Bayesian Network Based Imprecise Probability Estimation Method for Wind Power Ramp Events

Yuanchun Zhao, Wenli Zhu, Ming Yang, and Mengxia Wang

Abstract—Although wind power ramp events (WPREs) are relatively scarce, they can inevitably deteriorate the stability of power system operation and bring risks to the trading of electricity market. In this paper, an imprecise conditional probability estimation method for WPREs is proposed based on the Bayesian network (BN) theory. The method uses the maximum weight spanning tree (MWST) and greedy search (GS) to build a BN that has the highest fitting degree with the observed data. Meanwhile, an extended imprecise Dirichlet model (IDM) is developed to estimate the parameters of the BN, which quantificationally reflect the ambiguous dependencies among the random ramp event and various meteorological variables. The BN is then applied to predict the interval probability of each possible ramp state under the given meteorological conditions, which is expected to cover the target probability at a specified confidence level. The proposed method can quantify the uncertainty of the probabilistic ramp event estimation. Meanwhile, by using the extracted dependencies and Bayesian rules, the method can simplify the conditional probability estimation and perform reliable prediction even with scarce samples. Test results on a real wind farm with three-year operation data illustrate the effectiveness of the proposed method.

Index Terms—Bayesian network (BN), conditional probability, imprecise Dirichlet model (IDM), imprecise probability, wind power ramp events.

I. INTRODUCTION

LARGE-SCALE wind power has been continuously integrated into power systems. However, the inherent randomness and volatility of wind power cause an increasing impact on power system operation [1]. The significant changes of wind power in a short period are often referred to as wind power ramp events (WPREs) [2]. Although such events are relatively scarce, the unanticipated sudden increases or decreases in wind power could inevitably deteriorate the stability of power systems [3]-[5]. Meanwhile, the disturbances caused by WPRE may hurt the interests of energy traders by exposing them to the risk of financial penalties, and thus reduce the vendibility of wind energy [6].

Accurate warning of WPRE can not only provide indication for scheduling backup resources to mitigate the impacts [7], [8], but also facilitate market participants to better understand the risks involved in the trades [9]. Although the WPRE prediction has attracted widespread attention in recent years, the research on WPRE still has broad prospects and profound significance. The deterministic WPRE prediction can be roughly divided into two categories [10]. One is dedicated to providing deterministic numerical descriptions of ramp characteristics such as magnitude, duration, ramp rate and timing [9], [11]. The other regards the ramp event as a multi-state random variable and alerts the approaching ramp event by identifying the state with the highest probability [4], [12], [13].

However, the scarcity of WPRE may bring unavoidable statistical errors to the deterministic WPRE prediction [14]. Under this circumstance, the uncertain WPRE prediction, which can assist system operators and market participants to make more informed and advisable decisions, has attracted increasing attention. Reference [15] adopted the numerical weather prediction (NWP) ensemble to realize the uncertain estimation of ramp timing. Reference [16] provided temporal uncertainty information of WPRE using wind power scenarios generated from quantile prediction. Reference [17] predicted wind power for different scenarios based on the neural network method, so as to provide statistics on ramp swing, ramp timing and the duration. For the probabilistic prediction of the multi-state random WPRE, the intuitive way is to predict the wind power distribution at each prediction moment, and then the statistical random sampling on two adjacent distributions are used to estimate the probability of the possible power changes [18]. Nevertheless, it is pointed out in [18] that the results predicted by this way are lack of applicability due to the failure of capturing temporal correlations of the wind power at adjacent moments.

To overcome this deficiency, [18] converted the NWP ensemble into wind power ensemble using the random forest model, and then detected WPRE based on the predicted wind power members to improve the capture ratio. By categorizing the statistical scenarios, the principal component analysis (PCA) based method in [19] can directly estimate the ramp probability based on the observed wind speed series. The method does not need to predict the power before performing WPRE prediction.

Although significant progress has been made on the probabilistic prediction of WPRE, the historical samples are always assumed to be sufficient for making reliable probability distribution estimation. In other words, the precondition of the law of large numbers is always assumed to be satisfied. However, when counting the ramp probabilities under different meteorological conditions, due to the scarcity of WPRE, the estimation errors introduced by the finite sample statistics are inevitable, leading to unreliable probabilistic prediction results.

Under this circumstance, this paper proposes a Bayesian network (BN) based method to estimate the imprecise conditional probability of WPRE. The event is modeled as a multistate random variable, where different states correspond to different ramp magnitudes. Based on the historical observations of wind power and meteorological elements, the maximum weighted spanning tree (MWST) and greedy search (GS) are applied to objectively excavate the conditional dependencies among variables and construct the BN structure that best fits the samples. An extended imprecise Dirichlet model (IDM) is then developed to quantify these unclear dependencies and establish the imprecise conditional probability table (CPT) at each node. Thus, the interval probability of each ramp state under a given meteorological condition can be inferred using the BN-based probability inference algorithm. Case studies on a wind farm located in Ningxia, China verify the effectiveness of the method in prediction scenarios with sufficient and insufficient samples, respectively.

The advantages of the proposed method lie in:

1) Wind power prediction methods intentionally ignore the extreme samples and obtain a relatively smooth curve to minimize the overall error [10]. Therefore, the methods that detect WPRE from the predicted power series may underestimate ramp probability [7]. In contrast, the proposed method directly explores the ramp probabilities under different meteorological conditions, avoiding the cumulative errors caused by power prediction.

2) The paper applies imprecise probability to quantify the uncertainty of the probability estimation with insufficient samples. The prediction results represented by imprecise probability mass functions can effectively estimate the range of occurrence probability of each ramp state, providing more comprehensive information than the traditional deterministic probability prediction.

3) In most methods based on scenario categorization, the sample sizes corresponding to the extreme scenarios are usually small, which may cause unreliable estimation results [19]. By extracting the dependency between WPRE and each meteorological variable, the proposed BN-based method can increase the valid sample size in conditional prediction and improve the prediction reliability even under some extreme conditions.

The rest of the paper is organized as follows. Section II defines the WPRE quantitatively. Section III introduces the theoretical basis of BN, which includes the structure learning and probability inference algorithm. Section IV expresses the standard IDM (SIDM) and extends it for our research. Section V applies the BN and extended IDM (EIDM) on WPRE prediction. Test results on a real wind farm are analyzed in Section VI and conclusions are drawn in Section VII.

II. DEFINITION OF WPREs

The WPREs are defined as significant upward or downward wind power variations in fixed short-time intervals. The threshold of variation applied for detection can be within a specific megawatt or a percentage of the installed capacity [20]. The discriminant is expressed as:

$$|P_{t+\Delta t} - P_t| > P_\varepsilon \text{ or } |P_{t+\Delta t} - P_t| > rP_R \tag{1}$$

where P_t and $P_{t+\Delta t}$ are the observed wind power at moment t and $t+\Delta t$, respectively; P_ε is the threshold value; P_R is the installed capacity; r is the specified percentage; and Δt is the time interval.

At present, an extensive consensus for the setting of P_ε and r in (1) has not been achieved. In most cases, they should be carefully assigned according to the actual requirement of power grid. A variety of setting methods for the thresholds have been discussed in [10].

III. CONSTRUCTION AND INFERENCE ALGORITHMS OF BN

A. Structure of BN

The BN utilizes the directed acyclic graph (DAG) to express the conditional dependencies among variables [21]. In a BN, the nodes $X_1, X_2, ..., X_n$ represent the random variables and the directed edges represent the dependencies among these variables. The CPT attached to each node stores the conditional probability distribution of the variable given the instantiation of its parents $Pa(X_i)$, i. e., $P(X_i|Pa(X_i)), i = 1, 2, ..., N$.

By extracting the dominant dependencies among the variables, the BN simplifies the network-based inference, thereby avoiding dimension curse during the prediction. In this paper, the discrete BN is applied for WPRE prediction, whose variables all have discrete values.

B. BN Structure Learning Algorithms

BN structure learning is to construct a DAG that best fits the observations and expresses the hidden conditional dependencies abstractly. BN structure learning algorithms can be roughly divided into two categories:

1) The constraint-based structure learning algorithms check the dependencies by conditional independence tests (CITs) [22]. These approaches are simple and intuitive, but sensitive to the accuracy of CIT and easy to cause error propagation and accumulation during the learning.

2) The score-and-search-based structure learning algorithms judge the quality of the structure through scoring functions and search for the optimal structure intelligently [23]. However, the search space may exhibit exponential growth as the number of nodes increases, which aggravates the computation burden.

To integrate the above advantages, this paper applies an MWST-initialized GS algorithm to build the BN structure, which is denoted as MWST-GS for convenience.

Since searching for the optimal structure in the vast space of network structure is NP-hard [24], the heuristic search algorithms, typically represented by the GS [25], are usually adopted. Beginning with an initial structure, the GS algorithm locally updates the current structure in each iteration, and evaluates all new structures obtained by the scoring function. The quality of the structure is evaluated by the Bayesian information criterion index, which uses the likelihood function to describe the fitting degree, and punishes the complexity to avoid overfitting [26]. If the optimal candidate is superior to the current, replace the current with the optimal one and continue to search; otherwise, stop searching and the current structure is the final result.

In the GS algorithm, an unreasonable initial structure could lead to a complex search iteration process and even create a local optimal solution. To overcome this defect, the MWST algorithm is applied to build an initial tree structure based on the captured plain dependencies. Thus, the initial structure of the GS algorithm can be confined to the neighborhood of the global optimal structure.

MWST is a constraint-based structure learning algorithm. It uses the mutual information (MI) index to measure the dependency between every two variables.

$$MI(X,Y)=\sum_{x,y}\left(P(X=x,Y=y)\lg\frac{P(X=x,Y=y)}{P(X=x)P(Y=y)}\right) \quad (2)$$

where X and Y are the random variables; $P(X=x,Y=y)$ is the joint probability distribution function; and $P(X=x)$ and $P(Y=y)$ are the marginal probability distribution functions of X and Y, respectively.

When the MWST algorithm is executed, it firstly sorts the edges in descending order of MI values of the connected variables. Then, the edges can be successively added to construct the oriented tree, as long as the newly added edge does not form a cycle [27]. When all n nodes have been connected by $n-1$ edges, the algorithm stops.

The MWST-GS algorithm inherits the computation efficiency of the MWST algorithm, and the employed GS algorithm can eliminate the errors in the high-order independence tests of the MWST algorithm. Figure 1 provides the flowchart of the MWST-GS algorithm applied in this paper.

C. Probability Inference Algorithm

The BN inference is to estimate the concerned posterior probability according to the observed evidence and the network describing the dependencies among target and evidence variables. The inference rules can be intuitively demonstrated with the structure shown in Fig. 2.

Suppose A and C are both two-state nodes, and B is a three-state node, i.e., $A=\{A_i|i=1,2\}$, $B=\{B_k|k=1,2,3\}$, $C=\{C_d|d=1,2\}$. With the given evidence, e.g., $\{B_k,C_d\}$, the occurrence chance of the state A_i can be expressed as $P(A_i|B_k,C_d)$. According to Bayesian rules, this conditional probability can be rewritten as:

$$P(A_i|B_k,C_d)=\frac{P(A_i)P(B_k,C_d|A_i)}{\sum_{i=1}^{2}P(A_i)P(B_k,C_d|A_i)} \quad (3)$$
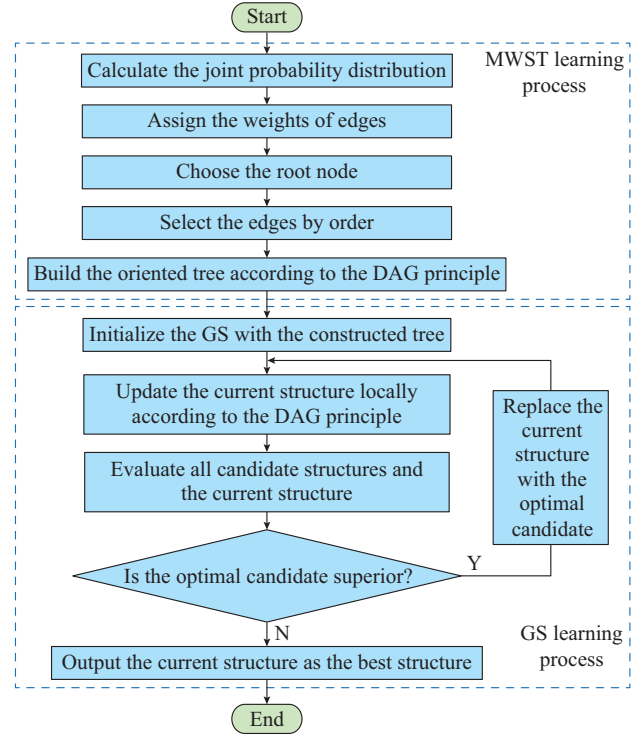

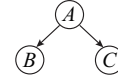Fig. 1. Flowchart of MWST-GS algorithm for BN structure learning.

Fig. 2. Simple three-node structure example.

With respect to the independence indicated in Fig. 2, the conditional joint probability $P(B_k,C_d|A_i)$ can be further factorized according to the chain rule as:

$$P(B_k,C_d|A_i)=P(B_k|A_i)P(C_d|A_i,B_k)=P(B_k|A_i)P(C_d|A_i) \quad (4)$$

Therefore, the conditional probability $P(A_i|B_k,C_d)$ can be calculated with respect to the CPT at each node as:

$$P(A_i|B_k,C_d)=\frac{P(A_i)P(B_k|A_i)P(C_d|A_i)}{\sum_{i=1}^{2}P(A_i)P(B_k|A_i)P(C_d|A_i)} \quad (5)$$

In summary, the Bayesian rules and the chain rule factorize the concerned posterior probability, and the implied conditional independence simplifies the expression. Thereby, the BN inference algorithm facilitates the posterior distribution estimation of the target variable and improves the computation efficiency.

IV. SIDM AND EIDM

The imprecise probability theory employs interval probability to express the occurrence chance of a random event, which describes the uncertainty in statistics. IDM is an effective method for imprecise probability estimation with respect to insufficient samples [28].

A. SIDM

In the process of deterministic multinomial distribution estimation, an arbitrary Dirichlet distribution is commonly ad-

opted as the prior distribution. However, without sufficient samples, an improper prior distribution may lead to subjective estimation results. The IDM develops the deterministic Dirichlet model in the context of imprecise probability, and employs a set of possible prior density functions to avoid the subjective results [29].

Consider a multi-nomial variable that has n states, and the occurrence chance of each state is represented by $\theta_i, i = 1, 2, ..., n$. Then, with respect to the observations, the imprecise posterior estimation of θ_i obtained by IDM is:

$$\theta_i \in \left[\frac{m_i}{M+s}, \frac{m_i+s}{M+s} \right] \qquad (6)$$

where m_i is the number of times that the ith state is observed; and M is the total number of observations. The hyper parameter s determines the degree of imprecision, i.e., the larger s is, the more reliable the estimation will be [30]. However, if s is too large, the estimation will be weakened. Till now, an extensive consensus for the setting of s has not been achieved. However, $s=2$ is generally considered to be enough cautious in various applications [28], [30].

Although the historical samples of ramp phenomena are usually insufficient, a considerable number of samples, e. g., a few hundreds, are still available for quantifying the dependencies among variables. In this case, as shown in Fig. 3, the interval estimated by (6) with $s=2$ is very narrow, which may cause unreliable results of ramp probability prediction. This phenomenon is analyzed below. It is also noted that in Fig. 3, u is a parameter of EIDM to be discussed in Section IV-B.
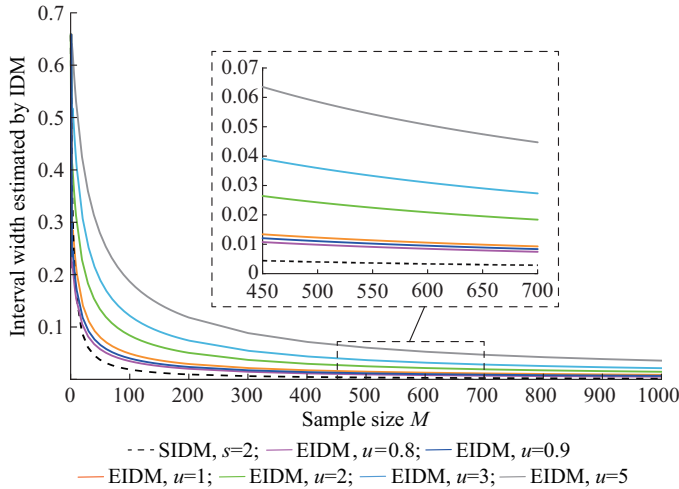

Fig. 3. Reduction of uncertainty, i. e., interval width, in SIDM and EIDM with increasing sample size M.

Suppose that the real occurrence probability of event Ω is $p \in [0, 1]$, and M samples are available. Therefore, the expected times that Ω occurs is $Mp = m$. Due to the randomness of sampling, the actual observation result might be $m+1$ (when M is large, this experimental bias is very likely to happen). In this case, the occurrence probability estimated by SIDM ($s=2$) is $\left[(m+1)/(M+2), (m+3)/(M+2) \right]$. It is found that the lower bound of the interval is larger than the target value p when $p<0.5$, so the estimated result fails to cover the real probability in this case.

On the contrary, if Ω is observed for $m-1$ times, the interval probability provided is $\left[(m-1)/(M+2), (m+1)/(M+2) \right]$. It is found that the upper bound of the interval is smaller than the target value p when $p>0.5$, indicating that the estimated result fails to cover the real probability.

In summary, for a relatively large M, when $p<0.5$ and the counted times of Ω are no less than $Mp+1$, or when $p>0.5$ and the counted times are no more than $Mp-1$, the intervals estimated by SIDM ($s=2$) will deviate from the real probability. The constant hyper parameter, e. g., $s=2$, cannot be suitable for all M, which is the main reason to raise this issue.

B. EIDM

The IDM is extended to overcome its deficiency and improve the chance of the predicted ramp interval probability covering the target probability. As shown in (7), the hyper parameter s in EIDM is designed to be a function of the sample size M, where the convergence speed of the interval probability is controlled by an exogenous parameter u that can be optimized to improve the results of WPRE prediction.

$$s = \begin{cases} 2 & M = 0, 1 \\ u \lg M & M \geq 2, u \in \mathbf{R}^+ \end{cases} \qquad (7)$$

Through the similar derivation process, the interval probability estimated by EIDM can be expressed as:

$$\theta_i \in \begin{cases} \left[\dfrac{m_i}{M+2}, \dfrac{m_i+2}{M+2} \right] & M = 0, 1 \\[4mm] \left[\dfrac{m_i}{M+u \lg M}, \dfrac{m_i+u \lg M}{M+u \lg M} \right] & M \geq 2, u \in \mathbf{R}^+ \end{cases} \qquad (8)$$

The rationality of EIDM is explained as below:

1) The advantages of SIDM are inherited for $M = 0, 1$.

The value of s for $M = 0, 1$ are inherited from SIDM. When no sample is available, i.e., $M = 0$, the estimated probability will be [0, 1], indicating the prior ignorance. On the other hand, when one sample is obtained and the concerned event occurs, i.e., $M = 1$ and $m = 1$, the estimated posterior probability will be [1/3, 1]. Otherwise, if the event does not occur, i. e., $M = 1$ and $m = 0$, the estimation result will be [0, 2/3]. As discussed in [28], the estimated interval can permit a relatively high degree of imprecision.

2) EIDM possesses the convergence property.

When M approaches infinity, according to the L'Hospital rule [31], the uncertainty measured by interval probability can reduce to zero, which conforms to the law of large numbers. And the property can be proved by:

$$\lim_{M \to \infty} \frac{u \lg M}{M + u \lg M} = \lim_{M \to \infty} \frac{u/M}{1 + u/M} = \lim_{M \to \infty} \frac{u}{M + u} = 0 \qquad (9)$$

3) A minimum u can be found to ensure the coverage of the target probability for any sampling result.

The interval probability estimated by EIDM is $\left[(m+a)/(M+u \lg M), (m+a+u \lg M)/(M+u \lg M) \right]$ when Ω occurs for $m+a$ ($a>0$) times. The target probability p can

be covered when $(m+a)/(M+u\lg M)\le p$ (or $u\ge a/(p\lg M)$) and $(m+a+u\lg M)/(M+u\lg M)\ge p$ (or $u\ge -a/[(1-p)\lg M]$).

Obviously, the second inequality above is always satisfied. Therefore, as long as $u\ge a/(p\lg M)$, EIDM can guarantee the coverage of p.

On the contrary, if event Ω is observed for $m-a$ $(a>0)$ times, the estimated result $\left[(m-a)/(M+u\lg M),(m-a+u\lg M)/(M+u\lg M)\right]$ will cover the target probability p when $(m-a)/(M+u\lg M)\le p$ (or $u\ge -a/(p\lg M)$) and $(m-a+u\lg M)/(M+u\lg M)\ge p$ (or $u\ge a/[(1-p)\lg M]$).

Here the first inequality is always satisfied. Thus, as long as $u\ge a/[(1-p)\lg M]$, the coverage can be guaranteed.

In summary, if the coverage of the target probability is required as long as the observed times of Ω is within $[m-a, m+a]$, u should satisfy the following condition:

$$u\ge\max\left(\frac{a}{p\lg M},\frac{a}{(1-p)\lg M}\right)=\begin{cases}\dfrac{a}{p\lg M} & p\le 0.5\\[2mm]\dfrac{a}{(1-p)\lg M} & p>0.5\end{cases} \quad (10)$$

Obviously, the minimum u cannot be directly obtained from (10) since p and a are both unknown. Instead, an approximate u can always be estimated according to the experiential data using heuristic optimization algorithms. The corresponding EIDM curves in Fig. 3 illustrate that by specifying a proper u, the uncertainty involved in the estimated interval probability can be preserved even for the modest samples. In fact, the absolute reliability is not necessary and too cautious interval probabilities may disguise the valid information. To avoid unreliable and over-cautious parameter setting, the proposed method seeks u that leads to the best prediction performance with respect to a comprehensive criterion.

Based on the algorithms mentioned above, the proposed prediction method can be divided into four key steps, as shown in Fig. 4.

V. Case Study

The BN and EIDM are applied for WPRE prediction carried out on a wind farm located in Ningxia, China. The details of the prediction are provided in this section.

A. Data Description

The installed capacity of the wind farm is 36 MW. The data including wind power and four meteorological measurements, i.e., wind speed, wind direction, temperature and humidity, are from January 1, 2015 to December 31, 2017, and the time resolution is 30 min. All the meteorological measurements are integrated into the BN as the candidate evidence variables. Moreover, since sudden changes in wind speed can easily trigger WPRE, the wind speed variation in 30 min calculated by (11) is selected as an additional candidate evidence variable.

$$V_t = S_t - S_{t-30} \quad (11)$$

where S_t and S_{t-30} are the wind speeds at moment t and the previous moment, respectively. The identification thresholds of WPRE are designed [32]:
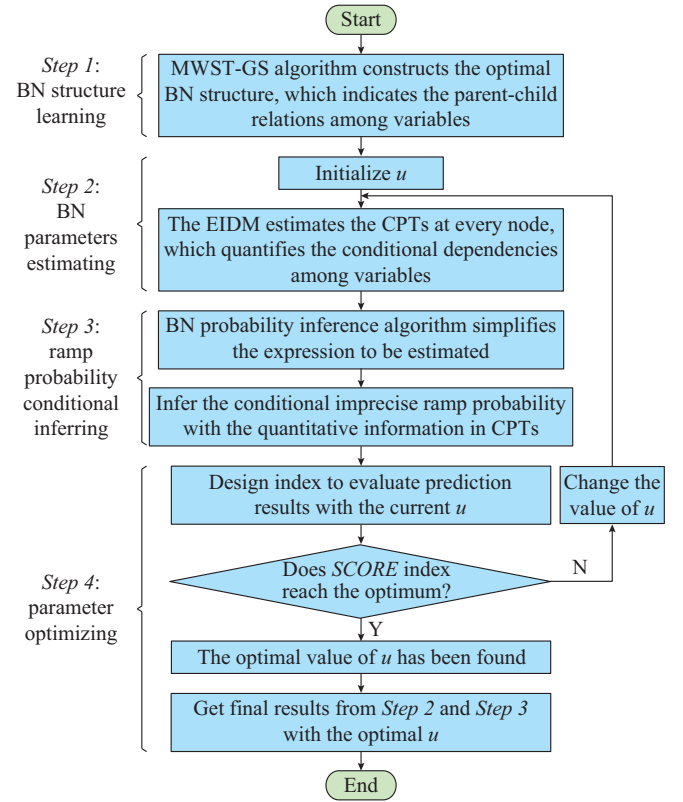

Fig. 4. Four key steps of proposed prediction method.

$$\begin{cases}-11\%\times P_R\le P_t-P_{t-30}\le 10\%\times P_R & \text{no events}\\ P_t-P_{t-30}>10\%\times P_R & \text{ramp up events}\\ P_t-P_{t-30}<-11\%\times P_R & \text{ramp down events}\end{cases} \quad (12)$$

where P_{t-30} is the observed wind power at the previous moment.

The dataset is divided into a training set and a validation set, as shown in Table I. The training set is used to build the prediction model and optimize the parameter u, while the validation set is used to verify the effectiveness.

TABLE I
DATASET DESCRIPTION

Dataset	Time span	Sample size	Ramp up events	Ramp down events
Training set	January 2015 to June 2016	25774	2864	2328
Validation set	July 2016 to December 2017	25084	2703	2304

Every candidate evidence variable is divided into three states by equal-frequency discretization process [33]. Table II provides the state list of the variables.

B. Construction of BN

A tree structure shown in Fig. 5(a) is firstly built by the MWST algorithm presented in Section III-B to extract the heuristic knowledge about the hidden dependencies for initializing the BN structure. With the initial structure, the GS algorithm can locally update the structure and evaluate whether the current structure is superior to the optimal candidate.

TABLE II
STATES OF VARIABLES IN CONSTRUCTED DISCRETE BN

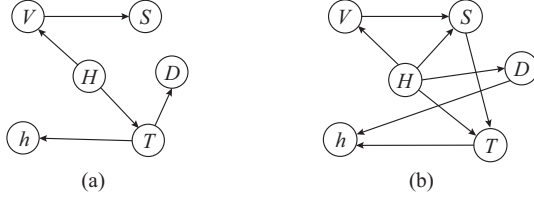| State	Wind speed variation V (m/s)	Wind speed S (m/s)	Wind direction D (°)	Temperature T (°C)	Humidity h (%)	Ramp events H (wind power variation) (MW)
1	V_1: [−11.2, −0.8)	S_1: [0, 3.6)	D_1: [1, 120)	T_1: [−21.7, 5.7)	h_1: [9.5, 36.5)	H_1: [−4.0, 3.6]
2	V_2: [−0.8, 0.8)	S_2: [3.6, 5.3)	D_2: [120, 240)	T_2: [5.7, 16.8)	h_2: [36.5, 58.5)	H_2: (3.6, 36]
3	V_3: [0.8, 16.7]	S_3: [5.3, 23.2]	D_3: [240, 360]	T_3: [16.8, 36.2]	h_3: [58.5, 98.5]	H_1: [−36, −4.0)

Fig. 5. Constructed BN structures. (a) Tree structure acquired from MWST. (b) Optimal structure acquired from GS.

According to the process in Fig. 1, the optimum network structure (as shown in Fig. 5(b)) that illustrates the dominant conditional dependencies can be established.

The constructed network structure indicates that WPRE are directly related to wind speed, wind direction, temperature and the wind speed variation. Therefore, the states of these four evidence variables can determine the probability distribution of WPRE.

C. Inference of Conditional Imprecise Probability of WPRE

All possible meteorological conditions can be expressed as $E_l = \{V_y, S_r, D_q, T_k, h_d\}$, where $y, r, q, k, d \in \{1, 2, 3\}$. There are totally 243 meteorological conditions. The conditional imprecise probability of the ramp state $H_w, w \in \{1, 2, 3\}$ can be inferred based on the BN shown in Fig. 5(b) as:

$$\begin{cases} \underline{P}(H_w|E_l) = \min \dfrac{P(H_w)P_{im}(V_y|H_w)P_{im}(S_r|H_w,V_y)P_{im}(D_q|H_w)P_{im}(T_k|H_w,S_r)}{\sum\limits_{w=1}^{3} P(H_w)P_{im}(V_y|H_w)P_{im}(S_r|H_w,V_y)P_{im}(D_q|H_w)P_{im}(T_k|H_w,S_r)} \\[4mm] \overline{P}(H_w|E_l) = \max \dfrac{P(H_w)P_{im}(V_y|H_w)P_{im}(S_r|H_w,V_y)P_{im}(D_q|H_w)P_{im}(T_k|H_w,S_r)}{\sum\limits_{w=1}^{3} P(H_w)P_{im}(V_y|H_w)P_{im}(S_r|H_w,V_y)P_{im}(D_q|H_w)P_{im}(T_k|H_w,S_r)} \end{cases} \quad (13)$$

In (13), the prior probability $P(H_w)$ can be counted as the occurrence frequency of H_w in the training set, and other quantitative information can be read from the CPTs estimated by EIDM.

D. Evaluation of Prediction Performance

The evaluation of the interval probability results mainly focuses on the reliability and the sharpness. The reliability indicates the capability of covering the target probability. The following criterion $SCORE_1$ is designed here for the reliability evaluation.

$$SCORE_1 = \begin{cases} 1 & P^*(H_w|E_l) \in \left[\underline{P}(H_w|E_l), \overline{P}(H_w|E_l)\right] \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where $[\underline{P}(H_w|E_l), \overline{P}(H_w|E_l)]$ is the predicted interval probability; and $P^*(H_w|E_l)$ is the target probability, which is replaced by the counted conditional frequency. A bigger $SCORE_1$ reflects a better performance on the reliability.

The sharpness evaluated by criterion $SCORE_2$ measures the imprecision degree of the interval. The smaller $SCORE_2$ is, the better the sharpness performance will be.

$$SCORE_2 = \overline{P}(H_w|E_l) - \underline{P}(H_w|E_l) \quad (15)$$

In addition, the weighted sum of $SCORE_1$ and $SCORE_2$ can be calculated to evaluate the comprehensive performance of the prediction method, which can be expressed as:

$$\begin{cases} SCORE = WT_1 \cdot SCORE_1 - WT_2 \cdot SCORE_2 \\ \text{s.t. } WT_1 > 0 \\ \quad WT_2 > 0 \\ \quad WT_1 + WT_2 = 1 \end{cases} \quad (16)$$

where WT_1 and WT_2 are the weights of $SCORE_1$ and $SCORE_2$, respectively. The weights can be specified according to the individual risk attitude. For a risk averter, a larger WT_1 should be selected to enhance the reliability of the predicted interval probabilities, and vice versa.

It can be observed that the larger the $SCORE$ value is, the better the comprehensive performance will be. Therefore, the optimization of parameter u is to find out u with the maximum $SCORE$. Then the optimized parameter u can be obtained by sensitivity analysis of $SCORE$. With the pre-set risk attitude, the optimal u can strike a balance between the reliability and sharpness.

VI. RESULT ANALYSIS

A. Comparison with Central Limit Theorem (CLT) Based Method

1) Prediction with Limited Samples

The CLT is commonly used for estimating the distribution of the statistical mean [34]. Suppose that μ and σ^2 represent the mean and variance of the samples, respectively. When the sample size M is large enough, CLT presents that the mean can be approximated by normal distribution $N(\mu, \sigma^2/M)$. Thus, the confidence interval of the mean can

be obtained accordingly.

Table III analyzes the average performance of WPRE prediction results under 243 meteorological conditions predicted by the CLT (90% confidence level) and BN. In the table, different weights, i.e., $WT_1 = 0.3$, 0.5, and 0.7, are applied to evaluate the performance of the methods.

TABLE III
ANALYSES OF PREDICTION RESULTS

WT_1	Prediction model	$SCORE_1$	$SCORE_2$	$SCORE$	Coverage rate (%)	Average width	Proportion of prediction results with interval width no less than 0.1 (%)	Proportion of prediction results with interval width no less than 0.2 (%)	Proportion of prediction results with interval width no less than 0.3 (%)
0.3	BN	570	88.7	108.9	78.20	0.122	60.50	8.20	0.80
	CLT	410	107.1	48.0	56.20	0.147	54.30	32.00	16.10
0.5	BN	608	105.2	251.4	83.40	0.144	70.30	16.80	2.40
	CLT	410	107.1	151.4	56.20	0.147	54.30	32.00	16.10
0.7	BN	646	133.2	412.2	88.60	0.183	84.00	38.30	8.20
	CLT	410	107.1	254.9	56.20	0.147	54.30	32.00	16.10

Risk seekers may choose $WT_1 = 0.3$, $WT_2 = 0.7$ to express their concerns about the sharpness. In this case, the average width of interval probabilities predicted by BN is only 0.122. More than 90% of the interval probabilities are narrower than 0.2, and only a minority of intervals are wider than 0.3, which are much better than those of the CLT model. Besides, during the test, 78.2% of the interval probabilities can cover the target probabilities, which are higher than those of the CLT model. This indicates the acceptable reliability performance of the BN model.

Contrarily, risk averters may choose $WT_1 = 0.7$, $WT_2 = 0.3$ to reflect their concerns about the reliability. In this case, the coverage rate obtained by BN model is close to 90%, indicating its remarkable reliability. However, the cost of pursuing high reliability is that when WT_1 increases from 0.3 to 0.7, the average width of the intervals increases by 50%.

The two models happen to provide almost the same average width when $WT_1 = WT_2 = 0.5$. In this case, the higher coverage rate of the BN model clearly reflects its superiority over the CLT model.

In summary, the following conclusions can be drawn from the test results shown in Table III.

1) The weights WT_1 and WT_2 reflect the individual prediction attitude. For a larger WT_1, more reliable prediction results will be provided. Meanwhile, the predicted interval probabilities will be relatively wider.

2) With respect to different risk attitudes, the parameter u in BN model can be optimized accordingly to obtain better performance. It is clearly observed that the interval probabilities predicted by BN model can be tuned according to the risk attitude, which indicates its flexibility.

3) Regardless of the individual risk attitude, the proposed model always exhibits a better performance, since it can always get a higher evaluation score.

To further verify the effectiveness of the proposed method, Fig. 6 provides the imprecise probabilities of predicted WPRE by using the BN and CLT under 8 different meteorological conditions. Under the meteorological conditions E_6, E_7, and E_8, weak prediction results are obtained by the CLT model. In these three cases, the narrower predicted intervals can always be obtained by the BN model, while all the target probabilities are well covered.
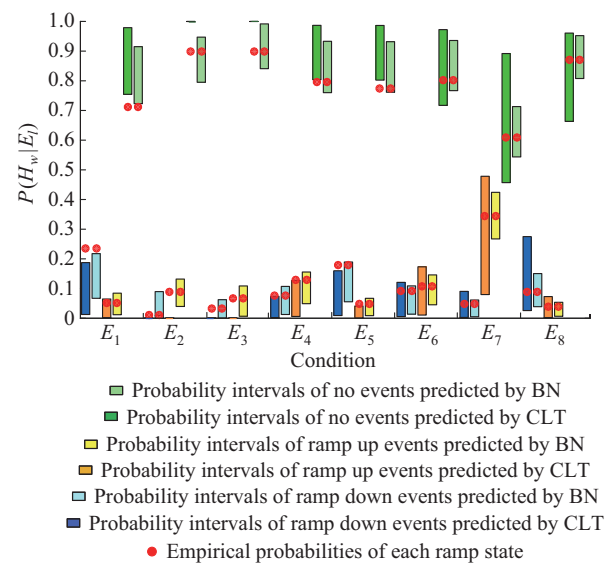

Fig. 6. Predicted imprecise probabilities of WPRE under 8 different meteorological conditions when $WT_1 = WT_2 = 0.5$.

Under the meteorological conditions E_4 and E_5, the counted empirical probabilities differ greatly between the training and validation sets because of the scarcity of samples in these meteorological conditions. In these cases, the CLT model does not work well, since its precondition, i.e., the number of samples is large enough, cannot be satisfied. On the contrary, by using the BN model, all the target probabilities can be covered in the predicted intervals, revealing the effectiveness of the proposed method with scarce samples.

Moreover, for the meteorological conditions E_2 and E_3, no ramp event occurs in the training set. In these cases, the CLT model can only generate deterministic prediction results, i.e., $P(H_1|E_l) = 1$ and $P(H_2|E_l) = P(H_3|E_l) = 0$, which leads to unbearable errors. On the contrary, it is observed that the proposed BN model still works well under these conditions and all the target probabilities can be covered.

Under the most adverse condition E_1, due to the severe scarcity of samples, gigantic differences exist in the counted empirical probabilities of training and validation sets, leading to the poor reliability performance of both the BN and

CLT models. However, less deviations corresponding to the proposed model indicate that the method can reflect the potential distribution more accurately even in this adverse case.

2) Prediction with No Samples

One feature of the proposed BN model is that it can still perform effective predictions without available samples under some extreme meteorological conditions. To illustrate this feature, the states of variable D, T and h are redefined in Table IV.

TABLE IV
ADJUSTMENT OF DISCRETIZATION OF VARIABLES D, T AND H

State	Wind direction D (°)	Temperature T (°C)	Humidity h (%)
1	D_1: [1, 200)	T_1: [−21.7, 14.0)	h_1: [9.5, 55.0)
2	D_2: [200, 318)	T_2: [14.0, 18.0)	h_2: [55.0, 75.0)
3	D_3: [318, 360]	T_3: [18.0, 36.2]	h_3: [75.0, 98.5]

Based on the new state definitions, no training sample exists for meteorological conditions $E_1 = \{V_1, S_3, D_2, T_3, h_3\}$, $E_2 = \{V_1, S_3, D_3, T_3, h_3\}$, $E_3 = \{V_3, S_1, D_3, T_2, h_3\}$ and $E_4 = \{V_3, S_1, D_3, T_3, h_3\}$. In these cases, the prediction results provided by the CLT model can only be $P_{im}(H_w|E_v) = [0, 1]$ where $w = 1, 2, 3$ and $v = 1, 2, 3, 4$. Obviously, the results cannot provide any useful information.

By using the proposed method, the network structure learned by MWST-GS algorithm with respect to all the training samples is shown in Fig. 7. The structure clearly illustrates that WPRE is directly related to wind speed, temperature and the wind speed variation in this situation, and the states of these three evidence variables will determine the probability distribution of WPRE.
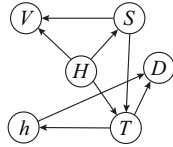

Fig. 7. BN structure constructed with redefined variable states.

Then, based on the BN probability inference rules explained in Section III-C, the interval probabilities of WPRE can be calculated by:

$$
\begin{cases}
\underline{P}(H_w|E_v) = \\
\quad \min \dfrac{P(H_w)P_{im}(S_r|H_w)P_{im}(V_y|H_w,S_r)P_{im}(T_k|H_w,S_r)}{\sum\limits_{w=1}^{3} P(H_w)P_{im}(S_r|H_w)P_{im}(V_y|H_w,S_r)P_{im}(T_k|H_w,S_r)} \\
\overline{P}(H_w|E_v) = \\
\quad \max \dfrac{P(H_w)P_{im}(S_r|H_w)P_{im}(V_y|H_w,S_r)P_{im}(T_k|H_w,S_r)}{\sum\limits_{w=1}^{3} P(H_w)P_{im}(S_r|H_w)P_{im}(V_y|H_w,S_r)P_{im}(T_k|H_w,S_r)}
\end{cases}
\tag{17}
$$

When $WT_1 = WT_2 = 0.5$, the prediction results corresponding to the extreme meteorological conditions are summarized in Table V. Although several predicted interval probabilities in Table V fail to cover the target probabilities, they can still pick out the ramp states that are most likely to happen, and

the tolerable deviations indicate the effectiveness of the proposed method under fresh meteorological conditions.

TABLE V
BN PREDICTION RESULTS UNDER EXTREME CONDITIONS

Conditional probability	Target probability	Interval probability
$P_{im}(H_1|E_1)$	0.846	[0.606, 0.833]
$P_{im}(H_2|E_1)$	0	[0.036, 0.140]
$P_{im}(H_3|E_1)$	0.154	[0.113, 0.301]
$P_{im}(H_1|E_2)$	0.667	[0.606, 0.833]
$P_{im}(H_2|E_2)$	0	[0.036, 0.140]
$P_{im}(H_3|E_2)$	0.333	[0.113, 0.301]
$P_{im}(H_1|E_3)$	0.600	[0.566, 0.952]
$P_{im}(H_2|E_3)$	0.400	[0.041, 0.357]
$P_{im}(H_3|E_3)$	0	[0, 0.163]
$P_{im}(H_1|E_4)$	1.000	[0.555, 0.954]
$P_{im}(H_2|E_4)$	0	[0.038, 0.349]
$P_{im}(H_3|E_4)$	0	[0, 0.199]

In summary, the constructed BN prediction model is applicable to all the meteorological conditions. For the condition with scarce samples or even no samples, the method can still obtain reliable prediction results. Actually, in the proposed method, the samples participated in the conditional prediction are not confined to the ones satisfying the concerned meteorological evidence, which makes it possible to fully excavate the statistical information of the training samples.

B. Comparison with Principal Component Analysis (PCA) Based Statistical Method

In order to evaluate the accuracy of the deterministic prediction results of the proposed method, a PCA-based statistical model is established [19] with respect to the training samples. The PCA uses an orthogonal transformation to convert the observations of possibly correlated variables into the values of linearly uncorrelated variables, i.e., principal components, thereby realizing the dimensionality reduction of multi-dimensional information. In [19], the first three principal components of wind speed time series are extracted and selected as the evidence variables to categorize the statistical conditions. The counted conditional frequencies of WPRE are used to evaluate the probabilities of the forthcoming WPRE.

The performances of the BN model and the PCA-based statistical model are evaluated using the validation set. Since the PCA-based statistical model can only provide deterministic results, the interval probabilities generated by BN model are firstly converted into their medians.

The ranked probability score (RPS) criterion [20], which compares the predicted cumulative density function (CDF) against the observed CDF, is selected to evaluate the performances of these two methods. For discrete cases, the RPS criterion can be expressed as:

$$
RPS = \frac{1}{K-1}\sum_{k=1}^{K}(F_k - O_k)^2
\tag{18}
$$

where K is the number of prediction categories; and $F_k = \sum_{i=1}^{k} f_i$ and $O_k = \sum_{i=1}^{k} o_i$ are the kth components of the predicted and observed CDF, respectively, f_i is the predicted probability of the event in category i, o_i is a binary variable that takes the value of 1 if the event is observed in category i. The RPS criterion calculated by (20) is a punitive function, and a smaller score reflects a better prediction performance.

By performing 25084 prediction tests with respect to the validation set, the score of the PCA-based statistical method is 3053.31, while the score of the BN-based prediction method is 2895.22. Therefore, although the proposed method does not take the deterministic probability prediction accuracy as an indicator to optimize the parameters, its outstanding capability of excavating statistical information from limited samples makes it perform better even for the deterministic probability prediction.

VII. Conclusion

In this paper, an imprecise probability estimation method for WPRE is proposed by combining the MWST-GS algorithm, EIDM, and the BN probability inference algorithm. The uncertainty of the ramp probability estimation can be quantificationally reflected by the interval probability. The method maps the WPRE to meteorological evidences directly, avoiding the prediction of wind power time series and the corresponding cumulative errors. The dominant conditional dependencies among the evidence and target variables are extracted by MWST-GS algorithm. Then, with respect to these dependencies, the method can perform reliable WPRE prediction by BN inference even with scarce samples. The EIDM developed in this paper can enhance the reliability of the estimated interval probability. Meanwhile, its exogenous parameter can be optimized according to the specified risk attitude to tune the prediction results, reflecting the flexibility of the method. Case studies of a wind farm located in Ningxia, China illustrate the effectiveness of the proposed method.

References

[1] Y. Gong, C. Y. Chung, and R. S. Mall, "Power system operational adequacy evaluation with wind power ramp limits," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2706-2716, May 2018.

[2] Y. Gong, Q. Jiang, and R. Baldick, "Ramp event forecast based wind power ramp control with energy storage system," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1831-1844, May 2016.

[3] M. Cui, C. Feng, and Z. Wang, "Statistical representation of wind power ramps using a generalized Gaussian mixture model," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 261-272, Jan. 2018.

[4] Y. Liu, Y. Sun, and D. Infield, "A hybrid forecasting method for wind power ramp based on orthogonal test and support vector machine (OT-SVM)," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 451-457, Apr. 2017.

[5] Y. Qi and Y. Liu, "Wind power ramping control using competitive game," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, pp. 1516-1524, Oct. 2016.

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

[7] Y. Liu, Y. Sun, and S. Han, "A WT-ARMA based method for wind power ramp events forecasting," in *Proceedings of 5th IET International Conference on Renewable Power Generation (RPG) 2016*, London,

UK, Sept. 2016, pp. 1-6.

[8] A. K. Nayak, K. C. Sharma, and R. Bhakar, "ARIMA based statistical approach to predict wind power ramps," in *Proceedings of 2015 IEEE PES General Meeting*, Denver, USA, Jul. 2015, pp. 1-5.

[9] M. Cui, J. Zhang, A. R. Florita *et al.*, "An optimized swinging door algorithm for identifying wind ramping events," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 150-162, Jan. 2016.

[10] C. Gallego-Castillo, A. Cuerva-Tejero, and O. Lopez-Garcia, "A review on the recent history of wind power ramp forecasting," *Renewable and Sustainable Energy Reviews*, vol. 52, no. 1, pp. 1148-1157, Dec. 2015.

[11] M. J. Cui, Y. Z. Sun, and D. P. Ke, "Wind power ramp events forecasting based on atomic sparse decomposition and BP neural networks," *Automation of Electric Power Systems*, vol. 38, no. 12, pp. 6-11, Jun. 2014.

[12] J. Zhang, M. Cui, and B. M. Hodge, "Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales," *Energy*, vol. 122, no. 1, pp. 528-541, Mar. 2017.

[13] H. Zareipour, D. Huang, and W. Rosehart, "Wind power ramp events classification and forecasting: a data mining approach," in *Proceedings of 2011 IEEE PES General Meeting*, Detroit, USA, Jul. 2011, pp. 1-3.

[14] C. W. Potter, E. Grimit, and B. Nijssen, "Potential benefits of a dedicated probabilistic rapid ramp event forecast tool," in *Proceedings of 2009 IEEE PES Power Systems Conference and Exposition*, Seattle, USA, Mar. 2009, pp. 1-5.

[15] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting uncertainty related to ramps of wind power production," in *Proceedings of 2010 European Wind Energy Conference and Exhibition*, Warsaw, Poland, Apr. 2010, pp. 9-17.

[16] Y. Li, P. Musilek, E. Lozowski *et al.*, "Temporal uncertainty of wind ramp predictions using probabilistic forecasting technique," in *Proceedings of 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications (BigDataService)*, Oxford, UK, Mar.-Apr. 2016, pp. 166-173.

[17] M. Cui, D. Ke, Y. Sun *et al.*, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 422-433, Apr. 2015.

[18] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting ramps of wind power production with numerical weather prediction ensembles," *Wind Energy*, vol. 16, no. 1, pp. 51-63, Jan. 2013.

[19] J. Heckenbergerova, P. Musilek, and M. Janata, "Sensitivity analysis of PCA method for wind ramp event detection," in *Proceedings of 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, Florence, Italy, Jun. 2016, pp. 1-4.

[20] T. Ouyang, X. Zha, and L. Qin, "A survey of wind power ramp forecasting," *Energy and Power Engineering*, vol. 5, no. 1, pp. 368-372, Jul. 2013.

[21] N. Friedman, G. Dan, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 1, pp. 131-163, Nov. 1997.

[22] X. Chen, G. Anantha, and X. Lin, "Improving bayesian network structure learning with mutual information-based node ordering in the K2 algorithm," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 628-640, May 2008.

[23] L. Bouchaala, A. Masmoudi, and F. Garouri, "Improving algorithm for structure learning in Bayesian networks using a new implicit score," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5470-5475, Jul. 2010.

[24] D. M. Chickering, D. Heckerman, and C. Meek, "Large-sample learning of bayesian networks is NP-hard," *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 1287-1330, Dec. 2004.

[25] L. Holder, "Greedy search approach of graph mining," in C. Sammut, and G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*, Boston, USA: Springer, Apr. 2017, pp. 483-489.

[26] L. M. Campos, "A scoring function for learning bayesian networks based on mutual information and conditional independence tests," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 2149-2187, Oct. 2006.

[27] D. Heckerman, G. Dan, and D. M. Chickering, "Learning Bayesian networks: the combination of knowledge and statistical data," *Machine Learning*, vol. 20, no. 3, pp. 197-243, Mar. 1994.

[28] P. Walley, "Inferences from multinomial data: learning about a bag of marbles," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 3-34, Feb. 1996.

[29] M. Yang, J. Wang, H. Diao *et al.*, "Interval estimation for conditional failure rates of transmission lines with limited samples," *IEEE Trans-

actions on Smart Grid, vol. 9, no. 4, pp. 2752-2763, Jul. 2018.

[30] J. Bernard, "An introduction to the imprecise Dirichlet model for multinomial data," *International Journal of Approximate Reasoning*, vol. 39, no. 2-3, pp. 123-150, Jun. 2005.

[31] I. Pinelis, "L'Hospital rules for monotonicity and the Wilker-Anglesio inequality," *The American Mathematical Monthly*, vol. 111, no. 10, pp. 905-909, Dec. 2004.

[32] C. Kamath, "Associating weather conditions with ramp events in wind power generation," in *Proceedings of 2011 IEEE PES Power Systems Conference and Exposition*, Phoenix, USA, Mar. 2011, pp. 1-8.

[33] S. Kotsiantis and D. Kanellopoulos, "Discretization techniques: a recent survey," *GESTS International Transactions on Computer Science and Engineering*, vol. 32, no.1, pp. 47-58, Jan. 2006.

[34] A. R. Barron, "Entropy and the central limit theorem," *The Annals of Probability*, vol. 14, no. 1, pp. 336-342, Jun. 1986.

Yuanchun Zhao received the B. E. degree and the M. E. degree in the School of Electrical Engineering, Shandong University, Jinan, China, in 2017 and 2019, respectively. She is currently working in the State Grid Jinan Power Company, Jinan, China. Her research interests include wind generation forecast and artificial intelligence techniques.

Wenli Zhu received her B. E. degree in automation control in 2017 from Hohai University, Nanjing, China. She is currently a Master student in Key Laboratory of Power System Intelligent Dispatch and Control, Shandong University, Jinan, China. Her areas of interest include renewable energy technology and artificial intelligence techniques.

Ming Yang received the B. E. and Ph. D degrees in electrical engineering from Shandong University, Jinan, China, in 2003 and 2009, respectively. He was an exchange Ph.D. student at Energy System Research Center, The University of Texas at Arlington, Arlington, USA, from October 2006 to October 2007. He conducted postdoctoral research at School of Mathematics of Shandong University from July 2009 to July 2011. From November 2015 to October 2016, he was a Visiting Scholar at the Energy Systems Division, Argonne National Laboratory, Argonne, USA. Currently, he is a Professor at Shandong University and an Associate Editor of IEEE Transactions on Industry Applications. His research interests are power system optimal operation and control.

Mengxia Wang received the Ph. D degree in electrical engineering from Shandong University, Jinan, China, in 2011. From July 2011 to July 2013, he conducted postdoctoral research with the Department of Electrical Engineering, Tsinghua University, Beijing, China. He is currently a Lecturer at Shandong University. His research interests are power system security analysis, optimal operation, and control.