IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 31, 2018, accepted September 8, 2018, date of current version October 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870430

Probability Interval Prediction of Wind Power Based on KDE Method With Rough Sets and Weighted Markov Chain

XIYUN YANG, XUE MA, NING KANG, AND MIERZHATI MAIHEMUTI
School of Control and Computer Engineering, North China Electric Power University, Beijing102206, China

Corresponding author: Xue Ma (13051635800@163.com)

This work was supported by the China National Science Foundation under Grant 51677067.

ABSTRACT Research on the uncertainty of wind power has a significant influence on power system planning and decision-making. This paper proposes a novel method for wind power interval forecasting based on rough sets theory, weighted Markov chain, and kernel density estimation (KDE) method. Since the wind power prediction is significantly correlated to its historical record, this method first applies the Markov chain method to predict the power at different steps based on historical power data, and then the overall power is calculated via rough set weighted summation. Finally, the obtained forecasting power is fed into the KDE forecasting model to obtain both upper and lower bounds of the probability interval of the wind power at a certain confidence level. The predicted interval coverage probability and average bandwidth are two of the criterions used to evaluate the proposed method. Moreover, the simulation results obtained via the Markov chain-KDE method and the weighted Markov chain-KDE method are compared against the results of the proposed method. These comparisons show that the proposed method based on rough set theory and weighted Markov chain KDE method offers unique advantages over the other methods for probability interval prediction of wind power, which are higher coverage, narrower average bandwidth, and a more accurate result.

INDEX TERMS Wind power, interval prediction, rough set, weighted Markov chain, KDE.

I. INTRODUCTION

Wind energy, which is considered as green and renewable, has been widely used in many countries all over the world. However, the randomness and instability of wind introduces great challenges on the safety and stability of power grid [1]. Hence, the ability to effectively predict the wind power is very important [2]. Physical and statistical methods [3] have extensively been explored by worldwide scholars. However, all these methods almost get obtain the point predictions, which usually has contain large errors in an actual project due to various uncertainties. Unlike conventional wind power forecasting, which usually only produces a single value and the probability and the fluctuation range of the predicted value cannot be accurately obtained, wind power probabilistic forecasting can provide much information on uncertainty. The interval models contain more beneficial information and can provide an effective range in which wind power output lies with a specified probability and thus offer more information for decision makers. It is of great significance for decision makers to make better decisions to plan and operate a power grid safely.

In recent years, many researchers have focused on probabilistic forecasting [4]. A detailed review of the method can be found in [5]. At present, four classes of probabilistic forecasting algorithms have been described in the literature: a) error distribution analysis [6], the problem of modeling the distribution of the forecast errors is addresses in [6] and a mixed distribution-based model develops to approximate the distribution of these errors. b) conditional kernel density estimation (KDE) [7], [8]. A time-adaptive quantile-copula estimator is constructed for kernel density forecast and how to select the adequate kernels for modeling are disscussed in [8]. c) machine learning [9]–[12]. In [9], it constructs PIs of the ANFIS models using the delta technique and a cost function is developed to train the ANFIS. d) quantile regression (QR) [13], [14]. In [13], it is shown how to build a model of the quantiles regression to extend forecast systems with information on uncertainty. The method of error distribution analysis usually takes the form of a probabilistic density function. Parametric and non-parametric approaches are two main techniques to extract probabilistic density function of error distribution. The parametric method is based on the assumed shape of sample data [15], such as a Gaussian distribution, a beta distribution, etc. The parametric method is more dependent on the prior model and the deviation of its error distribution will have a great impact on prediction accuracy. However, sometimes it is not reasonable to assume a specific distribution of the forecast errors for wind power. Nonparametric methods do not need to make any assumptions about the sample data distribution and are robust [16]; examples are the quantile regression and the kernel density estimation (KDE) method. In [16], a hybrid wind power probability density prediction method is presented based on quantile regression neural network and kernel function. QR is robust to estimate the conditional distribution of the explained variables. Reference [17] designs a quantile regression neural network (QRNN) model, which combines the advantage of ANN and QR model. Reference [18] proposes a robust neural network method to depict the condition density at any future time based on QR and ANN model. The KDE method can directly provide the probability density function and becomes a popular probability prediction method. In [7], a conditional kernel density estimation has been used to calculate wind power density forecasts and shown that its approach makes no distributional assumption for wind power. In [19], the KDE method was applied to predict the wind power in a wind farm located in France. The results showed that the predicted probability density distribution obtained via the KDE method has an improved reliability and sharpness compared to the quantile regression method. Therefore, the non-parametric kernel density method was adopted in this paper to model the wind power probability. The probabilistic density function of point prediction error is constructed by non-parametric kernel density estimation and the cumulative power error distribution function is calculated to achieve the upper and lower bounds of interval prediction.

To extract the probabilistic density function of power error distribution, firstly, a wind power prediction with high accuracy is needed to be generated. Wind speed is the meteorological factor of most relevance to wind power generation. The wind speed time-series has high fluctuations in a wide range of frequencies, and wind power is also a nonlinear and non-stable statistics process and has large random fluctuations. The Markov chain has higher accuracy and is better suited for describing large random fluctuation problems compared to other methods. The Markov chain method can predict trends of data series by studying the different states of initial probability and transition probability [20], [21]. Therefore, we can adopt Markov chain theory to accomplish the point forecasting of wind power. The weighted Markov chain [22]–[24] is a modified Markov chain method with weight improvement to improve data mining of the initial data so as to obtain more accurate precision. One of the relevant issues is that weights have a decisive effect on the accuracy of weighted Markov chain prediction. Rough set [25] does not require extra information in addition to the data set and it is not subjective either, it only needs to evaluate the predicted value of the model to obtain the weight coefficients. Therefore, this paper proposes to use the rough set method to calculate weights for establishing a combined Markov chain model of wind power prediction to further improve prediction accuracy. Predicted wind powers with Markov chain model at different steps are defined as condition attributes of rough set. Rough set calculates the significance coefficient of each attribute, thus obtaining the combined weights of predicted wind powers at different steps and establishes a combined Markov chain model.

The main contribution of this study presents a wind power interval probability prediction model that uses the non-parametric kernel density with rough set theory and weighted Markov chain method. The weighted Markov chain is capable of effectively processing a variety of information in random wind power systems. Rough set theory can obtain the objective importance of each condition attribute without prior information. The rough set is used to determine the weights in weighted Markov chains, which can improve the accuracy of the power prediction. The KDE method does not make any assumptions about the data distribution and directly provides the probability density function based on data characteristics. The KDE method is then applied to predict the wind power probability interval based on power error data. We utilized the predictive evaluation indexes reported in reference [26], predicted interval coverage probability and interval average bandwidth, as predictive evaluation indexes to validate the effectiveness of proposed KDE with rough set-weighted Markov chain model. Comparison with the KDE method based on Markov chain and KDE method based on weighted Markov chain are also presented.

The remainder of this paper is organized as follows: The second section introduces the theories; the third part establishes the wind power interval prediction model using the KDE method with both the rough set and the weighted Markov chain; the fourth section presents a case study and the last section provides the conclusion.

II. METHODOLOGY

A. MARKOV CHAIN

Wind power data are non-linear, non-stationary time series data. Compared with other methods, the Markov chain is better suited for describing random fluctuation problems. Markov chain [20], [27] is a special stochastic process, exhibiting the Markov property, in which the conditional probability distribution of future states of the process (given the present state and all past states) depends only upon the present state and not on any past states. Discrete time and state Markov processes are known as Markov chains.

Assuming a random process $\{X_{\mathrm{n}}\}$ and a discrete state $S = \{s_0, s_1, \cdots, s_n \cdots\}$, if the state at (n+k) state $X_{n+k} = s_{n+k}$ of a random process X_n only relates to the n state$X_n = s_n$, but not to the previous state, then

$$P\{X_{n+k} = s_{n+k}|X_0 = s_0, X_1 = s_1, \cdots, X_n = s_n\} = P\{X_{n+k} = s_{n+k}|X_n = s_n\} \quad (1)$$

Then, $\{X_n\}$ is considered as a Markov chain.

The conditional probability is

$$P_{ij(k)} = P\{X_{n+k} = s_j|X_n = s_i\} \quad (2)$$

In which: $P_{ij(k)}$ represents the transition probability from state n+k to s_j in the initial time n and state s_i.

$s_0, s_1, \cdots, s_n \cdots$ are different states, which represent the data series, then the conditional transition probability is:

$$P_{ij(k)} = \frac{M_{ij(k)}}{M_i} \quad (3)$$

In which: $M_{ij(k)}$ represents the number of data transited from state s_i to s_j which has k steps; M_i represents the original number of data when the time series is at state s_i. The conditional transition probability matrix at the k step:

$$P^k = \begin{bmatrix} P_{11(k)} & P_{12(k)} & \ldots & P_{1n(k)} \\ P_{21(k)} & P_{22(k)} & \ldots & P_{2n(k)} \\ \ldots & \ldots & \ldots & \ldots \\ P_{n1(k)} & P_{n2(k)} & \ldots & P_{nn(k)} \end{bmatrix} \quad (4)$$

The Markov chain prediction can be determined by its transition probability matrix and its initial distribution vector. The state transition probability matrix describes the dynamic characteristics of the Markov forecasting model. Assuming the distribution vector of the initial state s_0 is P_0 and the state distribution after k steps:

$$P(k) = P_0 P^k \quad (5)$$

In the paper, Markov chain theory is utilized to to predict wind power state.

B. WEIGHTED MARKOV CHAIN

The weighted Markov chain is a modified Markov chain with weight improvement, which enables to explore the impact of the transition probability matrix of each step and data mining of the initial data.

The step of Weighted Markov chain method [20] is as follows:

(1) Establish classification standards and determine the state of each data based on the standards.

(2) Establish the Markov chain transition probability matrix at different steps.

(3) Predict the state probability based on the corresponding transition probability matrix, as well as using the previous data as initial state.

(4) Determine the weight of each step in the Markov chain mode.

(5) The weighted sum of each predicted probability for the same state is used as final probability, and the state corresponding to the maximum probability is the predicted state of the sample. The prediction state is then converted to the corresponding predicted value. Adding the predicted value to the original data series, step (1) to step (5) are repeated for the next round of prediction.

In order to make full use of historical information, this paper uses the weighted Markov chain for wind power prediction.

C. ROUGH SET THEORY

The weight coefficients can affect the accuracy of predicted state. A rough set can determine a reasonable weight, because it does not require any prior information and can analyze the data to obtain the objective importance of each condition attribute.

Rough set theory [28] is a theory for data analysis proposed by Z. Pawlak of the Polish Academy of Sciences, which offers unique advantages for the processing and summarization of incomplete and indefinite data since it does not require prior information besides the data set itself. It is capable of effectively processing a variety of data and information in complex systems and can also be applied for data analysis and statistical inference. It has been widely used in data mining, decision analysis, and pattern recognition.

In the rough set theory, the knowledge system can be described as:

$$S = \{U, A, V, f\} \quad (6)$$

In which: $U = \{x_1, x_2, \cdots, x_n\}$ represents the universe of discourse; $A = C \cup D$ represents a set of attributes, C represents a set of condition attributes, and D represents a decision attribute; $V = \cup V_\alpha$ represents a set of attribute values and V_α represents the range of the attribute values in $\alpha \in A$. $f : U \times A \to V$ represents the information system. A knowledge expression system with both conditional attributes and decision attributes is often referred to as a decision system. The decision system is represented by a decision table, in which each row records the elements of the domain and each column represent different attributes.

Let $C_1 \in C$ be a condition attribute of the decision system. In the rough set theory, the dependency degree of the condition attribute C_1 to the decision attribute D is defined as

$$\gamma_{C_1}(D) = \frac{POS_{C_1}(D)}{|U|} \quad (7)$$

Where, POS_{C_1} (D) represents the positive domain of decision attribute D of the knowledge C_1, which is a set of objects. This domain can be accurately divided into equivalence classes of the relational D according to the classified U/C_1 information; $|U|$ represents the number of elements in the domain. γ_{C_1} (D) represents the proportion of objects that can be accurately classified into the decision class U/C under the condition attribute C_1, and describes the level of support of the condition attribute for the decision attribute D.

For a decision-making system, every condition attribute has a different degree of dependency to the decision attribute D, and the degree of dependency of decision attributes by condition attributes is called the significance of the condition attributes. In the rough set theory, the significance of the condition attributes is measured via the change of classification ability in the decision-making system after removing the condition attributes.

C_1 represents a condition attribute in a decision system and sig(C1,C;D)represents the significance of C_1:

$$sig\,(C_1, C; D) = \gamma_C\,(D) - \gamma_{C-C_1}\,(D) \tag{8}$$

A larger sig(C1,C;D) indicates that condition attribute C_1 exerts more influence on the decision in condition attribute set C and therefore, it is more important. In contrast, condition attribute C_1 has a smaller impact on the decision and is inevitably less important. In order to improve the predicted accuracy, this paper uses rough set to determine the weight of the weighted Markov chain.

D. NONPARAMETRIC KDE THEORY

The nonparametric estimation method [29] does not require a priori definition of the model, which is a great advantage without knowing which probability density standard parameters will follow. The KDE is one of the non-parametric estimation methods, compared with the parametric method, which has higher reliability and can well describe the continuous density function. The method does not make any assumptions about the data distribution, but purely studies the data distribution based on its characteristics. Therefore, this paper uses a non-parametric KDE to establish the error distribution model of the predicted wind power.

The KDE has the following expression for data series $y_1, \ldots, y_n$.

$$\widehat{f}_h = \frac{1}{nh}\sum_{q=1}^{n} K\left[\frac{y - y_q}{h}\right] \tag{9}$$

Where the kernel function K(·) is a weight function whose shape and range control the number of points and utilization when it is used to estimate the value of f(y) at point y. The estimation of kernel density depends on the choice of the kernel function and on bandwidth h. However, the kernel function has a far more significant influence on the kernel estimation than bandwidth. Gauss kernel, Earl's kernel, homogeneous kernel, etc. are all commonly used kernel functions. The Gauss distribution $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^2)$ is chosen in this paper.

Asymptotic integral mean square error (AMISE) can be used to select the bandwidth, which is calculated as follows

$$A(h) = \frac{R(K)}{nh} + \frac{h^4\sigma_k^4 R(f)}{4} \tag{10}$$

Here, $R(K) = \int K^2(u)du$, $R(f) = \int (f(x))^2 dx$.

The bandwidth h_{AMISE} is described as follows, when $A(h)$ is minimized at $\partial A(h)/\partial h = 0$.

$$h_{AMISE} = \left(\frac{R(k)}{n\sigma_k^4 R(f'')}\right)^{\frac{1}{5}} \tag{11}$$

Here, σ_k represents the standard deviation. $R(f'')$ includes the second derivation of an unknown probability density function. Since f is unknown, $R(f'')$ is estimated first before h_{AMISE} is known. Silverman proposed an elementary approach to swap the normal densities that have a mean of 0 and a variance of the sample variance matching the variance and the estimated variance, respectively.

$$R(f'') = \frac{3}{8\pi^{-\frac{1}{2}}\sigma^5} \tag{12}$$

Here, σ represents the standard deviation of the normal distribution

The bandwidth h_{pilot} is described as follows when K is the standard normal kennel distribution:

$$h_{pilot} = (4\pi)^{-\frac{1}{10}}\left(\frac{3}{8}\pi^{-\frac{1}{2}}\right)^{-\frac{1}{5}}\sigma n^{-\frac{1}{5}} = \left(\frac{4}{3n}\right)^{\frac{1}{5}}\sigma \tag{13}$$

III. WIND POWER INTERVAL PREDICTION MODEL BASED ON THE KDE METHOD WITH ROUGH SET AND WEIGHTED MARKOV CHAIN

This paper proposes a rough set and weighted Markov chain KDE method for wind power interval prediction; the overall flow chart is shown in Fig. 1. After the wind power is predicted via the rough sets and the weighted Markov chain model, the wind power interval prediction model is obtained by utilizing the nonparametric density method. Specific technical details are described below.

A. POWER PREDICTION BASED ON WEIGHTED MARKOV CHAIN WITH ROUGH SET

The Markov chain model with a step size of 1 and the initial state vector are used to calculate the absolute distribution of the future time, which is called the absolute distribution Markov chain prediction method. The data sampling point interval is 10 min. This paper explores the influence of historical power at t-1, t-2, t-3, t-4, t-5, and t-6 on the power at time t to predict the power at time t. The weighted Markov chain method is applied to obtain the predicted values of the steps 1, 2, 3, 4, 5, and 6, respectively. Finally, the rough set is used to determine the weights at different steps to calculate the final predicted power value. Fig. 2 shows the flow diagram of this method.

Steps are as follow:

1) The maximum wind farm power selected in this simulation is 200 MW. According to the power distribution, the processed data is divided into 20 states, which form a state sequence. For example, the power in the [0, 10], [10, 20], and [190, 200] interval is denoted as 0, 1, and 19 respectively.

2) Select the model data and normalize.

3) The state sequence is statistically calculated and the transition probability matrices at the step of 1, 2, 3, 4, 5, and 6 are obtained, respectively.

4) The states corresponding to the power values of the previous intervals are taken as the initial state vectors, respectively, and the state probabilities under different step lengths can be predicted according to formula (5) in combination with their corresponding transition probability matrices.

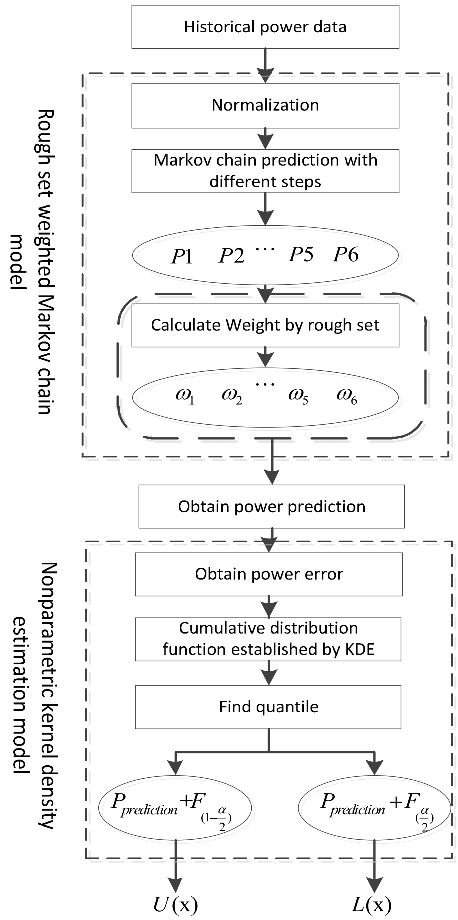

FIGURE 1. **Flowchart of wind power probability interval prediction for rough set, weighted Markov chain, and KDE method.**

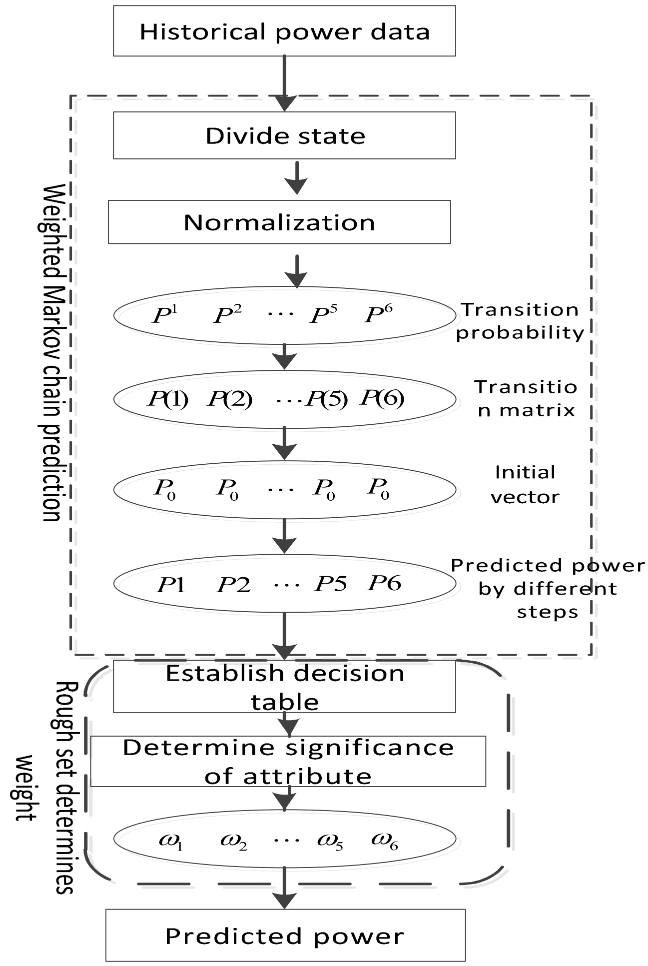

FIGURE 2. **Flowchart of the power prediction model based on weighted Markov chain and rough set.**

5) Predict the power at different steps. The state corresponding to the maximum transition probability is the state of the predicted value, which is then converted into the predicted value.

$$predicted\ vaule = \frac{upper\ value + lower\ value}{2} \tag{14}$$

For example, if the state is 0, the state value is set as 5; if the state is 1, the state value is 15, and the predicted value is added to the original sequence.

6) The decision table is established, and the predicted power obtained at steps 1, 2, 3, 4, 5, and 6 is taken as the conditional attribute, $C = \{C_1, C_2, \cdots, C_6\}$, $C_i\ (i = 1, 2, \cdots, 6)$ is the condition attribute; the real wind power is taken as the decision attribute D. Let x_t be the element in the domain $U, x_t = \{C_{1,t}, C_{2,t}, \cdots, C_{n,t}; D_t\}$. $C_{i,t}$, and D_t are the predicted value and real value of the i th prediction model at t, respectively. Then, the significance of every condition attributes is calculated by formula (8).

7) Final power prediction = predicted value at different step × corresponding significance of condition attribute.

B. WIND POWER PROBABILITY CONFIDENCE LEVEL PREDICTION BASED ON NONPARAMETRIC KERNEL DENSITY ESTIMATION

After the predicted wind power is achieved from section 3.1, the prediction error can be obtained. The non-parametric kernel density method is capable of obtaining the probability density function of the prediction error without requiring any distributional assumptions on the sample data. This paper uses it to establish the probability prediction interval of wind power.

The window width of the non-parametric kernel density method determines the accuracy and smoothness of this estimation. A large h may lead to high smoothness of the probability density function, which leads to a large estimation error; In contrast, a small h may lead to exceedingly high probability density function fluctuations (especially the tail of the probability density curve). In this paper, the mean square error of the asymptotic integral method is used to determine the window width of the non-parametric kernel density method.

The Gaussian kernel density function has excellent robustness, which is chosen to estimate the cumulative distribution function in this paper. The cumulative distribution function of P is the probability that the random distribution will take a value less than or equal to P, providing the cumulative probability of each value along the probability density function. For all real numbers v, the cumulative distribution function is defined as follows:

$$F(V) = P(V \leq \mathrm{v}) \tag{15}$$

Where, F(v) is the probability satisfying the condition. Taking the wind power error as variable V, if the real number v has a probability of 90%, thus satisfying the wind power, then equation (16) can be rewritten as:

$$F(v_{90\%}) = P(V \leq v_{90\%}) \tag{16}$$

Where the confidence level is $1-\alpha$, $\mathrm{F}\left(\frac{\alpha}{2}\right)$, $\mathrm{F}\left(1-\frac{\alpha}{2}\right)$ can be obtained according to equation (16). Therefore, the lower and upper bound of the wind power probability interval prediction can be approximately described as:

$$(P_{prediction} + F(\frac{\alpha}{2}), P_{prediction} + F(1-\frac{\alpha}{2})) \tag{17}$$

C. WIND POWER PROBABILITY INTERVAL PREDICTION STEPS

The specific steps of probability interval prediction (using non-parametric density estimation method based on rough set and weighted Markov chain) are as follows:

1) Remove dead pixels and other preprocessing for historical data.

2) According to the historical wind power, the state space is evenly divided to obtain the transition matrix from step one to step six and the Markov chain models with different steps are established. The weights of all six steps are determined to obtain the final power prediction value based on the rough set method.

3) The prediction error is calculated based on the predicted power using training data, and a non-parametric kernel density estimation model is established for the power error to obtain the cumulative power error distribution function.

4) Equation (15) can determine the cumulative distribution function at quantile $\alpha/2$ and $1-\alpha/2$, which is then converted into the upper and lower limits of the confidence interval of the predicted value according to equation (17);

5) The upper and lower limit of the confidence level at each point forms two envelops and obtain the probability interval prediction.

IV. CASE STUDY

A wind farm in the northwest of China is used as an example; the farm has a total installed capacity of 199.5 MW and a temporal resolution of 10 min. Wind power data collected from the site at the wind farm in 2014 are used to verify the feasibility of the proposed method. The KDE with Markov chain method and the KDE with the weighted Markov chain method are compared to verify the superiority of this method.

TABLE 1. Weights of combination model based on rough set.

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
0. 433	0.233	0.133	0.087	0.069	0.045

TABLE 2. RMSE of different forecasting models.

Model	Mean squared error (%)
Absolute distribution Markov chain	7
Weighted Markov chain	6.45
Rough set weighted Markov chain	4.93
BP neural network	7.35
Time series	7.77

A. WIND POWER PREDICTION BASED ON ROUGH SET AND WEIGHTED MARKOV CHAIN

1) WEIGHT SELECTION BASED ON ROUGH SET

Weighted Markov chain prediction at steps 1, 2, 3, 4, 5, and 6 is selected in this paper. The conditional attributes of the rough set are the six Markov predicted powers, and the decision attributes are the true power output values. Table 1 shows the significance of conditional attributes obtained via simulation, which are the weights of the weighted Markov chain.

2) WEIGHTED MARKOV CHAIN PREDICTION

To verify the effectiveness of rough set selection weights, it is compared against the absolute Markov chain power prediction simulation and weighted Markov chain power prediction, where the weight is based on parametric estimation. Among these, the weights ω_k determined by the parameter estimation method are obtained according to formulas (7) and (8):

$$r_k = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{18}$$

Where r_k is autocorrelation Function of each rank.

$$\omega_k = \frac{|r_k|}{\sum |r_k|} \tag{19}$$

The comparison results of the three prediction models are shown in Table 2. The prediction work based on BP neural network and time series method are also carried to compare the performance of rough set weighted Markov chain. Table 2 shows that the rough set-weighted Markov chain can fully consume the information of all aspects by using reasonable weights, which can effectively improve prediction accuracy.

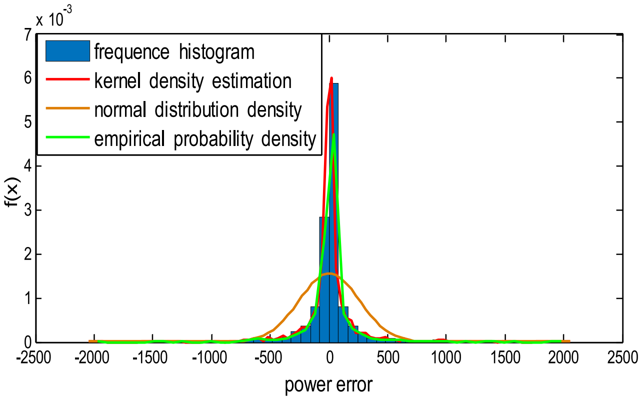

FIGURE 3. Power error frequency histogram and kernel density estimation.

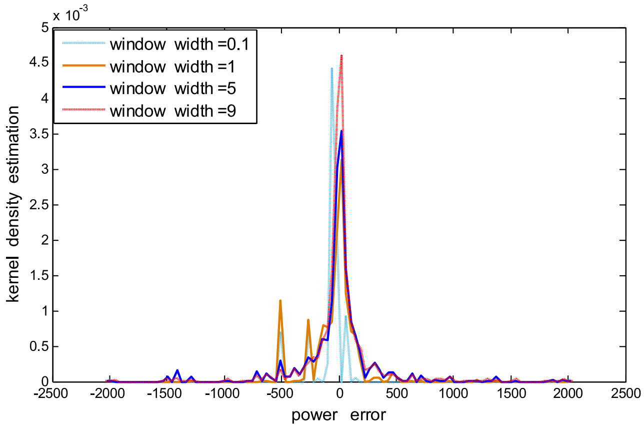

FIGURE 4. Nonparametric Kernel Density Estimation of power errors with different window sizes.

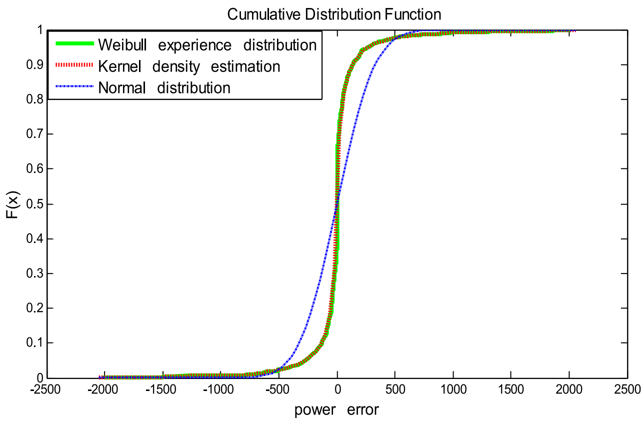

FIGURE 5. Wind power error cumulative distribution.

B. UPPER AND LOWER BOUND SELECTION OF POWER INTERVAL PREDICTION

Taking the 80% confidence level as an example, utilizing predicted power, the power error model is established by using the non-parametric kernel density method. The cumulative distribution function is calculated and F(10%) and F(90%) are substituted into formula (17) to obtain the upper and lower bounds of the probability interval prediction of wind power.

1) KERNEL ESTIMATION PARAMETERS SELECTION

Fig. 3 shows the statistical histogram and non-parametric kernel density estimation, normal distribution density fitting, and Weibull empirical distribution density after statistical analysis of the rough set and weighted Markov chain prediction results.

Fig. 3 shows that the probability density distribution of the KDE is closer to the actual distribution with good fitting accuracy. It offers good accuracy and smoothness without ignoring the global peak; it is also in line with the description of its basic characteristics of the distribution requirements.

The smoothness of estimation is influenced by the non-parametric KDE window width. The simulation of different window widths is conducted by using the method of asymptotic integral mean square error, as shown in Fig. 4. The best fitting result is obtained for a window width of 9, which is used as the kernel width.

Fig. 5 shows the distributions generated via the Weibull empirical distribution, kernel density estimation with Gaussian kernel function, and normal distribution in the MATLAB platform using window width of 9. The result shows that the kernel density estimation method is capable of producing a distribution function with good fitness, smoothness, and accuracy, effectively reflecting the distribution of power prediction errors.

C. INTERVAL PREDICTION RESULTS AND ANALYSIS

The power system always requires a higher confidence level to obtain more accurate information, to ensure the safety of the system in practical applications. Therefore, confidence levels of 80%, 85%, and 90% are selected in this study. According to the cumulative distribution function of power error provided in Fig. 5, interval prediction can be calculated using equation (17). Figs. 6, 7, and 8 show the wind power predicted interval of 500 points based KDE with 80%, 85%, and 90% confidence levels. In comparison with KDE, a prediction intervals by normal distribution is also carried, whose PI is obtained by equation (20) and (21).

$$L(x_i) = \hat{y}(x_i) - z_{\alpha/2}\sigma \tag{20}$$
$$U(x_i) = \hat{y}(x_i) + z_{1-\alpha/2}\sigma \tag{21}$$

Where $L(x_i)$ is lower bound of predition interval, $U(x_i)$ is upper bound of prediction interval, $\hat{y}(x_i)$ is prediction value by rough set weighted Markov chain, σ is the variance of a normal distribution, Z is coffeicent, which can be obtained by looking up standard normal distribution table at specified confidence level.

As seen in these figures, the proposed model can effectively track the change of wind power. The model can cover the actual value around the confidence level and the deviation between the actual value falling outside and the PI boundary are relatively small. The PIs by normal distribution present equal width for upper and lower bound and show the wider

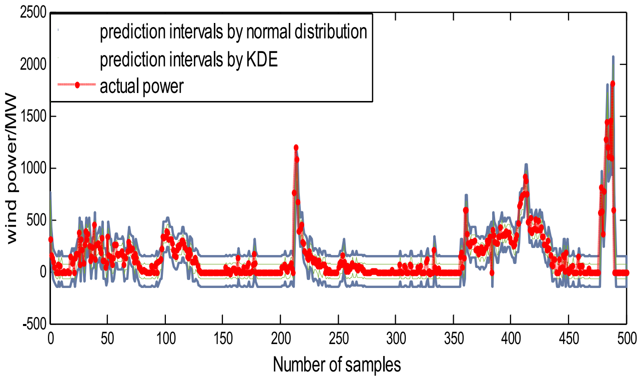

FIGURE 6. Prediction intervals under 80% confidence level.

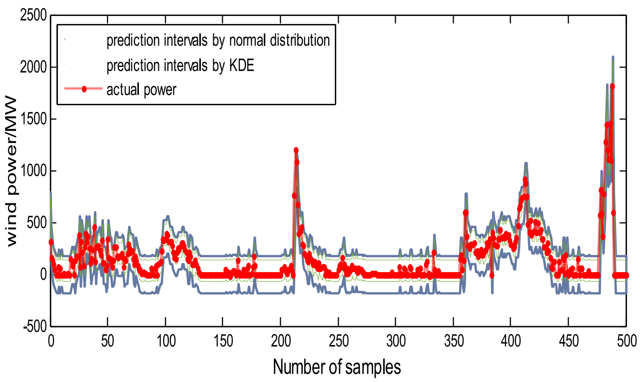

FIGURE 7. Prediction intervals under 85% confidence level.

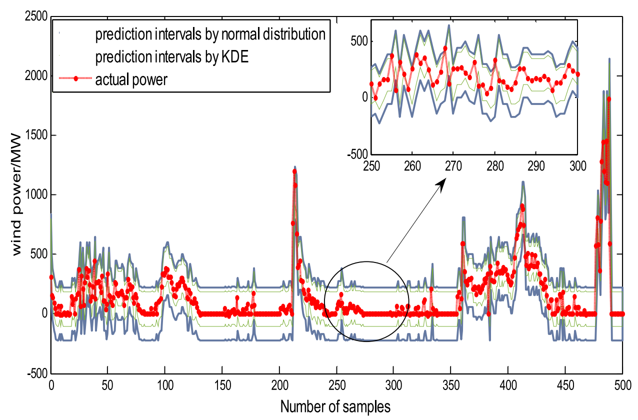

FIGURE 8. Prediction intervals under 90% confidence level.

band width. Since the error distribution might not follow the normal distribution in real wind power prediction, the performance of PIs by normal distribution is not satisfied. It can be seen from the the window in figure 8 that KDE can provide a diffrient upper and lower bound and get the narrow band width.

To further evaluate the performance of two prediction methods in Figs 6,7 and 8, the Prediction interval coverage probability (PICP) and the Prediction interval normalized average width (PINAW) introduced from reference [26] are used for the prediction evaluation.

TABLE 3. PI evaluation index.

CI/%		PICP/%	PINAW
90	KDE	91.02	289.1067
	Normal Distribution	91.91	332.1713
85	KDE	85.83	239.6367
	Normal Distribution	86.97	269.8892
80	KDE	80.04	157.2267
	Normal Distribution	81.98	217.9874

$$PICP = \frac{1}{N_t}\sum_{i=1}^{N_t} k_i^{(\alpha)} \tag{22}$$

In which N_t represents the sample number; $k_i^{(\alpha)}$ represents the Boolean value, 1: prediction falls within the prediction interval, 0: outside of the prediction interval

$$PINAW = \frac{1}{N_t}\sum_{i=1}^{N_t}[U_t^{(\alpha)}(x_i) - L_t^{(\alpha)}(x_i)] \tag{23}$$

Where: $U_t^{(\alpha)}$ represents the prediction interval upper bound; $L_t^{(\alpha)}$ represents the prediction interval lower bound. A small PINAW value indicates a narrow interval and high prediction precision if PICP is constant.

The PI evaluation index is shown in Table 3:

Table 3 shows that at each confidence level, the true value can effectively fall within the given prediction interval for both methods, and the prediction bandwidth widens as the confidence level increases, which is consistent with theoretical knowledge. Normal distribution method has wider band width than KDE method for higher PICP. Table 3 shows that the proposed method is stable and effective.

D. COMPARATIVE VALIDATION

To further validate the effectiveness of the algorithm, the results from KDE with the Markov chain method, KDE with weighted Markov chain, and KDE with rough set-weighted Markov chain are compared based on selected same training samples and test samples. The PICP and PINAW are used as criterions to evaluate the quality of the interval. The PICP and PINAW values of the three algorithms under different confidence levels are shown in Table 4, in which MK represents KDE with the Markov chain method, WMK represents the KDE with weighted Markov chain, and RWMK represents the KDE with rough set-weighted Markov chain method.

As seen in Table 4, the proposed method provides satisfactory PI performance. RWMK has a smaller PINAW: 157.2267, 239.6367, 289.1067 at PICP = 80%, 85%, 90% confidence level, respectively, which shows that the smaller the interval average bandwidth, the higher the prediction accuracy, and the smaller the degree of uncertainty. The three proposed methods almost produce identical PICP reliability results, which obtain better coverage of real power values

TABLE 4. Performance index comparison of different algorithms.

CI (%)		PICP (%)	PINAW
80	MK	80	294.9
	WMK	81.24	247.3067
	RWMK	80.04	157.2267
85	MK	85.8.	349
	WMK	86.42	335.2267
	RWMK	84.83	239.6367
90	MK	90	436
	WMK	91.82	432.8933
	RWMK	91.02	289.1067

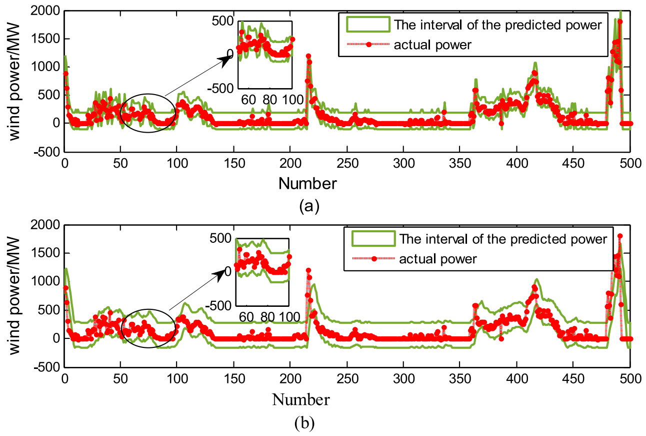

FIGURE 9. Prediction intervals by RWMF and WMF at 85% confidence level. (a)Prediction intervals by RWMK method. (b)Prediction intervals by WMK method.

with better PI reliability. In conclusion, the proposed method offers the best prediction result.

Fig. 9 further shows a comparison of prediction interval with different weight methods, which are calculated via the predicted WMK and WMK intervals at a confidence probability of 85%. (a) shows a prediction interval diagram for determining the weights of a rough set, and (b) shows a prediction interval diagram for determining weights via the parameter estimation method.

Fig. 9 shows that the proposed weighted Markov chain weight model based on rough set can ensure the tracking interval of wind power, while having narrow upper and lower bounds, thus it can provide decision makers with better uncertainty information. It utilizes the advantage of the rough set of dealing with the uncertain information. Furthermore, the attribute significance can analyze the effect of the non-synchronous Markov chain on the predicted true value, thus generating the weighted weight of a more reasonable Markov model.

V. CONCLUSION

This paper presents a wind power interval probability prediction model that uses the non-parametric kernel density with rough set theory and weighted Markov chain method. The rough set is used to determine the weights in weighted Markov chains, which can improve the accuracy of the power prediction. The KDE method is then applied to predict the wind power probability interval based on power error data. In addition, the method proposed in this paper for wind power interval prediction has the following characteristics:

1) The range of fluctuations that may occur in the predicted power at any given confidence level can be obtained without an assumption of the error distribution of the wind power.

2) The weighted Markov chain model applied to different samples of modeling and forecasting is capable to reveal the influence of the transition probability matrix at each step on the prediction and effectively data-mine the initial data contained in the information.

3) The use of the rough set can determine the reasonable weight, which does not require any prior information and is capable of obtaining the objective significance of each conditional attribute by analyzing the data.

4) The wind power interval prediction model based on the Gaussian distribution kernel function is developed. Moreover, the prediction interval coverage probability and the interval average bandwidth are used to evaluate interval precision. Finally, a comparison shows the superiority of the method introduced in this paper.

REFERENCES

[1] N. W. Miller, D. Guru, and K. Clark, "Wind generation," *IEEE Ind. Appl. Mag.*, vol. 15, no. 2, pp. 54–61, Mar. 2009.

[2] L. Landberg, "Short-term prediction of the power production from wind farm," *J. Wind Eng. Ind. Aerodyn.*, vol. 80, nos. 1–2, pp. 207–210, 1999.

[3] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 849–856, Oct. 2013.

[4] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[5] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew Sustain. Energy Rev.*, vol. 32, no. 5, pp. 255–270, 2014.

[6] S. Tewari, C. J. Geyer, and N. Mohan, "A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2031–2039, Nov. 2011.

[7] J. W. Taylor and J. Jeon, "Forecasting wind power quantiles using conditional kernel estimation," *Renew. Energy*, vol. 80, pp. 370–379, Aug. 2015.

[8] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renew. Energy*, vol. 40, pp. 29–39, Apr. 2012.

[9] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 983–988, Oct. 2011.

[10] X. Peng, W. Zheng, D. Zhang, Y. Liu, D. Lu, and L. Linb, "A novel probabilistic wind speed forecasting based on combination of the adaptive ensemble of on-line sequential ORELM (outlier robust extreme learning machine) and TVMCF (time-varying mixture copula function)," *Energy Convers. Manage.*, vol. 138, pp. 587–602, Apr. 2017.

[11] X. S. Lu, X. Yan, and J. Chen, "Ultra-short-term wind power intervals prediction considering randomness of wind power generation," *ActaEnergiae Solaris Sinica*, vol. 38, no. 5, pp. 1307–1315, 2017.

[12] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb. 2014.

[13] 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.

[14] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, pp. 95–108, Jan. 2006.
[15] G. Zhang, Y. Wu, K. P. Wong, Z. Xu, Z. Y. Dong, and H. H. C. Iu, "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2706–2715, Sep. 2015.
[16] 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.
[17] J. W. Taylor, "A quantile regression neural network approach to estimating the conditional density of multiperiod returns," *J. Forecast.*, vol. 19, no. 4, pp. 299–311, 2000.
[18] Y. Feng, R. Li, A. Sudjianto, and Y. Zhang, "Robust neural network with applications to credit portfolio data analysis," *Statist. Interface*, vol. 3, no. 4, p. 437, 2010.
[19] J. Juban, L. Fugon, and G. Kariniotakis, "Uncertainty estimation of wind power forecasts: Comparison of probabilistic modelling approaches," in *Proc. Eur. Wind Energy Conf. Exhib. EWEC*, Brussels, Belgium, Mar. 2008, p. 10.
[20] J. Tang, A. Brouste, and K. L. Tsui, "Some improvements of wind speed Markov chain modeling," *Renew. Energy*, vol. 81, pp. 52–56, Sep. 2015.
[21] Z. Song, X. Geng, A. Kusiak, and C. Xu, "Mining Markov chain transition matrix from wind speed time series data," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10229–10239, 2011.
[22] G. D'Amico, F. Petroni, and F. Prattico, "First and second order semi-Markov chains for wind speed modeling," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 5, pp. 1194–1201, 2013.
[23] J. Chen and Y. Yang, "SPI-based regional drought prediction using weighted Markov chain model," *Res. J. Appl. Sci. Eng. Technol.*, vol. 4, no. 21, pp. 4293–4298, 2012.
[24] D. Sengupta, U. Maulik, and S. Bandyopadhyay, "Weighted Markov chain based aggregation of biomolecule orderings," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 3, pp. 924–933, May 2012.
[25] Z. Pawlak, "Rough sets and intelligent data analysis," *Inf. Sci.*, vol. 147, nos. 1–4, pp. 1–12, Nov. 2002.
[26] 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.
[27] A. Carpinone, M. Giorgio, R. Langella, and A. Testa, "Markov chain modeling for very-short-term wind power forecasting," *Electr. Power Syst. Res.*, vol. 122, pp. 152–158, May 2015.
[28] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[29] Y. Zhang, J. Wang, and X. Luo, "Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel," *Energy Convers. Manage.*, vol. 96, pp. 440–451, May 2015.

XIYUN YANG was born in Tongliao, Neimenggu, China, in 1973. She received the Ph.D. degree from North China Electric Power University, Beijing, China, in 2004. Her research interests include renewable energy, energy analytics, smart power systems, and electricity markets.

She is currently a Professor with the School of Control and Computer Engineering, North China Electric Power University.

XUE MA was born in Tangshan, Hebei, China, in 1993. She received the bachelor's degree from North China University of Science and Technology, Hebei, in 2016.

She is currently a Researcher with the School of Control and Computer Engineering, North China Electric Power University. Her research interests include renewable energy and wind power forecasting, and complex system modeling.

NING KANG was born in Neimenggu, China, in 1995. She received the bachelor's degree from North China Electric Power University, Beijing, China, in 2017.

She is currently a Researcher with the School of Control and Computer Engineering, North China Electric Power University. Her research directions are renewable energy and wind power forecasting.

MIERZHATI MAIHEMUTI was born in Kuche, Xinjiang, China, in1995. He received the bachelor's degree from North China Electric Power University, Beijing, China, in 2017.

He is currently a Researcher with the School of Control and Computer Engineering, North China Electric Power University. His research interests include renewable energy and wind power forecasting, and data mining.