IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted November 4, 2019, date of publication November 8, 2019, date of current version November 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952555

Wind Power Prediction Based on LSTM Networks and Nonparametric Kernel Density Estimation

BOWEN ZHOU, (Member, IEEE), XIANGJIN MA, YANHONG LUO, (Member, IEEE), AND DONGSHENG YANG, (Senior Member, IEEE)

College of Information Science and Engineering, Northeastern University, Shenyang 110000, China

Corresponding author: Yanhong Luo (luoyanhong@ise.neu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703081, in part by the Liaoning Revitalization Talents Program under Grant XLYC1801005, in part by the Natural Science Foundation of Liaoning Province under Grant 20170520113, and in part by the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources under Grant LAPS19005.

ABSTRACT Wind energy is a kind of sustainable energy with strong uncertainty. With a large amount of wind power injected into the power grid, it will inevitably affect the security, stability and economic operation of the power grid. High-precision wind power spot prediction and fluctuation interval information can provide more adequate decision-making support for grid scheduling and optimization. Hence, this paper proposes a K-Means-long short-term memory (K-Means-LSTM) network model for wind power spot prediction, and a nonparametric kernel density estimation (KDE) model with bandwidth optimization for wind power probabilistic interval prediction. The long short-term memory (LSTM) network has a strong memory function, which can establish the correlation between the data before and after, so as to improve the prediction accuracy. The K-Means clustering method forms different clusters of wind power impact factors to generate a new LSTM sub-prediction model. The optimization of the bandwidth in the nonparametric KDE is implemented by the mean integrated squared error criterion. In addition, a part of the dataset is deliberately demarcated from the wind power historical dataset to generate reasonable wind power prediction errors. The simulation results show that the proposed K-Means-LSTM network model has higher prediction accuracy than the back propagation (BP) neural networks, Elman neural networks, support vector regression (SVR) and LSTM network models. Compared with the KDE model with random bandwidth and the Gaussian distribution model, the bandwidth optimization model proposed in this paper has more narrow prediction intervals with higher interval coverage rates.

INDEX TERMS Wind power prediction, Kernel density estimation, long short-term memory, K-means clustering, probabilistic interval prediction.

I. INTRODUCTION

In recent years, with the increasing attention of countries around the world to the development of renewable energy, the research and development of wind energy is increasing day by day [1]. Wind energy is a clean energy source from nature. However, it has strong uncertainties [2]. At present, the research related to wind energy can be mainly divided into wind speed prediction [3], wind power prediction [4], wind power ramp event prediction [5], [6], and the optimal configuration of wind energy [7], [8], etc. Along with the massive grid connection of wind power, it has a huge impact on the safe and stable operation of the power grid [9]. Therefore, to realize the highly accurate prediction of wind power will provide more beneficial information for the optimal dispatch of the power grid.

At present, the research directions of wind power prediction can be divided into spot prediction and probabilistic interval prediction [10]. There are many spot prediction methods for wind speed and wind power of wind farms. In reference [11], a prediction method combining wavelet transform (VT) method and support vector machine (SVM) was proposed. Wang *et al.* [12] proposed the maximum correntropy criteria (MCC) algorithm instead of the traditional minimum square error (SE) algorithm to improve the back propagation (BP) neural network and to realize wind power

The associate editor coordinating the review of this manuscript and approving it for publication was Mingjian Cui.

prediction. Huang *et al.* [13] proposed an enhanced harmony search (EHS) algorithm to implement the selection of support vector regression (SVR) model parameters. Chen *et al.* [14] made correlation research on wind speed prediction, which was based on extreme learning machine (ELM), Elman Neural Network and LSTM Network. Recently, A large number of new methods based on LSTM for renewable energy and load forecasting are proposed. Han et al. [15] proposed a prediction model based on variational mode decomposition and LSTM. Yu *et al.* [16] proposed an enhanced forget-gate LSTM Model. Han *et al.* [17] proposed a model based on Copula function and LSTM, which can effectively extract wind power impact factors and achieve better prediction results. However, the research objects of the above literatures are based on wind power historical data, and the wind power output series at the next moment is obtained by constructing supervised learning. Therefore, the actual predicted objects are only the wind power values at the next moment, and there is a significant delay of single unit time. In addition, due to the strong randomness of wind energy, wind power prediction without the combination of numerical weather prediction (NWP) data generally lacks practicality. Thus, this paper will conduct an in-depth study from the perspective of the NWP data and the LSTM network model to achieve wind power spot prediction.

The wind power spot prediction still poses a great threat to the safety and stability of the power grid. Thus, the probabilistic interval prediction of wind power emerges. Probabilistic interval prediction presents the fluctuation range of wind power prediction error and provides more beneficial auxiliary information for grid scheduling and energy optimization [18], [19]. In general, the probabilistic interval prediction of wind power can be obtained by parametric and nonparametric methods [20]. The parametric method usually assumes that the sample data obeys a certain type of distribution, such as Beta distribution, Gaussian distribution, etc. Yuan *et al.* [21] assumed that the prediction error obeyed the Beta distribution, and used the particle swarm optimization (PSO) algorithm to optimize the parameters in the Beta distribution to find the minimum probability interval. Zhang *et al.* [22] proposed a mixed Gaussian distribution model to fit the wind power prediction error. Besides, the learning model represented by sparse Bayesian learning (SBL) belongs to the parametric methods [23], which usually requires an accurate prior knowledge of the relevant parameters. The SBL model is complex with many parameter settings, which also has a strong dependence on the sample data. The parametric method is often constrained by the hypothetical prediction error model. If the assumption is unreasonable, it will lead to unsatisfactory interval prediction results.

On the contrary, the nonparametric method is more precise than the parametric method, since the nonparametric method does not need to make the assumptions of the wind power error distribution. As a result, the conditional constraints of the nonparametric method are limited. It is suitable for many applications, and its versatility is stronger than the parametric method. Nonparametric methods mainly include quantile regression and kernel density estimation (KDE) methods. Haque *et al.* [24] proposed the firefly algorithm to optimize the combined prediction algorithm, using the quantile regression method to achieve probability prediction. In Reference [25], an online quantile regression method based on Reproducing Kernel Hilbert Space was proposed, which enabled online learning and online calibration. The KDE method can provide the probability density function (PDF) of wind power prediction error. In recent years, it has attracted much attention in the area of wind power probability prediction. Taylor and Jeon [26] used the conditional KDE method to achieve wind power density prediction, which showed that the KDE method had great application advantages. He and Li [27] obtained the PDF of wind power on the prediction period by using the prediction results based on different prediction conditions as the input of the KDE model. In reference [28], the KDE method was used to predict the wind power based on the data of a wind farm in France. It was concluded that the nonparametric KDE had better performance in sensitivity and reliability than the quantile regression method. Therefore, this paper will apply the KDE method to study the probabilistic interval prediction of wind power.

Different spot prediction methods have different prediction advantages. Combined forecasting methods can take advantage of each sub-model and have attracted much attention in recent years [29], [30]. In the current researches of probabilistic interval prediction methods, the wind power error data comes from the deviation between the predicted values and the real values within the prediction method adopted by the wind farms. However, there is a problem worthy of further study, which is the rationality of the error data. The wind power spot prediction methods applied in many literatures are often inconsistent with the theoretical research methods in practical applications. Such error data cannot represent the error dataset generated by the actual prediction method. In view of the above problems, this paper takes the measured data of wind farms as the micro-scale NWP data, mainly focusing on the following contributions:

a) The abnormal data of wind power historical data has a great influence on wind power prediction. This paper proposes an anomaly data processing method combining density clustering of applications with noise (DBSCAN) and 3σ technology to achieve efficient denoising.

b) A spot prediction method based on long short-term memory (LSTM) networks for wind power is proposed. Combined with the K-Means clustering method due to its fast and accurate performance in prediction [31], wind power impact factors clustering is implemented to generate K-Means-LSTM combined prediction model. Meanwhile, the inclusion of the NWP data of the current time eliminates the step delays in the prediction series.

c) The wind power historical dataset is divided into training, verification, error generation, and test dataset, where the error dataset is generated to ensure the rationality for the

nonparametric KDE method with bandwidth optimization, so as to obtain the rational wind power probabilistic interval prediction. Compared with the nonparametric KDE method with random bandwidth and the parametric method with the assumed error obeying the Gaussian distribution, the validity of the nonparametric KDE method with bandwidth optimization has been verified.

The remains of this paper are organized as follows. Section II initially introduces preliminary data processing by using DBSCAN and 3σ technology. The dataset division technique is also provided. Next, Section III delivers the proposed spot prediction method based on the K-Means-LSTM network. Afterward, Section IV proposes the nonparametric KDE-based probabilistic interval prediction method. The prediction evaluation indices are presented, respectively. Section V provides numerical case studies, and finally, Section VI concludes the whole paper.

II. DATA PROCESSING

A. RAW DATASET PROCESSING

In the actual operation of the wind farm, some wind power abnormal data will be generated due to wind turbine load shedding and outage, equipment failure, communication interference, and human malfunction, which are represented by not a number (NAN) and missing data. Abnormal data processing is usually performed by methods such as culling, replacement, and interpolation. The method of substitution (or interpolation) is further divided into replacement (or interpolation) of adjacent data of the same type, mean value replacement (or interpolation), and replacement (or interpolation) based on maximum likelihood estimation.

Since the number of the abnormal data samples is tiny compared to the sample capacity, the impact caused by the distribution and application of the dataset is negligible. Hence, this paper uses the culling method as abnormal data processing.

B. IMPACT FACTORS ANALYSIS

Pearson correlation coefficient analysis can explore the degree of correlation between various impact factors and wind power, so as to select the appropriate impact factors for the model as input data. The application formula for the Pearson correlation coefficient r_{jk} is:

$$r_{jk} = \frac{\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)\left(x_{ik} - \bar{x}_k\right)}{\sqrt{\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2} \sqrt{\sum_{i=1}^{n} \left(x_{ik} - \bar{x}_k\right)^2}} \tag{1}$$

where, x_{ij} and x_{ik} represent the i-th value of the class j and class k data, respectively; $\bar{x}_j$ and $\bar{x}_k$ represent the mean of the class j and class k data, respectively.

The data in this paper is derived from the measured data of two wind farms in the northeast of China. In addition to wind power, it also includes wind speed, wind direction, temperature, humidity, and pressure. The wind speed data adopts the average value of the measured wind speed. The wind direction data uses the measured average wind direction cosine value, and the remaining impact factors are the measured values at the moment. Fig. 1 shows the heat map of Pearson correlation coefficient analysis for a set of datasets. It can be seen from Fig.1 that the correlation between wind speed and power is the highest, reaching nearly 0.9, followed by the wind direction and air pressure, which are negatively correlated with power. The difference between the two is small, and the correlation coefficient values are -0.1592 and -0.1570, respectively. The degree of correlation between humidity and power has not yet reached 0.1, and the degree of correlation is weak.

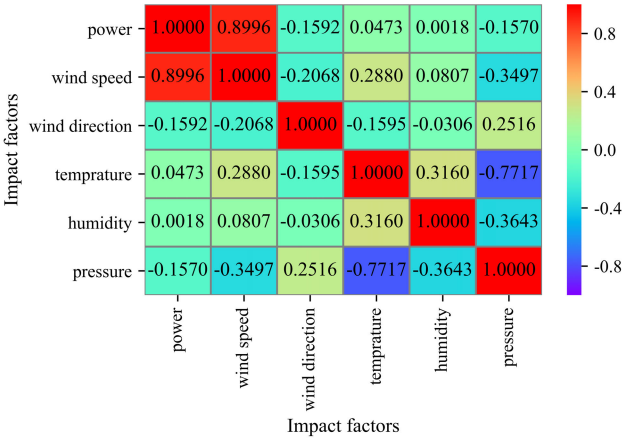


FIGURE 1. Pearson correlation coefficients between wind power and impact factors.

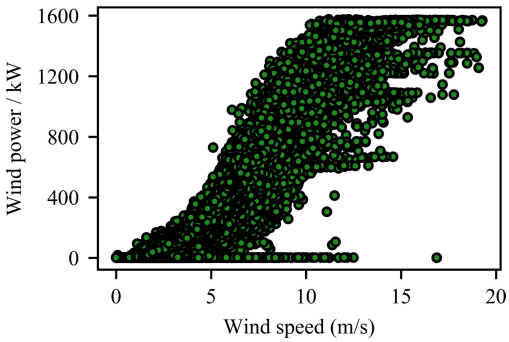


FIGURE 2. Wind speed-wind power curve of the wind turbine.

C. DATA PROCESSING BASED ON DBSCAN AND 3σ TECHNOLOGY

Wind speed is the most relevant impact factor, and its relationship with power needs to be clarified. Fig. 2 shows the relationship between wind speed and power. Some data deviates significantly from the speed-power curve of the wind turbine, in the form of various types of abnormal noises and data groups.

It is apparently not effective to directly use the raw dataset to ensure the reliability of wind data. Therefore, this paper adopts a data processing method based on DBSCAN and

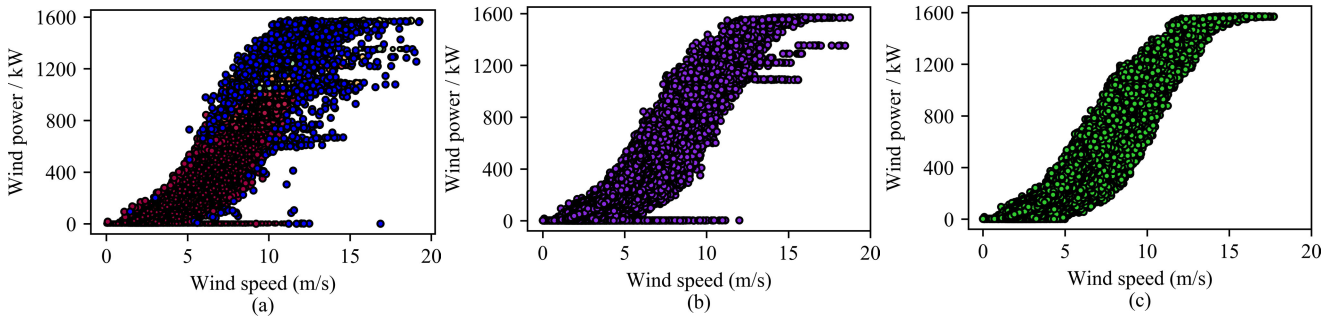


FIGURE 3. Wind speed-wind power curve changes during abnormal data processing. (a) In the process of DBSCAN. (b) Denoising based on DBSCAN. (c) Abnormal value culling based on 3σ technology.

3σ technology, which can effectively identify and eliminate noises and abnormal data clusters.

For a set of sample datasets $D = (x_1, x_2, \ldots, x_i)$, which contains only wind speed and power, and where, x_i represents the i-th data and the value range of i is: $i = 1, 2, \ldots, n$. The neighborhood radius value ε and the neighborhood density threshold M_t are determined according to the data type, and appropriate distance metrics are selected, such as Euclidean distance, Manhattan, Chebyshev, etc. The main steps of the specific DBSCAN processing include: establishment of an initialization state, finding all core objects, detection of core objects, generation of cluster sample sets, judgment on whether the generated clusters are completed, and finding all ε neighborhood subsample sets.

Although the effect of clustering to remove abnormal data is significant, some abnormal data cannot be effectively eliminated. Thus, this paper introduces 3σ technology to remove such data. The principle of the method is to determine whether the power value is stable within the range of $\pm 3\sigma$ of the average power $\bar{p}$ for a set of wind power $\{p_n\}$, which means the range is $[\bar{p} - 3\sigma, \bar{p} + 3\sigma]$, where, σ is the standard deviation of the power sequence and $\bar{p}_t$ is calculated by multi-step sliding calculation.

$$\bar{p}_t = \alpha p_t + (1 - \alpha)\, \bar{p}_{t-1} \tag{2}$$

where, $\bar{p}_t$ is the mean of power in step t, the starting power point p_0 is the $\bar{p}_0$ value, and α is the sliding weight parameter, which needs to be provided in advance. The calculated power averages of the respective steps are respectively determined according to (3).

$$\bar{p}_{t-1} - k\sigma < p_t < \bar{p}_{t-1} + k\sigma \tag{3}$$

where, k is the allowable deviation coefficient, which can be taken as 1, 2 and 3. Fig. 3 shows the effect of DBSCAN and 3σ techniques to remove abnormal data.

D. DIVISION OF DATASETS

The division of the dataset is an important part, and it is the preliminary step to realize the training of wind power data. In this section, the measured meteorological data around the wind turbine is adopted to represent the NWP data,

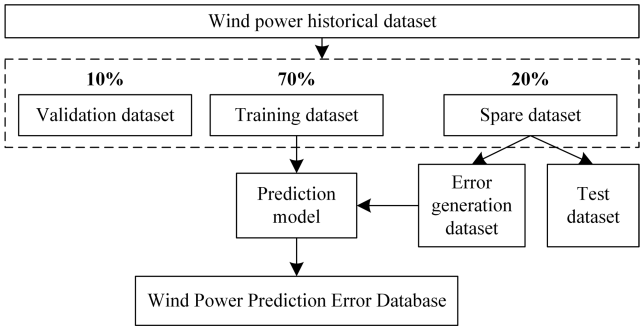


FIGURE 4. Division of wind power historical datasets.

and together with the wind power data constitutes the wind power historical dataset. After the abnormal data processing, the wind power historical dataset generally needs to be divided into three sub-datasets, such as the training dataset, the verification dataset, and the test dataset.

In the general probabilistic interval prediction study, the wind power error dataset takes the differences between the actual values and the theoretical power prediction values. However, the theoretical prediction method may not be the same as the spot prediction method used in actual research. In addition, the spot power values obtained by different prediction methods are different, which will result in the data distribution of the wind power prediction error also being different due to different prediction methods. Thus, the wind power prediction error dataset based on theoretical prediction values is not suitable for general research.

In order to obtain the reasonable wind power prediction error dataset, the wind power dataset is divided into the following parts according to the ratio of 7:1:2 according to the steps shown in Fig. 4: the training dataset, the validation dataset and the spare dataset. Among them, the spare dataset contains the error generation dataset and the test dataset. The number of samples in the test dataset is m. Compared with the general division of dataset, a part of the error generation dataset is obtained, which can obtain the related predicted power error distribution based on different prediction algorithms, and prepare for the follow-up data work for the wind power probabilistic interval prediction.

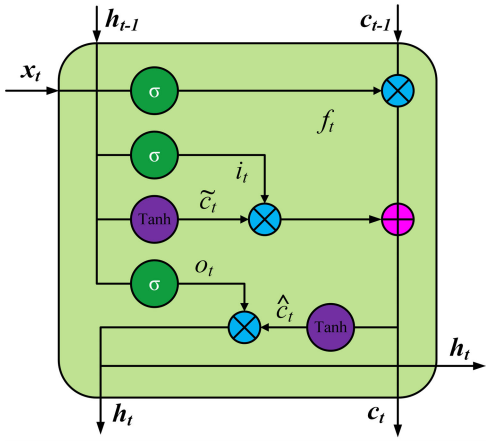


FIGURE 5. Internal unit structure of the LSTM network.

III. WIND POWER SPOT PREDICTION

A. LONG SHORT-TERM MEMORY

The LSTM network is a special recurrent neural network (RNN) with three thresholds, namely the input gate, the output gate and the forgetting gate. The unit structure of the LSTM network is shown in Fig. 5.

The forgetting gate determines the increase or decrease of the data flow by setting the threshold, which means reservation and forgetting. Since the hidden layer of the RNN has only one state, there are serious problems with gradient disappearance and gradient explosion. On the basis of RNN, LSTM adds the structure of the cell state, which can realize the long-term preservation of the state and highlights the strong memory function of the LSTM network. In the case of massive wind power data mining, the network can greatly improve the accuracy of wind power prediction.

In the forward propagation process of the LSTM network, the output value of the forgetting gate f_t can determine the trade-off of the unit state information, and the related functional relationship is obtained as (4).

$$f_t = \sigma\left(w_f h_{t-1} + u_f x_t + b_f\right) \tag{4}$$

Two variables i_t and $\tilde{c}_t$ are produced by the input gate, which are related to the previous moment. The expressions are as shown as (5) and (6).

$$i_t = \sigma\left(w_i h_{t-1} + u_i x_t + b_i\right) \tag{5}$$

$$\tilde{c}_t = \tanh\left(w_c h_{t-1} + u_c x_t + b_c\right) \tag{6}$$

Cell state c_t is the transmission center of the cell state before and after LSTM, which has the following functional relationship:

$$c_t = c_{t-1} \odot f_t + i_t \odot \tilde{c}_t \tag{7}$$

The output h_t of the output gate comes from two parts. One part is the output of the previous moment and the input of the current moment, and the other part is the information of the current cell state, and the specific expression form is achieved as (8) and (9).

$$o_t = \sigma\left(w_0 h_{t-1} + u_0 x_t + b_0\right) \tag{8}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{9}$$

where, u and w are the weight values; b is the bias values; σ is the activation function and the sigmoid function is applied in this paper; $\odot$ is the Hadamard product.

B. K-MEANS CLUSTERING PREDICTION MODEL

Section II.B has illustrated that wind speed is the key factor affecting wind power among the five factors, and the daily wind speed section is widely distributed. In order to fully exploit the data information, based on the wind speed segment, this paper uses the K-Means clustering algorithm to divide the wind power and the impact factor dataset into K subclasses. Before the clustering, because of the great differences in numerical grades of each type of data, the effective training of the late model is a huge challenge. Hence, the data is normalized before the K-Means clustering, and the normalization used in this paper is shown as (10).

$$z' = \frac{z - z_{\min}}{z_{\max} - z_{\min}} \tag{10}$$

where, $z_{\max}$ and $z_{\min}$ represent the maximum and minimum values of certain types of data, respectively. z and z' represent the values before and after normalization of the data, respectively.

The flow chart of the K-Means clustering algorithm is shown in Fig. 6, which can be divided into the following four parts:

a) Given a set of input datasets D, determine the number of clusters K and the appropriate distance metrics.

b) Randomly extract K samples from the dataset as the initial cluster center, denoted as $m_1 = \{m_{11}, m_{12}, \ldots, m_{1k}\}$. Then, the distances of each sample in the dataset D to m_{1n} are calculated and classified into the category with the shortest distance to form a cluster set $C_1 = \{C_{11}, C_{12}, \cdots, C_{1k}\}$.

c) Calculate a new cluster center $m_2 = \{m_{21}, m_{22}, \cdots, m_{2k}\}$ for each cluster C_{1n} in the set C_1, and calculate the distance from each sample in the dataset D to m_{2n} as described in Step b to generate a new cluster $C_2 = \{C_{21}, C_{22}, \ldots, C_{2k}\}$.

d) Repeat Step c until the cluster center does not change anymore, namely $m_i = m_{i+1}$, indicating that the clustering has been completed and then output the clustering result.

The K-Means clustering method is simple in principle, easy to implement, and has obvious clustering effects. Even if the application object is massive data, it performs very fast and is an efficient clustering method.

C. SPOT PREDICTION PROCESS

The wind power spot prediction framework proposed in this paper is illustrated in Fig. 7, which can be divided into the following five parts: abnormal data processing, dataset division and acquisition, K-Means clustering based on the impact factors, LSTM network modeling, and testing experiments.

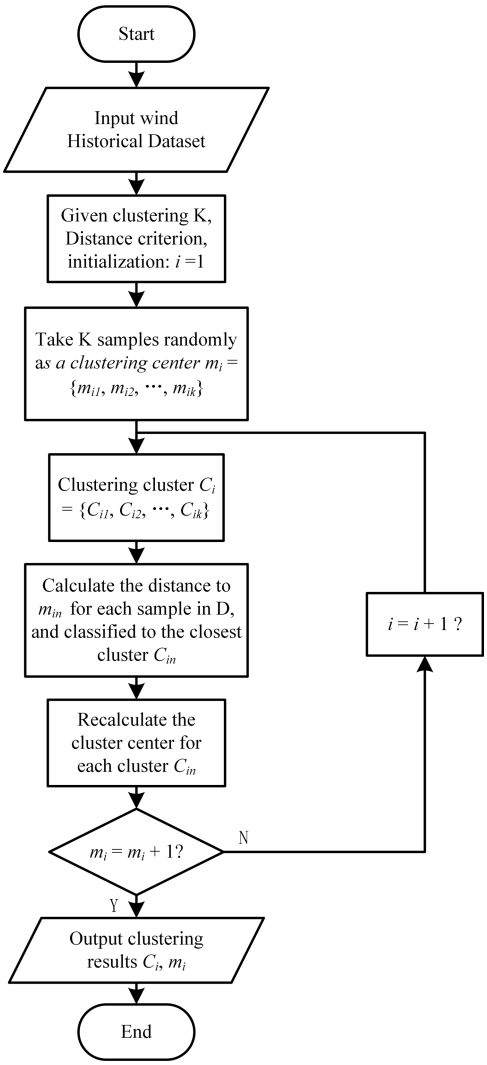


FIGURE 6. K-Means clustering process of impact factors and wind power.

a) Abnormal data processing. For a set of wind power historical data, DBSCAN and 3σ technology are used to deal with the abnormal values.

b) Division and acquisition of data sets. In order to meet the needs of experimental methods, the training dataset, the validation dataset, and the test dataset are extracted according to the data division method in section II.D.

c) K-Means clustering based on the impact factors. Firstly, the normalization method is applied to eliminate the numerical differences of the dataset. And then, K-Means clustering of the impact factors is realized to form K clustering categories.

d) LSTM network modeling. The LSTM network model is established based on each clustering category. The normalized impact factor and the wind power are used as the input and the output of the prediction model, respectively. In LSTM network training, the Adam algorithm [32] is used to optimize the loss function, and the Dropout technology [33] is used to prevent model over-fitting.

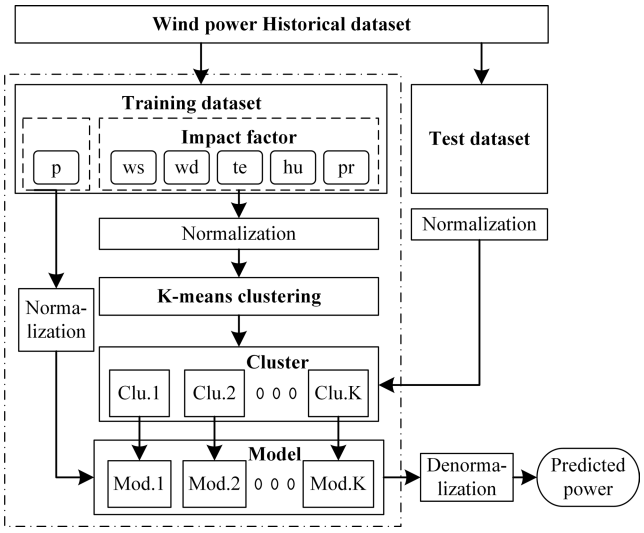


FIGURE 7. Prediction process of K-Means-LSTM combined forecasting model.

e) Test experiments. Normalize the impact factors in the test dataset as the input of the trained prediction model to realize the prediction of the related spot power, and combine the prediction error evaluation criteria to evaluate the pros and cons of the model. The errors are summarized as the error datasets for probabilistic interval prediction.

D. PREDICTION ERROR EVALUATION CRITERIA

In order to more quantitatively quantify the prediction error size numerically, several popular error evaluation criteria are used, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The first two are dimensioned in kilowatts (kW), the latter being dimensionless, expressed as a numerical percentage. The specific error calculation formulas are expressed as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_t - y_t\right| \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_t - y_t\right)^2} \tag{12}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100\% \tag{13}$$

where, n represents the number of test samples; $\hat{y}_t$ and y_t represent the predicted power and real power at time t, respectively.

IV. PROBABILISTIC INTERVAL PREDICTION

The distribution of spot power is easily affected by objective conditions such as geographic information and the meteorological factors of wind farms. However, the distribution of power prediction errors is relatively less affected by these objective conditions. Unlike parametric methods,

the nonparametric KDE [34] does not need to make certain assumptions about the distribution of the dataset. It depends on the kernel function and the bandwidth to achieve the fitting of the probability density, which is not only universal but also makes the fitted PDF closer to the real information.

KDE is a typical nonparametric method. It does not need to make assumptions about the PDF of the data in advance as prior information. It relies on the characteristics of data to obtain its distribution. Compared with the parametric method, it has more accurate expression ability and wider application occasions. Therefore, this method is applied to establish the probability distribution model of wind power prediction error.

A. NONPARAMETRIC KERNEL DENSITY ESTIMATION

For a group of wind power prediction error data $p = \{p_1, p_2, \cdots p_n\}$, n is the number of power error samples. Based on the principle of the nonparametric KDE, the PDF of wind power prediction error is estimated as (14).

$$\hat{f}(p,h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{p-p_i}{h}\right) \tag{14}$$

where, $\hat{f}(p,h)$ is the KDE of the prediction errors; h is the bandwidth, which determines the interval division of the error data distribution; p_i is the i-th prediction error sample point; function $K(p,h)$ denotes the kernel function used, and the power and the bandwidth are independent variables of the function.

The key factors affecting the nonparametric KDE are the selection of the kernel function and the selection of the bandwidth h. The selectivity of kernels is abundant. The commonly used kernels include uniform kernels, gamma kernels, Epanechnikov kernels, etc. Compared with the bandwidth, the selection of the kernel function has a less significant impact on the KDE. Gaussian kernel is suitable for many application scenarios for its smooth characteristics, which is usually considered to be the first choice. In this paper, the Gaussian kernel is applied, and it can be expressed as (15).

$$K(p) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-p^2}{2}\right) \tag{15}$$

Substitute (15) into (14), the nonparametric KDE is thus obtained as (16).

$$\hat{f}(p,h) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n}\exp\left(-\frac{1}{2}\left(\frac{p-p_i}{h}\right)^2\right) \tag{16}$$

B. BANDWIDTH OPTIMIZATION MODEL

In the nonparametric KDE, the PDF of the power error is strictly affected by the bandwidth. If the bandwidth is too small, the local fluctuation will be severe, and it will be affected by some special sample points, which affects the distribution of the estimated model. If the bandwidth is too large, the PDF will be too smooth to reasonably fit the distribution. As a brief summary, the appropriate selection of the bandwidth has a crucial impact on the nonparametric KDE.

At present, the bandwidth h is mainly based on the mean squared error (MSE) criterion, the integrated squared error (ISE) criterion, and the mean integrated squared error (MISE) criterion. However, the first two are susceptible to some points and sample properties, and MISE is more global than the other two. Therefore, in order to prevent the extreme phenomenon caused by the inappropriate selection of the bandwidth h, the MISE criterion can be used to measure the suitability of the bandwidth when selecting the bandwidth.

$$\begin{aligned} MISE(h) &= E\int\left(\hat{f}_h(x) - f(x)\right)^2 dx \\ &= \int\left(E\hat{f}(x) - f(x)\right)^2 dx + \int \operatorname{var}\hat{f}(x)\,dx \end{aligned} \tag{17}$$

The estimation of the kernel density can be composed of deviations and variances. After Taylor expands, MISE can be obtained as (18).

$$MISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^4 l^2(K)\,R(f'') + o\left(h^4 + \frac{1}{nh}\right) \tag{18}$$

where, $R(g)$ and $l(K)$ satisfy the following relationship:

$$\begin{cases} R(g) = \int g^2(u)\,du \\ l(K) = \int x^2 K(x)\,dx \end{cases} \tag{19}$$

After discarding the infinitesimal quantity of (19), the bandwidth optimization equation based on asymptotic integral mean square error (AMISE) can be obtained as (20).

$$A(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^4 l^2(K)\,R\left(f''\right) \tag{20}$$

According to (20), the partial derivative is obtained. When $\partial A(h)/\partial h = 0$, $A(h)$ converges and there is an optimal bandwidth h_{AMISE}.

$$h_{AMISE} = \left(\frac{R(K)}{nl^2(K)R(f'')}\right)^{\frac{1}{5}} \tag{21}$$

In (21), there exists an unknown function term $R(f'')$, containing the second-order partial derivative of KDE. For this reason, Silverman [35] pointed out that when f came from a normal population with mean $\mu = 0$ and variance σ, $R(f'')$ could be obtained as (22).

$$R(f'') = \frac{3}{8\sqrt{\pi}}\sigma^{-5} \tag{22}$$

C. INTERVAL PREDICTION PROCESS

The probabilistic interval prediction of wind power based on the nonparametric KDE can be divided into three steps: a) acquiring wind power error datasets; b) using the nonparametric KDE method to obtain the quantile at the confidence of (1-α); c) obtaining the wind power interval prediction values at different times, which is based on the spot prediction.

The interval prediction process of wind power is shown in Fig. 8. In order to obtain a reasonable dataset of wind power prediction errors, based on the spot prediction process shown in Fig. 7, the error generation dataset is replaced by

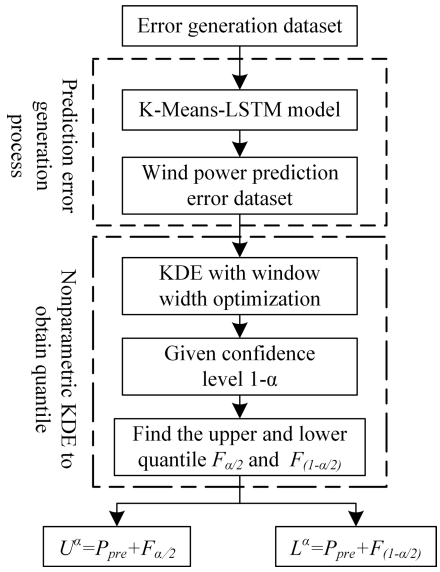


FIGURE 8. Wind power interval prediction process based on the power error dataset.

the test dataset to generate a set of wind power prediction error dataset. In addition, KDE with the optimal bandwidth is obtained by combining the bandwidth optimization described in Section IV.B. Given the confidence level of (1-α), the upper and lower quantile spot $F_{\alpha/2}$ and $F_{(1-\alpha/2)}$ are obtained. Finally, the wind power interval $\tau = [U^{\alpha}, L^{\alpha}]$ at the confidence level of (1-α) is obtained. The expressions of the upper and lower bounds are shown as (23).

$$\begin{cases} U^{\alpha} = P_{pre} + F_{\alpha} \\ L^{\alpha} = P_{pre} + F_{(1-\alpha/2)} \end{cases} \tag{23}$$

where, U^{α} and L^{α} are the upper and lower boundaries of the predicted power interval at the confidence level of (1-α), respectively, and P_{pre} is the spot prediction value of wind power.

D. PREDICTION EFFECT EVALUATION

The interval coverage rate and the sensibility index are the popular performance evaluation index [36], which are currently used to measure the prediction effect of the nonparametric KDE method.

The interval coverage rate R_{cover} can be used to measure the proportion of the real power falling within the prediction interval, and be defined as (24).

$$R_{cover} = \frac{1}{N}\sum_{i=1}^{N}\xi_i, \quad \xi_i = \begin{cases} 0 & P_i^{real} \notin \left[L_i^{\alpha}, U_i^{\alpha}\right] \\ 1 & P_i^{real} \in \left[L_i^{\alpha}, U_i^{\alpha}\right] \end{cases} \tag{24}$$

where, P_i^{real} is the real power from the i-th predicted sample; U_i^{α} and L_i^{α} are the upper and lower boundaries of the predicted interval of the i-th predicted sample at the confidence level of (1-α), respectively; ξ_i is an intermediate variable used to determine whether the real spot power falls within the range of the predicted interval; and N is the number of the prediction samples.

The sensibility index is used to measure the average width of the prediction intervals. The larger the interval coverage rate, the smaller the average width, which is the best performance of the prediction model. The sensitivity index ζ_{mean}^{α} is defined as (25).

$$\zeta_{mean}^{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\left(U_i^{\alpha} - L_i^{\alpha}\right) \tag{25}$$

V. CASE STUDY AND DISCUSSION

The wind power historical data used in this paper is derived from two wind turbines, which are respectively from two wind farms in the northeast of China. For convenience, they are respectively recorded as wind farms A and B. The rated capacity of a wind turbine is 1.5 MW, and the time span of the dataset is From Dec. 2017 to Dec. 2018, a total of 12 months of the dataset. The sampling time-frequency of each piece of data is 10 minutes.

A. PREDICTION EFFECT OF DATA PROCESSING

Nowadays, SVR, BP, and Elman neural networks are widely applied in wind power spot prediction. SVR is a prediction algorithm based on the idea of structural risk minimization, which has strong robustness. BP neural network is a classic neural network. Although it is easy to fall into local optimum during network training, the learning rate is fast and widely used in many research areas. The structure of the Elman neural network is similar to that of the BP neural network. However, the output of the hidden layer is connected to its input layer, which increases the steps of data delay and storage, and plays a certain role in the global optimization of the network. In this paper, wind power prediction models are established for these kinds of prediction algorithms and compared with the proposed methods.

In order to explore the effectiveness of DBSCAN and 3σ technology processing, this paper respectively takes 1000 samples of the same time period as the test dataset, which is from before and after the abnormal data processing. The prediction results of the four models are provided in Table 1. After the abnormal data processing, the MAE and the RMSE are greatly reduced, and the MAPE is reduced by nearly 12%-16%. It shows that the prediction effects of the four prediction models have been greatly improved. In addition, compared with different prediction errors, LSTM has smaller prediction errors than Elman, BP, and SVR. Here, the experiment illustrates the importance of abnormal data processing and also verifies that the DBSCAN and 3σ technology processing methods proposed in this paper are effective and feasible.

In addition, Fig. 9 shows the wind power fluctuations under 4 typical days of 4 seasons in wind farm A, and the prediction errors based on 1000 test data for different seasons are provided in Table 2. As shown in Fig. 9 and Table 2, the wind power in spring and autumn is larger than that in summer, and

TABLE 1. Prediction effect before and after of abnormal data processing within wind farms A and B datasets.

Model		Before abnormal data processing			After abnormal data processing		
		MAE	MAPE	RMSE	MAE	MAPE	RMSE
Wind farm A	LSTM	130.689	37.567	144.853	81.157	23.174	93.962
	SVR	143.153	41.635	158.864	95.034	27.846	110.855
	Elman	197.306	46.189	202.184	118.315	34.234	132.867
	BP	189.575	45.164	210.705	142.003	33.019	160.984
Wind farm B	LSTM	53.946	35.943	69.295	44.653	20.168	62.533
	SVR	73.534	42.892	90.698	54.853	30.275	65.894
	Elman	80.347	44.739	98.326	55.638	30.893	68.152
	BP	77.584	45.326	93.263	57.931	33.914	70.627

TABLE 2. Prediction error of different seasons based on wind farms A and B datasets.

Dataset	Season	Model	Before abnormal data processing			After abnormal data processing		
			MAE	MAPE	RMSE	MAE	MAPE	RMSE
Wind farm A	Spring	LSTM	124.407	35.689	137.85	77.349	22.015	89.514
	Summer	LSTM	116.563	33.435	129.169	72.480	20.625	83.876
	Autumn	LSTM	121.791	34.937	134.963	75.726	21.552	87.635
	Winter	LSTM	134.860	38.694	149.449	83.882	23.869	96.931
Wind farm B	Spring	LSTM	51.749	34.146	66.330	42.920	19.160	59.906
	Summer	LSTM	48.512	31.989	62.173	40.241	17.950	56.154
	Autumn	LSTM	50.670	33.427	64.944	42.127	18.756	58.656
	Winter	LSTM	55.964	37.021	71.874	46.493	20.773	64.909

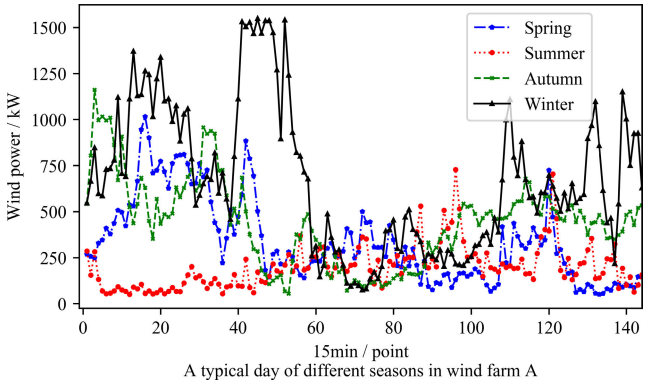


FIGURE 9. PDF of predicted error within wind farm A.

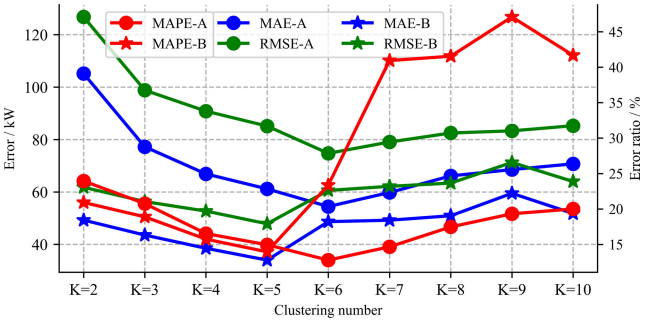


FIGURE 10. Predicted errors of different K values within wind farms A and B datasets.

the prediction error in spring and autumn is a little higher than that in summer. The wind power in winter is generally the largest, and the prediction error is also the highest among the four seasons. However, all the results indicate that abnormal data processing has a significant influence on the prediction results, and the prediction error is less affected by seasons.

B. EFFECTS OF SPOT PREDICTION

The combined prediction model of K-Means and LSTM proposed in this paper needs to determine the optimal number of clusters K. However, the unreasonable selection of the K value may cause inconspicuous or deteriorated results, and it is generally possible to find it through multiple value tests. In this paper, 2200 test samples are selected from wind farms A and B, respectively, in which, 2000 test samples are used to find the best value of K, and the other 200 test samples are used for the effectiveness test of the K-Means-LSTM model. The K value of the cluster number is taken from 2 to 10, and a total of 9 natural numbers are considered. The results of the three types of error evaluation of the model are shown in Table 3, and the related fluctuation trend is shown in Fig. 10. Different K values have a great influence on the prediction error. Whether it is MAE, MAPE or RMSE, the prediction effect is not satisfied when the K value is small, which is because the classification is not refined. When the K value is large enough, the error within wind farm A shows an increase, while the error within wind farm B appears to

TABLE 3. Wind power predicted error of different K-value prediction models based on wind farms A and B datasets.

Cluster		K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Wind farm A	MAE	105.154	77.227	66.887	61.176	54.435	59.791	66.119	68.541	70.725
	MAPE	23.945	20.755	16.531	14.950	12.783	14.661	17.471	19.323	20.003
	RMSE	126.752	98.814	90.825	85.127	74.732	79.044	82.488	83.309	85.266
Wind farm B	MAE	49.262	43.555	38.524	34.002	48.667	49.226	50.920	59.475	51.786
	MAPE	20.920	18.891	15.745	13.952	23.361	40.941	41.553	47.095	41.685
	RMSE	61.781	56.334	52.698	47.872	60.579	62.150	63.391	71.255	63.988

TABLE 4. Prediction error of each model based on wind farms A and B datasets.

Model	Wind farm A					Wind farm B				
	BP	Elman	SVR	LSTM	K=6_LSTM	BP	Elman	SVR	LSTM	K=5_LSTM
MAE	144.335	120.317	96.026	82.065	53.182	58.446	56.300	55.480	49.559	33.445
MAPE	33.441	34.516	28.101	23.413	12.490	34.215	31.261	30.627	20.579	14.117
RMSE	164.042	134.583	111.992	95.180	73.017	71.255	68.861	66.727	63.791	47.088

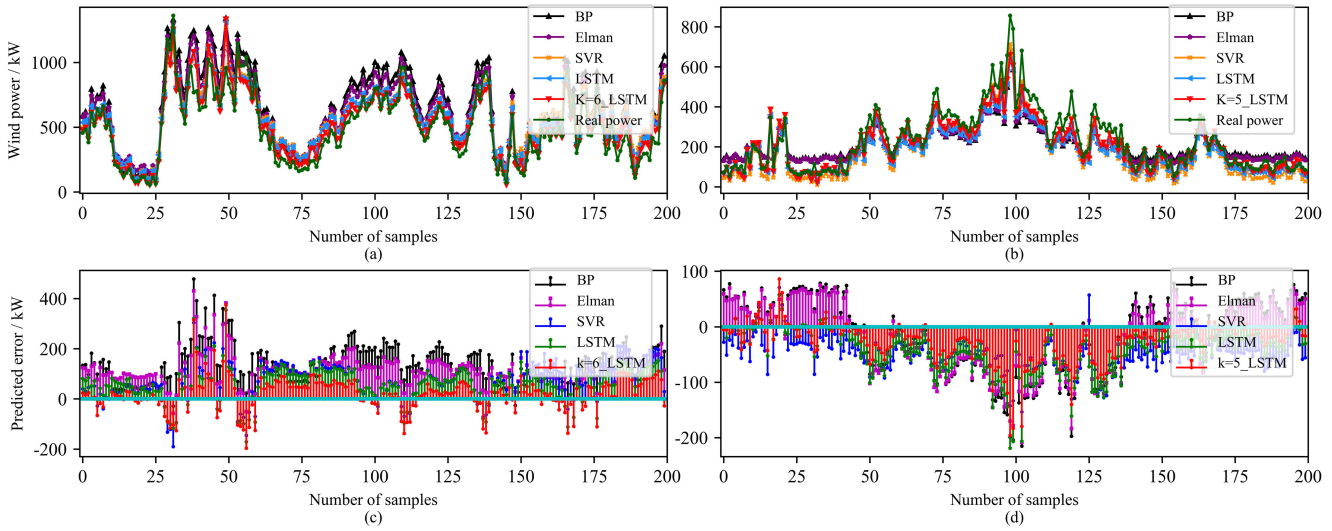


FIGURE 11. Spot prediction results for 200 samples within wind farms A and B datasets. (a) Wind power spot prediction in wind farm A dataset. (b) Wind power spot prediction in wind farm B dataset. (c) Wind power spot prediction error in wind farm A dataset. (d) Wind power spot prediction error in wind farm B dataset.

be close to a certain large error value. The key reason is that the number of samples in the same category set is too sparse when the classification is too much, and the sample data with more classification ambiguity is easily generated at the classification boundary. For wind farm A, there is a minimum error value when $K = 6$, and $K = 6$ is taken as the optimal cluster number of the related prediction model within wind farm A. Similarly, $K = 5$ is the optimal number of clusters for wind farm B.

In order to evaluate the advantages and disadvantages of the K-Means-LSTM network model, this section compares the prediction effects of the K-Means-LSTM network model, the LSTM network model, and the other prediction models mentioned before.

The prediction results of each model under another 200 test samples based on wind farms A and B are shown in Table 4. Fig. 11 shows the real-time prediction values and errors of the related prediction models.

It can be seen from Table 4 and Fig. 11 that several types of prediction errors of each model within wind farm B data are smaller than those of A. The main factor is that the numerical fluctuation range of wind power within wind farm B is relatively small and gentle. Among them, the prediction error of SVR is larger, and the prediction effect is second to that of the LSTM network. The prediction effect of the Elman network is similar to that of the BP network, which is worse than SVR and LSTM. However, the prediction results of both types of networks are easy to fall into local optimum. After the reasonable search of K value, the optimal K-Means-LSTM prediction model is obtained. The prediction effect of this model is the closest to the real power curve. Therefore, the validity of the combined prediction method proposed

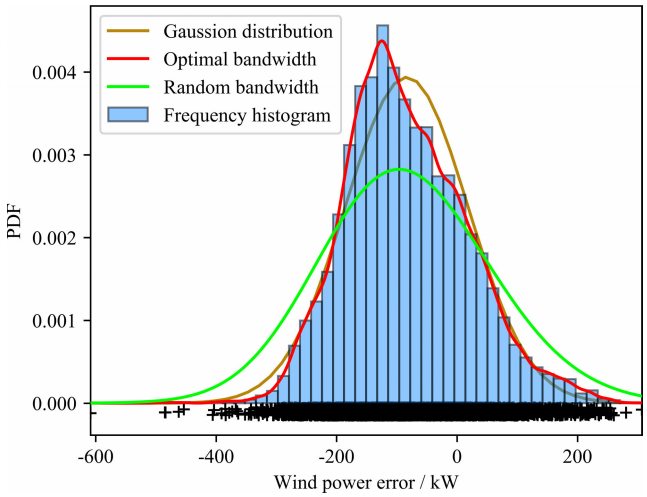


FIGURE 12. PDF of predicted error within wind farm A.

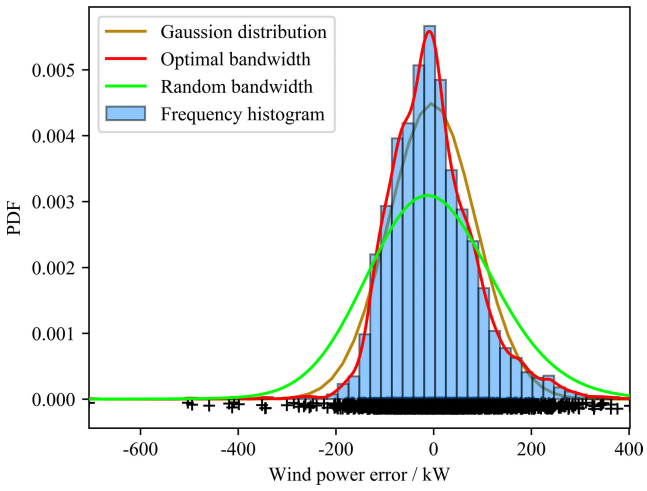


FIGURE 13. PDF of predicted error within wind farm B.

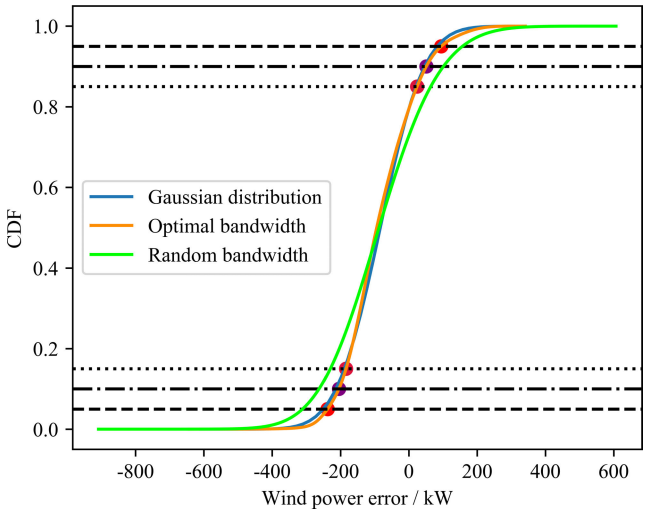


FIGURE 14. CDF of predicted error within wind farm A.

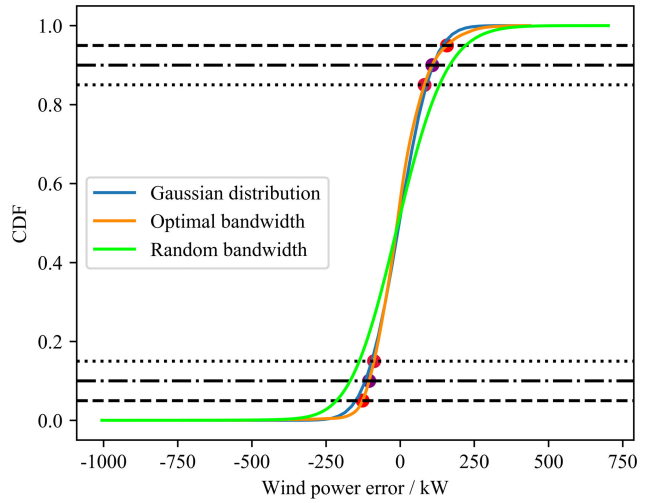


FIGURE 15. CDF of predicted error within wind farm B.

in this paper is verified. The main advantage of adopting the LSTM network model is that the network model can link the data of the previous time and the present time, so that the fluctuation trend of the data is relatively flat. In this paper, the Adam optimization algorithm and the dropout technology are adopted for the LSTM network. The former guarantees the efficiency and stability of the LSTM network model training, and the latter can well solve the over-fitting problem of the LSTM network model, so that the predicted power fluctuation range is closer to the real power fluctuation.

C. PROBABILISTIC INTERVAL PREDICTION EFFECT

Fig. 12 and Fig. 13 show the effect of three models within wind farm A and B datasets which are the nonparametric KDE model with the optimal bandwidth, the random bandwidth, and the Gaussian distribution model, respectively.

It can be seen from Fig. 12 and Fig. 13 that the nonparametric KDE method with the optimal bandwidth has the best fitting effect and is the closest to the distribution of the real power error dataset. The related cumulative distribution function (CDF) is obtained for each PDF fitted in Fig. 14 and Fig. 15. Based on the definition of the quantile, the upper and lower quantiles within different confidence levels are shown in Fig. 14 and Fig. 15.

As can be seen from Fig. 14 and Fig. 15, the CDF obtained is largely different due to the inconsistency in the PDF.

The quantile points at $\alpha/2 = 5\%$, $\alpha/2 = 10\%$, and $\alpha/2 = 15\%$ are respectively taken for each CDF to obtain the prediction power fluctuation range at the confidence level of 90%, 80%, and 70%. Combined with the K-Means-LSTM spot prediction model in Section III, Fig. 16, Fig. 17, and Fig. 18 show the fluctuation effect of the wind power prediction interval within wind farm A. And the prediction results of the wind farm B are shown in Fig. 19, Fig. 20, and Fig. 21.

In order to quantitatively describe the application effects of the PDFs of different wind power predicted error, based on the wind power fluctuation range, combined with the

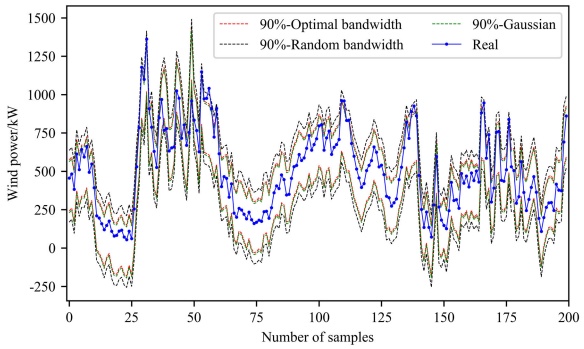


FIGURE 16. Interval prediction within wind farm A of 90% confidence.

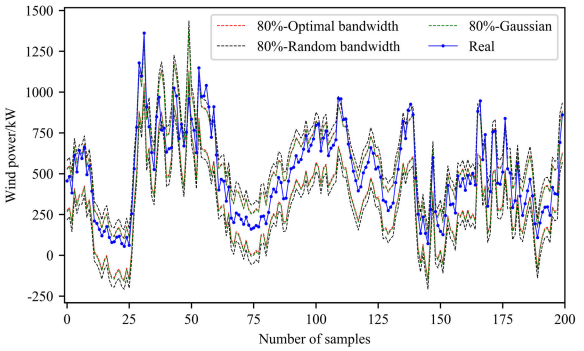


FIGURE 17. Interval prediction within wind farm A of 80% confidence.

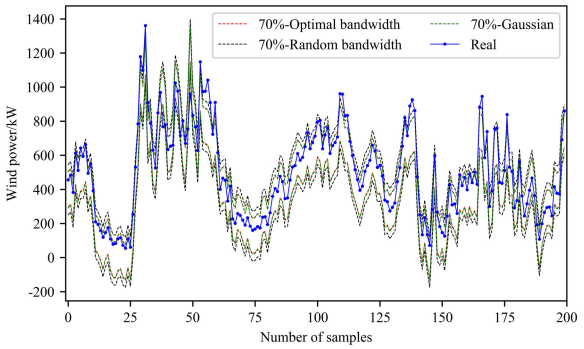


FIGURE 18. Interval prediction within wind farm A of 70% confidence.

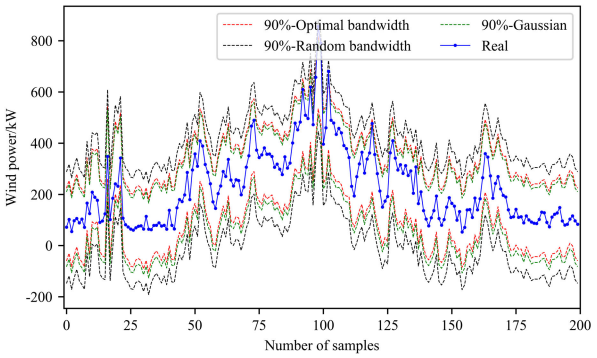


FIGURE 19. Interval prediction within wind farm B of 90% confidence.

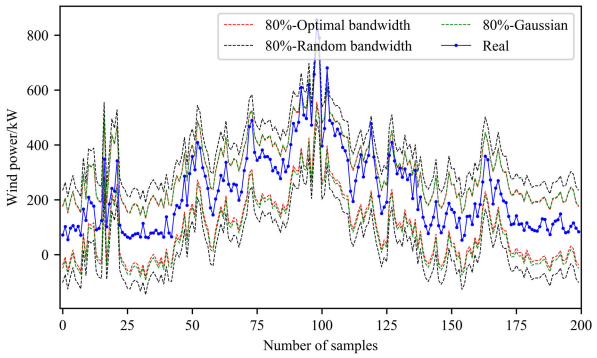


FIGURE 20. Interval prediction within wind farm B of 80% confidence.

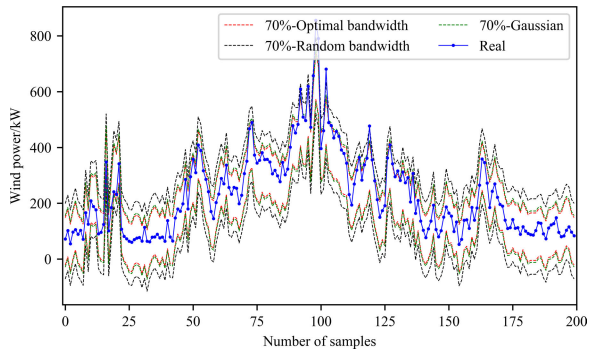


FIGURE 21. Interval prediction within wind farm B of 70% confidence.

evaluation criteria of Section IV.D, the evaluation results of the interval prediction at the confidence of 90%, 80%, and 70% are calculated, shown in Table 5.

As can be seen from Table 5, as the confidence level decreases, the width of the prediction interval decreases while the interval coverage decreases. The nonparametric KDE with the random bandwidth model has very high coverage at the confidence levels of 90%, 80%, and 70%. However, the ζ^{α}_{mean} evaluation index is very large. Compared with the other two methods, the information provided by this method does not have high utilization value. Within the wind farm A dataset, the ζ^{α}_{mean} evaluation index in the nonparametric KDE with the optimal bandwidth model at the different confidence levels is smaller than that of the Gaussian distribution model, and R_{cover} evaluation index is higher than that of the Gaussian distribution model. Within wind farm B data, ζ^{α}_{mean} of the optimal bandwidth model at the different confidence levels is significantly smaller than that of the Gaussian distribution model, which is beneficial to the auxiliary information on the grid side. Only when R_{cover} is slightly lower than the Gaussian distribution model at the confidence level of 70%, and the occurrence of such incidents is allowed.

It is because the upper and lower boundaries of the fluctuation intervals obtained by different models are different. Thus, their interval coverage rates also have some differences. In general, the interval prediction performance of the optimal bandwidth model is higher than that of the

TABLE 5. Effect evaluation of interval prediction.

Confidence	Method	Wind farm A		Wind farm B	
		R_{cover}	ζ^{α}_{mean}	R_{cover}	ζ^{α}_{mean}
90%	Optimal bandwidth	91.5%	332.854	98.5%	284.123
	Gaussian	89.5%	333.290	98.5%	292.377
	Random bandwidth	98.5%	467.963	100.0%	435.905
80%	Optimal bandwidth	85.5%	255.895	97.0%	212.546
	Gaussian	84.5%	259.676	97.0%	227.799
	Random bandwidth	91.5%	363.914	98.5%	336.134
70%	Optimal bandwidth	80.0%	207.978	93.5%	169.928
	Gaussian	79.5%	210.009	94.5%	184.229
	Random bandwidth	88.0%	294.089	98.0%	270.411

Gaussian distribution model, which can provide a narrower prediction interval and higher interval coverage at the same confidence levels. In addition, the information obtained by the random bandwidth model does not have a reliable and beneficial auxiliary effect.

VI. CONCLUSION

The method of the wind power prediction within the wind power historical data in the way of constructing supervised learning has serious delays, and the delay effect differs by a single unit time. Moreover, wind power prediction not combined with NWP is inadequate for the strong randomness of wind power. In order to avoid the above problems, this paper introduces the historical datasets of wind power with meteorological data as the research object, which comes from two wind farms in the Northeast of China. The meteorological data is regarded as NWP data in the micro-scale, and the wind power prediction is realized by combining LSTM and K-Means clustering algorithm and the nonparametric KDE method with bandwidth optimization. The main work of this paper can be summarized as follows:

a) This paper proposed an abnormal data processing method based on DBSCAN and 3σ technology for wind power historical data. The experimental results show that the method is effective and feasible, and the prediction accuracy of wind power has been greatly improved after the abnormal data processing.

b) Based on the varieties between the various factors of wind power, the K-Means clustering method is used to generate different clusters, and the wind power prediction is realized by the combination of each cluster forming an LSTM network model alone. Compared with BP, Elman neural network, SVR, and LSTM network prediction models, the K-Means-LSTM network model has higher accuracy for wind power spot prediction.

c) In order to obtain the wind power prediction error for the nonparametric KDE based interval prediction, a certain proportion of data has been divided from the wind power historical dataset as the error generation dataset by means of data set division. This method makes the power error fluctuation characteristics match the prediction preferences of the model, and thus generates a reasonable wind power prediction error dataset. A nonparametric KDE model with optimal bandwidth is proposed to realize the probabilistic interval prediction of wind power. Compared with the KDE model with the random bandwidth and Gaussian distribution model, the predicted interval width of the best bandwidth model is not only smaller, but also has a higher coverage rate, which can provide more reliable wind power auxiliary information.

ACKNOWLEDGMENT

The authors are grateful to Mr. Weiming Lang of Guodian Northeast Electric Power Co., Ltd. for providing the wind power historical datasets.

REFERENCES

[1] S. A. Vargas, G. R. T. Esteves, P. M. Maçaira, B. Q. Bastos, F. L. C. Oliveira, and R. C. Souza, "Wind power generation: A review and a research agenda," *J. Cleaner Prod.*, vol. 218, pp. 850–870, May 2019.

[2] L. Yang, M. He, J. Zhang, and V. Vittal, "Support-vector-machine-enhanced Markov model for short-term wind power forecast," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 791–799, Jul. 2015.

[3] Y. Zhang, B. Chen, Y. Zhao, and G. Pan, "Wind speed prediction of IPSO-BP neural network based on lorenz disturbance," *IEEE Access*, vol. 6, pp. 53168–53179, 2018.

[4] T. Mahmoud, Z. Y. Dong, and J. Ma, "An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine," *Renew. Energy*, vol. 126, pp. 254–269, Oct. 2018.

[5] M. Cui, J. Zhang, Q. Wang, V. Krishnan, and B.-M. Hodge, "A data-driven methodology for probabilistic wind power ramp forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1326–1338, Mar. 2019.

[6] M. Cui, V. Krishnan, B.-M. Hodge, and J. Zhang, "A copula-based conditional probabilistic forecast model for wind power ramps," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3870–3882, Jul. 2019.

[7] B. Zhou, D. Xu, C. Li, C. Y. Chung, Y. Cao, K. W. Chan, and Q. Wu, "Optimal scheduling of biogas-solar-wind renewable portfolio for multicarrier energy supplies," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6229–6239, Nov. 2018.

[8] C. Li, H. Shi, Y. Cao, Y. Kuang, Y. Zhang, D. Gao, and L. Sun, "Modeling and optimal operation of carbon capture from the air driven by intermittent and volatile wind power," *Energy*, vol. 87, pp. 201–211, Jul. 2015.

[9] W. Xie, P. Zhang, R. Chen, and Z. Zhou, "A nonparametric Bayesian framework for short-term wind power probabilistic forecast," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 371–379, Jan. 2019.

[10] P. Du, J. Z. Wang, W. D. Yang, and T. Niu, "A novel hybrid model for short-term wind power forecasting," *Appl. Soft. Comput.*, vol. 80, pp. 93–106, Jul. 2019.

[11] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis," *IEEE Trans. Ind. Appl.*, vol. 48, no. 4, pp. 1136–1141, Jul./Aug. 2012.

[12] Z. Wang, B. Wang, C. Liu, and W.-S. Wang, "Improved BP neural network algorithm to wind power forecast," *J. Eng.*, vol. 2017, no. 13, pp. 940–943, 2017.

[13] C.-M. Huang, C.-J. Kuo, and Y.-C. Huang, "Short-term wind power forecasting and uncertainty analysis using a hybrid intelligent method," *IET Renew. Power Gener.*, vol. 11, no. 5, pp. 678–687, 2017.

[14] M.-R. Chen, G.-Q. Zeng, K.-D. Lu, and J. Weng, "A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6997–7010, Aug. 2019, doi: 10.1109/JIOT.2019.2913176.

[15] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," *IET Renew. Power Gener.*, vol. 13, no. 10, pp. 1690–1700, 2019.

[16] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features," *Future Gener. Comput. Syst.*, vol. 93, pp. 33–42, Apr. 2019.
[17] S. Han, Y.-H. Qiao, J. Yan, Y.-Q. Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Appl. Energy*, vol. 239, pp. 181–191, Apr. 2019.
[18] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.
[19] Z. Gao, J. Geng, K. Zhang, Z. Dai, X. Bai, M. Peng, and Y. Wang, "Wind power dispatch supporting technologies and its implementation," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1684–1691, Sep. 2013.
[20] F. Ge, Y. Ju, Z. Qi, and Y. Lin, "Parameter estimation of a Gaussian mixture model for wind power forecast error by Riemann L-BFGS optimization," *IEEE Access*, vol. 6, pp. 38892–38899, 2018.
[21] X. Yuan, C. Chen, M. Jiang, and Y. Yuan, "Prediction interval of wind power using parameter optimized Beta distribution based LSTM model," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105550.
[22] J. Zhang, J. Yan, D. Infield, Y. Liu, and F.-S. Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Appl. Energy*, vol. 241, pp. 229–244, May 2019.
[23] M. Yang, S. Fan, and W.-J. Lee, "Probabilistic short-term wind power forecast using componential sparse Bayesian learning," *IEEE Trans. Ind. Appl.*, vol. 49, no. 6, pp. 2783–2792, Nov./Dec. 2013.
[24] A. U. Haque, M. H. Nehrir, and P. Mandal, "A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1663–1672, Jul. 2014.
[25] C. L. Gallego-Castillo, R. Bessa, L. Cavalcante, and O. Lopez-Garcia, "On-line quantile regression in the RKHS (Reproducing Kernel Hilbert Space) for operational probabilistic forecasting of wind power," *Energy*, vol. 113, pp. 355–365, Oct. 2016.
[26] J. W. Taylor and J. Jeon, "Forecasting wind power quantiles using conditional kernel estimation," *Renew. Energy*, vol. 80, pp. 370–379, Aug. 2015.
[27] Y. He and H. Li, "Probability density forecasting of wind power using quantile regression neural network and kernel density estimation," *Energy Convers. Manage.*, vol. 164, pp. 374–384, May 2018.
[28] J. Juban, L. Fugon, and G. Kariniotakis, "Uncertainty estimation of wind power forecasts: Comparison of probabilistic Modelling approaches," in *Proc. Eur. Wind Energy Conf. (EWEC)*, Brussels, Belgium, Mar. 2008, p. 10.
[29] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, pp. 28309–28318, 2019.
[30] H. Li, J. Wang, H. Lu, and Z. Guo, "Research and application of a combined model based on variable weight for short term wind speed forecasting," *Renew. Energy*, vol. 116, pp. 669–684, Feb. 2018.
[31] W. Wu and M. Peng, "A data mining approach combining K -means clustering with bagging neural network for short-term wind power forecasting," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 979–986, Aug. 2017.
[32] Z. Chang, Y. Zhang, and W. Chen, "Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform," *Energy*, vol. 187, Nov. 2019, Art. no. 115804.
[33] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018.
[34] X. Yang, X. Ma, N. Kang, and M. Maihemuti, "Probability interval prediction of wind power based on KDE method with rough sets and weighted Markov chain," *IEEE Access*, vol. 6, pp. 51556–51565, 2018.
[35] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.
[36] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 602–610, Jul. 2013.

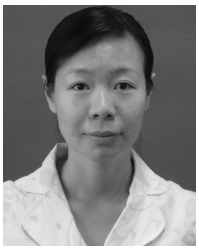
BOWEN ZHOU (S'12–M'16) received the B.Sc. and M.Sc. degrees from Wuhan University, Wuhan, China, in 2010 and 2012, respectively, and the Ph.D. degree from Queen's University Belfast, Belfast, U.K., in 2016, all in electrical engineering.

He joined the Institute of Electric Automation, College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2016, where he is currently working as a Lecturer. His research interests include power system operation, stability and control, electric vehicles, vehicle to grid, energy storage, demand response, renewable energy, and energy internet.

Dr. Zhou is currently a member of the IET, CSEE, and CCF. He is also a member of the Transformer Committee of the IEEE PES China. He is an Active Reviewer of some top journals, such as the *Applied Energy*, the *IET Generation, Transmission & Distribution*, and the IEEE TRANSACTIONS ON SMART GRID. He has served as the session chairs and TC/PC members for ten international conferences.

XIANGJIN MA was born in Lishui, Zhejiang, China, in 1995. He received the B.S. degree in electrical engineering and its automation from the Jiangxi University of Science and Technology, Ganzhou, China, in 2017.

He is currently pursuing the M.S. degree in electrical engineering from the College of Information Science and Engineering, Northeastern University, China. His research interests include renewable energy forecasting and control, machine learning, and data mining.



YANHONG LUO (M'09) received the B.S. degree in automation control, and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University, Shenyang, China, in 2003, 2006, and 2009, respectively.

She is currently with Northeastern University as an Associate Professor. Her research interests include power system optimization, energy internet, fuzzy control, neural networks adaptive control, approximate dynamic programming, and their industrial application.

DONGSHENG YANG (M'16–SM'19) received the B.S. degree in testing technology and instrumentation, the M.S. degree in power electronics and electric drives, and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 1999, 2004, and 2007, respectively.

He is currently a Professor with Northeastern University. He was supported by the Program for New Century Excellent Talents in University. He has authored or coauthored around 70 papers published in academic journals and conference proceedings, three monographs, and has co-invented 80 patents. His current research interests include distributed generation, multienergy power system, and artificial intelligence-based fault diagnosis and protection.