	-0.0299	88.28%	-0.0165
BELM	88.02%	-0.0331	88.80%	-0.0191
CPPI	88.80%	-0.0305	98.96%	-0.0182
HMELM	85.16%	-0.0298	98.70%	-0.0174
PM ($K = 0$)	88.54%	-0.0287	92.19%	-0.0152
PM (Optimal K)	89.32%	-0.0281	89.58%	-0.0148

Numerical comparisons of the existing methods are given in the case studies to reveal the effectiveness of the proposed method with 60-min and 90-min look-ahead times for very short-term forecasts of wind power. The utilization values of NWP data and the prediction performances vary with different look-ahead times. Thus, to reveal the PIs performances of

different look-ahead times is needed. Since 60-min and 90-min look-ahead times are both within a very short-term scale, thus the related forecasting performances are studied based on the same data utilization strategy. That is, in the proposed method, only the wind power observations are considered as model input for the deterministic prediction of low correlation wind farm output and probabilistic prediction, and only when studying the deterministic prediction of high correlation wind farm output, the wind power and wind speed data from NWP are considered as model input. To further verify the effectiveness of the proposed method, from June to November with PINCs of 90% and 95%, CPPI, MLLP, and the conventional PIs for regional outputs including the smoothing method [33] and the statistical upscaling method [29] are utilized for numerical comparison. Both the smoothing method and statistical upscaling method quantify PIs based on deterministic prediction and error assumption of normal distribution. The smoothing method quantifies the deterministic prediction based on the similarity of spatio-temporal correlation between wind speed and wind power. In the statistical upscaling method, representative wind farms are obtained based on the prediction accuracies and correlations, then, the regional output can be quantified. The PIs performances are shown in Table II and III with 60-min and 90-min look-ahead times, respectively. ‘None’ for AO means that all actual observations are within the PIs, that is, PICP is 100% [40]. Considering the dataset of NWP in inputs has low reliability for hourly forecasting [18], only the wind power observations are utilized as inputs of the existing probabilistic prediction approaches, including CPPI, MLLP and the statistical upscaling method. Among all the results of prediction models, the proposed PIs have the best reliability and overall performance. In the proposed method, OPNPIS are studied to optimize the uncertainty quantification, which helps improve PIs performance. The forecasting performance of the smoothing method is greatly influenced by the accuracy of NWP. Inaccurate prediction of wind speed will result in severe prediction deviation. The statistical upscaling method focuses on representative wind farms and the calculation formula of regional outputs. However, it is difficult to accurately select the representative wind farms and quantify the relationship between the outputs of representative wind farms and regional outputs. Besides, when the outputs of wind farms from the same cluster have a sudden and synchronously great change, the forecasting performance of the cluster will

TABLE II
COMPARISON OF PIS PERFORMANCES WITH 60-MIN LOOK-AHEAD TIME

PINC	Month	Smoothing method				CPPI				Statistical upscaling method				MLLP				Proposed method			
		PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score
90%	Jun.	100%	0.4086	None	-0.0817	88.80%	0.1171	0.0157	-0.0305	95.83%	0.2183	0.0111	-0.0456	88.54%	0.1204	0.0127	-0.0299	89.32%	0.1102	0.0143	-0.0281
	Jul.	100%	0.2906	None	-0.0581	80.47%	0.0736	0.0135	-0.0252	80.99%	0.0735	0.0126	-0.0243	95.31%	0.1142	0.0122	-0.0251	90.10%	0.0932	0.0114	-0.0232
	Aug.	92.45%	0.3906	0.0209	-0.0844	91.41%	0.0809	0.0182	-0.0225	88.54%	0.1818	0.0212	-0.0461	92.45%	0.0995	0.0119	-0.0235	89.58%	0.0789	0.0121	-0.0208
	Sep.	84.38%	0.2190	0.0557	-0.0786	94.01%	0.0750	0.0103	-0.0175	70.31%	0.1577	0.0540	-0.0957	92.19%	0.0843	0.0077	-0.0193	90.36%	0.0704	0.0075	-0.0170
	Oct.	97.14%	0.3095	0.0535	-0.0680	92.45%	0.1112	0.0206	-0.0284	84.90%	0.1772	0.0411	-0.0603	91.15%	0.0996	0.0151	-0.0253	90.36%	0.0909	0.0139	-0.0235
95%	Nov.	100%	0.3422	None	-0.0684	96.88%	0.0912	0.0140	-0.0187	92.45%	0.1679	0.0149	-0.0381	88.28%	0.0656	0.0072	-0.0165	89.58%	0.0605	0.0065	-0.0148
	Jun.	100%	0.4613	None	-0.0461	93.49%	0.1427	0.0145	-0.0181	99.48%	0.2608	0.0050	-0.0262	96.61%	0.1622	0.0082	-0.0171	94.27%	0.1300	0.0135	-0.0161
	Jul.	100%	0.3482	None	-0.0348	88.28%	0.0877	0.0126	-0.0147	88.28%	0.2871	0.0436	-0.0492	97.92%	0.1323	0.0123	-0.0143	95.83%	0.1141	0.0130	-0.0136
	Aug.	98.70%	0.4345	0.0088	-0.0439	93.75%	0.0949	0.0197	-0.0144	94.01%	0.2180	0.0203	-0.0267	96.09%	0.1224	0.0122	-0.0164	95.83%	0.1073	0.0144	-0.0131
	Sep.	98.18%	0.2947	0.0092	-0.0301	96.88%	0.0898	0.0113	-0.0104	73.44%	0.1872	0.0475	-0.0692	97.66%	0.1112	0.0029	-0.0117	94.79%	0.0881	0.0067	-0.0102
	Oct.	98.44%	0.3555	0.0131	-0.0364	95.57%	0.1332	0.0160	-0.0162	91.15%	0.2113	0.0343	-0.0368	95.83%	0.1371	0.0120	-0.0152	94.79%	0.1189	0.0128	-0.0146
	Nov.	100%	0.5826	None	-0.0583	98.44%	0.1088	0.0038	-0.0111	96.09%	0.2001	0.0062	-0.0210	93.75%	0.0876	0.0069	-0.0105	94.27%	0.0740	0.0046	-0.0084

TABLE III
COMPARISON OF PIS PERFORMANCES WITH 90-MIN LOOK-AHEAD TIME

PINC	Month	Smoothing method				CPPI				Statistical upscaling method				MLLP				Proposed method			
		PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score	PICP	AW	AO	Score
90%	Jun.	96.35%	0.3681	0.0388	-0.0793	87.76%	0.1564	0.0218	-0.0420	97.40%	0.2445	0.0084	-0.0498	87.76%	0.1604	0.0184	-0.0411	89.58%	0.1573	0.0182	-0.0391
	Jul.	97.14%	0.4759	0.0262	-0.0982	79.17%	0.1014	0.0165	-0.0340	83.85%	0.2478	0.0522	-0.0832	93.49%	0.1429	0.0176	-0.0332	91.15%	0.1253	0.0167	-0.0310
	Aug.	98.44%	0.4326	0.0235	-0.0880	84.64%	0.1091	0.0200	-0.0341	91.15%	0.2061	0.0263	-0.0505	95.57%	0.1421	0.0209	-0.0321	90.63%	0.1101	0.0161	-0.0281
	Sep.	94.01%	0.2494	0.0371	-0.0588	91.93%	0.1067	0.0174	-0.0270	69.58%	0.1742	0.0553	-0.0998	91.41%	0.1233	0.0101	-0.0281	89.84%	0.1061	0.0119	-0.0261
	Oct.	92.71%	0.3010	0.0641	-0.0791	91.93%	0.1578	0.0195	-0.0379	85.94%	0.2074	0.0441	-0.0663	91.15%	0.1476	0.0129	-0.0373	90.36%	0.1368	0.0183	-0.0344
95%	Nov.	98.70%	0.3301	0.0343	-0.0678	98.17%	0.1338	0.0054	-0.0272	93.75%	0.2018	0.0127	-0.0435	91.15%	0.0949	0.0104	-0.0226	90.89%	0.0947	0.0071	-0.0215
	Jun.	100%	0.5009	None	-0.0501	92.71%	0.1876	0.0162	-0.0235	100%	0.2912	None	-0.0291	96.09%	0.2063	0.0167	-0.0229	94.01%	0.1825	0.0181	-0.0226
	Jul.	100%	0.3377	None	-0.0338	87.50%	0.1211	0.0158	-0.0200	87.50%	0.2961	0.0423	-0.0508	97.66%	0.1394	0.0152	-0.0193	94.79%	0.1467	0.0186	-0.0185
	Aug.	99.74%	0.5102	0.0180	-0.0512	96.35%	0.1445	0.0343	-0.0195	94.27%	0.2510	0.0165	-0.0289	98.70%	0.1844	0.0259	-0.0198	95.57%	0.1569	0.0114	-0.0177
	Sep.	96.61%	0.3179	0.0165	-0.0340	94.01%	0.1386	0.0151	-0.0175	75.00%	0.2073	0.0449	-0.0656	96.35%	0.1497	0.0057	-0.0158	94.27%	0.1217	0.0121	-0.0149
	Oct.	98.70%	0.3639	0.1480	-0.0441	96.61%	0.1889	0.0205	-0.0217	91.15%	0.2476	0.0431	-0.0400	96.35%	0.2043	0.0092	-0.0218	94.01%	0.1682	0.0161	-0.0207
	Nov.	100%	0.4018	None	-0.0402	100%	0.1633	None	-0.0163	98.44%	0.2398	0.0030	-0.0242	96.88%	0.1230	0.0084	-0.0133	95.05%	0.1157	0.0060	-0.0127

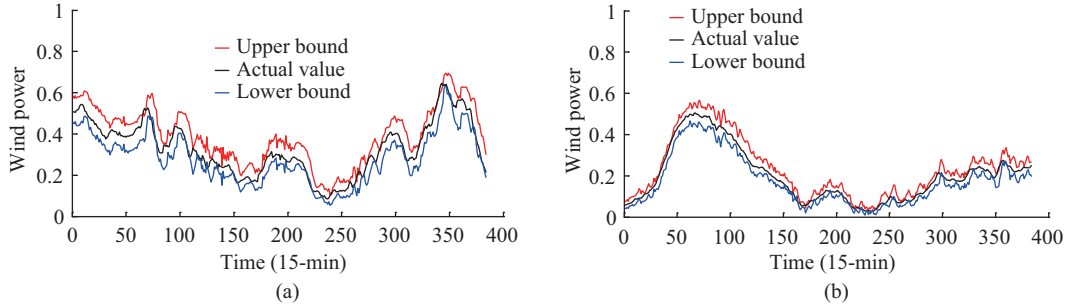


Fig. 6. PIs with 60-min look-ahead time and PINC of 90%. (a) June. (b) November.

have low accuracy. In practice, the combination of outputs from different wind farms considering the smoothing effect is helpful to reduce the influence caused by the sudden change of outputs [31]. A PC with Intel(R) Core(TM) i7-7700 CPU @ 2.8 GHz and 8 GB RAM is used for computations. The training times of the proposed deterministic prediction and probabilistic prediction are less than 200 and 300 seconds, respectively, while the prediction time of the deterministic and probabilistic prediction is less than 5 seconds. The proposed method has good computational efficiency and forecasting performance, so it is suitable for very short-term probabilistic prediction.

In Fig. 6, PIs with a 60-min look-ahead time and PINC of 90% in June and November are depicted. It can be found that most actual values lie within the lower and the upper bounds with a narrow average width of PIs, representing good sharpness. Also, as the given PINC is 90%, theoretically, there should be nearly 10% of wind power observations not covered by the PIs. The PICPs of Fig. 6(a) and (b), which are respectively, 89.32% and 89.58% given in Table II, reveal the good reliability of the proposed PIs. It is obvious that the

PIs of the proposed method have accurate coverage and good sharpness which can fulfill the practical application.

V. CONCLUSION

In this paper, an approach of very short-term nonparametric PIs for regional wind power generation based on OPNPIs has been proposed. This paper has mainly focused on the optimized uncertainty analysis based on the deterministic prediction errors, considering the nonparametric quantification for the universality and the optimal performance including the overall performance and reliability of PIs, specifically, to complete the following four tasks: First, the utilization analysis of NWP has been studied, with the numerical analysis of DPCCA between the wind speed and wind power, for the reasonable division of wind farms. In addition, the necessary discussion for the utilization of NWP has been revealed by the numerical comparison of deterministic prediction. Second, OPNPIs based on deterministic prediction have been proposed to fulfill the uncertainty quantification with excellent and efficient nonparametric PIs. Third, the OPNPIs have been enhanced by the optimal regulating coefficient to further

improve the forecasting performance, which has been validated by numerical results. Finally, effectiveness of proposed PIs, based on OPNPIS for regional outputs, has been verified by the numerical comparison with other existing forecasting approaches. The obtained results have illustrated the proposed uncertainty analysis based on OPNPIS can provide better forecasting performance based on the prediction errors.

Since several efficient deterministic prediction approaches have been studied, the proposed uncertainty analysis has high potential for practical applications in risk assessment and systems operation. Hence, future work may focus on the advanced optimization of uncertainty analysis, and the optimized criteria of PIs.

REFERENCES

- [1] S. X. Wang, X. Zhao, H. Wang, and M. Li, "Small-world neural network and its performance for wind power forecasting," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 2, pp. 362–373, Jun. 2020.
- [2] Y. Wang, Y. H. Sun, and V. Dinavahi, "Robust forecasting-aided state estimation for power system against uncertainties," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 691–702, Jan. 2020.
- [3] G. Mu, F. Z. Wu, X. P. Zhang, Z. H. Fu, and B. Xiao, "Analytic mechanism for the cumulative effect of wind power fluctuations from single wind farm to wind farm cluster," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 5, pp. 1290–1301, Sep. 2022.
- [4] L. Ye, C. H. Zhang, Y. Tang, W. Z. Zhong, Y. N. Zhao, P. Lu, B. X. Zhai, H. B. Lan, and Z. Li, "Hierarchical model predictive control strategy based on dynamic active power dispatch for wind power cluster integration," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4617–4629, Nov. 2019.
- [5] C. F. Zhao, C. Wan, and Y. H. Song, "An adaptive bilevel programming model for nonparametric prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 424–439, Jun. 2020.
- [6] X. Y. Feng, S. J. Lin, Y. T. Liang, G. S. Fan, and M. B. Liu, "Distributed and risk-averse ADP algorithm for stochastic economic dispatch of power system with multiple offshore wind farms," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 5, pp. 1977–1993, Sep. 2024.
- [7] J. J. Yang, M. Yang, M. X. Wang, P. J. Du, and Y. X. Yu, "A deep reinforcement learning method for managing wind farm uncertainties through energy storage system control and external reserve purchasing," *International Journal of Electrical Power & Energy Systems*, vol. 119, pp. 105928, Jul. 2020.
- [8] X. Fang, B. M. Hodge, E. S. Du, C. Q. Kang, and F. X. Li, "Introducing uncertainty components in locational marginal prices for pricing wind power and load uncertainties," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2013–2024, May 2019.
- [9] Y. N. Zhao, L. Ye, P. Pinson, Y. Tang, and P. Lu, "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5029–5040, Sep. 2018.
- [10] C. Wan, Z. Xu, and P. Pinson, "Direct interval forecasting of wind power," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4877–4878, Nov. 2013.
- [11] Y. B. Chen, Z. Zhang, H. Chen, and H. P. Zheng, "Robust UC model based on multi-band uncertainty set considering the temporal correlation of wind/load prediction errors," *IET Generation, Transmission & Distribution*, vol. 14, no. 2, pp. 180–190, Jan. 2020.
- [12] Y. Lin, M. Yang, C. Wan, J. H. Wang, and Y. H. Song, "A multi-model combination approach for probabilistic wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 226–237, Jan. 2019.
- [13] S. Wang, Y. H. Sun, S. M. Zhang, Y. Zhou, D. C. Hou, and J. X. Wang, "Very short-term probabilistic prediction of PV based on multi-period error distribution," *Electric Power Systems Research*, vol. 214, pp. 108817, Jan. 2023.
- [14] Y. S. Xue, C. Yu, K. Li, F. S. Wen, Y. Ding, Q. W. Wu, and G. Y. Yang, "Adaptive ultra-short-term wind power prediction based on risk assessment," *CSEE Journal of Power and Energy Systems*, vol. 2, no. 3, pp. 59–64, Sep. 2016.
- [15] G. Sideratos and N. D. Hatzigiorgiou, "Probabilistic wind power forecasting using radial basis function neural networks," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1788–1796, Nov. 2012.
- [16] T. Y. Hu, W. C. Wu, Q. L. Guo, H. B. Sun, L. B. Shi, and X. W. Shen, "Very short-term spatial and temporal wind power forecasting: A deep learning approach," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 2, pp. 434–443, Jun. 2020.
- [17] Y. X. Yu, X. S. Han, M. Yang, and J. J. Yang, "Probabilistic prediction of regional wind power based on spatiotemporal quantile regression," *IEEE Transactions on Industry Applications*, vol. 56, no. 6, pp. 6117–6127, Nov./Dec. 2020.
- [18] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033–1044, May 2014.
- [19] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
- [20] Y. Lan, Y. C. Soh, and G. B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing*, vol. 72, no. 13–15, pp. 3391–3395, Aug. 2009.
- [21] Y. H. Sun, P. Wang, S. W. Zhai, D. C. Hou, S. Wang, and Y. Zhou, "Ultra short-term probability prediction of wind power based on LSTM network and condition normal distribution," *Wind Energy*, vol. 23, no. 1, pp. 63–76, Oct. 2020.
- [22] C. Wan, J. Lin, J. H. Wang, Y. H. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, Jul. 2017.
- [23] C. Wan, J. Lin, Y. H. Song, Z. Xu, and G. Y. Yang, "Probabilistic forecasting of photovoltaic generation: An efficient statistical approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2471–2472, May 2017.
- [24] L. Y. Liu, Y. Zhao, D. L. Chang, J. Y. Xie, Z. Y. Ma, Q. Sun, H. Y. Yin, and R. Wennersten, "Prediction of short-term PV power output and uncertainty analysis," *Applied Energy*, vol. 228, pp. 700–711, Oct. 2018.
- [25] J. Vilar, G. Aneiros, and P. Raña, "Prediction intervals for electricity demand and price using functional data," *International Journal of Electrical Power & Energy Systems*, vol. 96, pp. 457–472, Mar. 2018.
- [26] C. F. Zhao, C. Wan, Y. H. Song, and Z. J. Cao, "Optimal nonparametric prediction intervals of electricity load," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2467–2470, May 2020.
- [27] C. Wan, J. H. Wang, J. Lin, Y. H. Song, and Z. Y. Dong, "Nonparametric prediction intervals of wind power via linear programming," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1074–1076, Jan. 2018.
- [28] Y. Zhou, Y. H. Sun, S. Wang, L. Q. Bai, D. C. Hou, R. J. Mahfoud, and P. Wang, "Very short-term probabilistic prediction method for wind speed based on ALASSO-nonlinear quantile regression and integrated criterion," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 6, pp. 2121–2129, Nov. 2023.
- [29] M. B. Ozkan and P. Karagoz, "Data mining-based upscaling approach for regional wind power forecasting: Regional statistical hybrid wind power forecast technique (regionalSHWIP)," *IEEE Access*, vol. 7, pp. 171790–171800, Nov. 2019.
- [30] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and C. Cornaro, "Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data," *Solar Energy*, vol. 158, pp. 1026–1038, Dec. 2017.
- [31] M. Yang, L. B. Zhang, Y. Cui, Y. Zhou, Y. L. Chen, and G. G. Yan, "Investigating the wind power smoothing effect using set pair analysis," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1161–1172, Jul. 2020.
- [32] M. Yang, X. X. Chen, and B. Y. Huang, "Ultra-short-term multi-step wind power prediction based on fractal scaling factor transformation," *Journal of Renewable and Sustainable Energy*, vol. 10, no. 5, pp. 053310, Sep. 2018.
- [33] M. G. Lobo and I. Sanchez, "Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1990–1997, Nov. 2012.
- [34] G. Sideratos and N. D. Hatzigiorgiou, "An advanced statistical method for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 258–265, Feb. 2007.
- [35] Q. Y. Xu, D. W. He, N. Zhang, C. Q. Kang, Q. Xia, J. H. Bai, and J. H. Huang, "A short-term wind power forecasting approach with

adjustment of numerical weather prediction input by data mining,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1283–1291, Oct. 2015.

- [36] A. Kusiak and W. Y. Li, “Estimation of wind speed: A data-driven approach,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 98, no. 10–11, pp. 559–567, Oct./Nov. 2010.
- [37] G. F. Gu and W. X. Zhou, “Detrending moving average algorithm for multifractals,” *Physical Review E*, vol. 82, no. 1, pp. 011136, Jul. 2010.
- [38] B. Mohandes, M. S. El Moursi, N. Hatzigargyriou, and S. El Khatib, “A review of power system flexibility with high penetration of renewables,” *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3140–3155, Jul. 2019.
- [39] J. K. Zhu, S. C. H. Hoi, and M. R. T. Lyu, “Robust regularized kernel regression,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 6, pp. 1639–1644, Dec. 2008.
- [40] Y. Zhou, Y. H. Sun, S. Wang, R. J. Mahfoud, H. H. Alhelou, N. Hatzigargyriou, and P. Siano, “Performance improvement of very short-term prediction intervals for regional wind power based on composite conditional nonlinear quantile regression,” *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 1, pp. 60–70, Jan. 2022.
- [41] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95-International Conference on Neural Networks*, Perth, WA, Australia, 1995, pp. 1942–1948.



Sen Wang is currently pursuing the Ph.D. degree in Electrical Engineering with Hohai University, Nanjing, China. His research interests include data analysis, renewable energy grid-connected control, power system operation optimization, and renewable energy generation forecasting.



Rabea Jamil Mahfoud received the Ph.D. degree in Electrical Engineering at Hohai University, Nanjing, China, in 2020. His current research interests include optimal operation and planning of power systems, optimization algorithms, and data analysis.



Yan Zhou received the Ph.D. degree from Hohai University, Nanjing, China, in 2020. His research interests include data analysis, machine learning, and renewable energy generation forecasting.



Dongchen Hou received the B.S. degree in Electrical Engineering and Automation from North China University of Water Resources and Electric Power, Henan, China, in 2017. He is currently pursuing the Ph.D. degree in Electrical Engineering with Hohai University, Nanjing, China. His research interests include theoretical and algorithmic studies in power system estimation, parameters identification, power system dynamics, signal processing, and cyber security.



Yonghui Sun received the Ph.D. degree from the City University of Hong Kong, Hong Kong, China, in 2010. He is currently a Professor with the College of Energy and Electrical Engineering, Hohai University, Nanjing, China. He has authored more than 80 papers in refereed international journals. His research interests include stability analysis and control of power systems, optimal planning and operation of integrated energy system, optimization algorithms, and data analysis.



Jianxi Wang is currently pursuing the Ph.D. degree in Electric Engineering with Hohai University, Nanjing, China. His research interests include multi-energy load forecasting, planning and operation optimization of integrated energy systems, optimization algorithms and data analysis.