IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 1, 2018, accepted November 25, 2018, date of publication December 3, 2018, date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2884539

Research of Wind Power Correlation With Three Different Data Types Based on Mixed Copula

YURONG WANG, (Member, IEEE), AND YANNAN LUO
School of Electrical Engineering, Southeast University, Nanjing 210096, China

Corresponding author: Yurong Wang (wangyurong@seu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51507031.

ABSTRACT With the increasing integration of large-scale wind farms, the stochastic characteristics of wind speed and the coupling relationship of geographically distributed wind farms become ineligible. Correlation investigation of wind farms based on copula theory is able to lay good foundation for further optimization and schedule in power systems. In this paper, three data types, including wind speed, calculated wind power, and actual wind power, are applied to explore the correlation of two geographically close wind farms. After graphical analysis and comparison of the time series, the structural and compositional characteristics of correlation by different data types are investigated based on mixed copula. Moreover, a two-stage filtration method is proposed to evaluate different types of copulas. Finally, taking into account the practical conditions after careful examination of actual wind power dataset, the practical conditions are applied to the calculated wind power dataset. Correlation research based on the adjusted calculated wind power dataset is further explored and revealed that it is more practical and prospective to provide reference for further power system operation with high penetration of wind farms.

INDEX TERMS Actual wind power, calculated wind power, correlation, mixed copula, two-stage filtration method, wind speed.

I. INTRODUCTION

As one of the most important form of clean energy generation, wind power generation grows rapidly in the total installed capacity all around the world in recent years and has become a considerable industry. However, due to the intermittency and stochastic characteristics of wind itself, it is difficult for wind energy to be used as well as traditional energy [1]. At the same time, with the gradual integration of large-scale wind farms into power grid, research on power output correlation between geographically distributed wind farms become more important and has gradually attracted the attention of many scholars [2]–[4].

The study on correlation of wind power output is necessary and useful for power forecast, operation and dispatch [5]–[8]. In the research field of wind power correlation, the copula function has caused extensive research because the marginal distribution of each wind farm can be separated from copula correlation, and the rank correlation index and tail correlation index of copula can better describe the nonlinearity, asymmetry, upper and lower tail relationship between wind farms. Based on copula function, research in [9] drawn a conclusion that as the distance increases, the spatial correlation of wind power output between wind farms decreases gradually, and then, [10] reached a further conclusion about spatio-temporal correlations of clustered wind power. Correlation structure models of wind farms established through copula theory is able to be used in probabilistic load flow analysis [11], optimal power flow [7], and power system uncertainty analysis [12], [13], and finally contribute on dispatch and optimization of power systems.

In order to obtain the comprehensive correlation structure of wind farms, selection and evaluation of copula model is important. Most of the literature used single copula function from Elliptic copula family or Archimedean copula family [14], [15], pair-copula [16] and vine copula [17], [18] are also involved in describing the correlation structure among wind farms. However, the characteristics of asymmetry and tail correlation cannot always be described by these copulas satisfactorily. Mixed copula [19] is a kind of copula function that consists of several different types of copula, it is prospectively able to obtain good characteristics of copula functions since it includes and describes information more

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

comprehensive and accurate. However, researches based on mixed copula is still far from extensive.

In terms of datasets, the correlations of geographically distributed wind farms are usually estimated based on the dataset of wind speed [20], [21] or actual wind power [22], [23]. At the same time, copula-based wind farm correlation analysis using wind speed can be further divided into two application methods [24]: a) Establish copula model using wind speed data directly, and then analyze wind power output based on the obtained copula correlation structure, b) Convert wind speed data to wind power data through the relationship of wind speed and wind power at first, and then establish copula model according to these calculated wind power data. Using wind speed or wind power data in copula-based correlation analysis of wind farms may lead to different results. It is meaningful to analyze and compare wind farm correlation based on different data types, however, few work has been conducted.

This paper investigates on the correlation analysis of wind farms using wind speed, calculated wind power, and actual wind power, respectively. The organization of this article is as follows. In Section II, the relationship and difference among the three datasets of wind farms is explored. In Section III, basic theory of copula function is introduced, and the mixed copula model is employed, moreover, a two-stage filtration method is proposed to evaluate copulas. Then, in the case study in Section IV, mixed copula models are established based on the three types of dataset mentioned above, respectively. Finally, in order to make good use of correlation in practice, the practical conditions from the analysis in actual dataset are investigated and applied to the calculated wind power dataset, and then a satisfactory correlation is obtained to verge to the actual condition. Section V concludes the findings.

II. FEATURE ANALYSIS OF DIFFERENT DATASETS

A. DESCRIPTION OF DATASETS

Wind speed and actual wind power can be obtained from the record collected from wind farms directly, and calculated wind power can be calculated from wind speed. The basic characteristic and relationship of these 3 data types are firstly analyzed in this section.

1) WIND SPEED DATASET

Wind speed data is the most intuitive standard for evaluating wind resources in a region, and also is one of the decisive factors in evaluating the allocation of wind farms.

For practical wind farms, it is necessary to use historical measured wind speed data as a reference to forecast the wind speed of the wind farm and then forecast wind power generation, and further be used to determine the operation plan of the wind farm with the grid conditions.

For the wind farms in the preparatory phase of construction, it is necessary to consider the security risks and economic burden that may imposed on the power system after connecting to power grid. since the actual wind power are not available in this situation, wind speed data collected from the bureau of meteorology can be used to evaluate wind power output.

2) CALCULATED WIND POWER DATASET

Whether it is a practical wind farm or a wind farm under the construction or preparatory stage, the ultimate purpose of collecting wind speed data is to predict the output of wind farm. The typical relationship between wind speed and wind power output is calculated as follows:

$$P_{WTG} = \begin{cases} 0 & 0 \le v \le v_{ci} \quad or\ v_{co} \le v \\ P_{WTG,r}\dfrac{(v - v_{ci})}{(v_r - v_{ci})} & v_{ci} \le v \le v_r \\ P_{WTG,r} & v_r \le v \le v_{co} \end{cases} \tag{1}$$

where v_{ci}, v_r and v_{co} are the cut-in, the rated, and the cut-out wind speed, respectively; $P_{WTG,r}$ is the rated active power output for a single wind turbine.

In a wind farm, the wind turbines are equipped in different places with different wind speed, angle, topography and geomorphology, the wind power output of wind farm is not simply the multiplication of the single output and the number of units. In addition, the outage and maintenance of wind turbine cannot be ignored. As a result, it is difficult to make calculated wind power verge to actual wind power.

In this paper, first of all, the whole wind farm is treated as an entity, and the outline characteristics of practical wind power output is employed to adjust the parameters in (1) to minimize the error [25], as shown in Fig.1. Moreover, practical conditions are investigated and more details are considered for calculated wind power dataset in part C of Section IV.

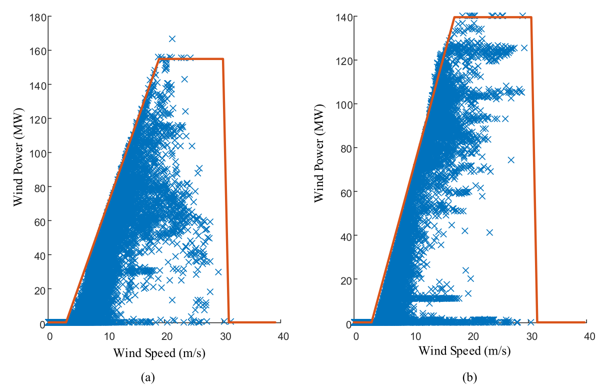

FIGURE 1. Relationship of wind speed dataset and wind power dataset at (a) wind farm 1 and (b) wind farm 2.

3) ACTUAL WIND POWER DATASET

In practical operation, the wind power output is obtained by recording the historical transmission power in the point of common coupling (PCC) of wind farm.

In practice, in order to extract maximum power and optimize the performance, some control strategy is studied to track maximum power points of the wind turbine and improve the power coefficient [26]–[28]. However, although wind power has the priority of dispatch, due to the randomness, intermittency and anti-peaking aspect of wind speed, wind curtailment is an inevitable and necessary solution to satisfy power balance, and this become the main cause of the difference between actual power output and calculated power output [25]. So, comparatively, actual wind power reflects the practical relationship, while calculated wind power of wind farms represents the ideal situation.

As a result, the three dataset might reveal different trends whilst they still have strong relationship.

B. RELATIONSHIP OF THREE TYPES OF DATASET

For two geographically distributed wind farms, the basic relationship between the three types of dataset can be found in Fig.2 graphically, it shows a typical time series of 72-hour segment diagram of wind speed, calculated wind power, and actual wind power. It is obvious that:

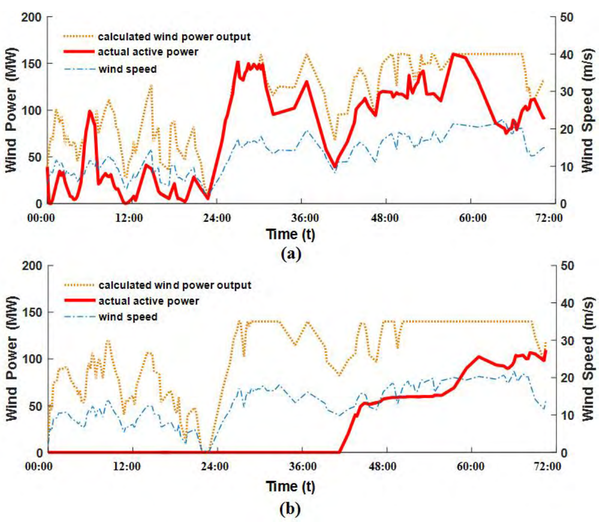

FIGURE 2. 72-hour time series diagram of three data types from (a) wind farm 1 and (b) wind farm 2.

a) The trend of wind speed and calculated wind power datasets are consistent. However, when wind speed is lower than the cut-in value or higher than the cut-out value (the wind power output is 0 in these conditions), or the wind speed is higher than the rated value but smaller than the cut-out value (the wind power output is rated wind power in this condition), the fluctuation of wind speed can no longer influence the output power, therefore, there is still a certain difference between the wind speed curve and the calculated active power curve.
b) Due to the effect of wind curtailment or maintenance, the actual wind power of wind farm is less than the forecasted in most cases.
c) There is also a noteworthy phenomenon that besides the situation that both wind farms are in operation, at some time, one of the wind farm is out of service while the other one is in service.

It can be found that there is consistency in the trend of wind speed, calculated wind power, and actual wind power. However, because of the non-linear relationship between wind speed and calculated power and the existence of wind curtailment, using different types of dataset in copula-based wind farm correlation analysis will lead to different conclusions. Further comparison and investigation on the difference of the copula models can bring helpful information to power system dispatch, operation and control.

C. CORRELATION DESCRIPTION BASED ON DIFFERENT DATASETS

The primary linear correlation between wind farms can be observed by scatter plot. The more scattered points concentrated on the diagonal of the graph, the stronger linear correlation exist between the dataset, respectively.

Although wind speed is the fundamental driver of total wind farm output, due to the nonlinear relationship from (1), the correlation using wind speed reveals different when compared with the correlation using calculated wind power.

In addition, there is a difference in output power between the actual value and the calculated value which is mainly caused by deliberate wind curtailment for either wind farm, this lead to different correlations of the two wind farms based on these two types of datasets.

Fig.3 is the scatter diagram of wind speed, calculated wind power and actual power from two geographically distributed wind farms. The wind speed of the two wind farms depicted in Fig.3(a) is basically distributed along the diagonal line, which represents strong linear correlation, and the stronger the wind speed is, the stronger linear correlation is.

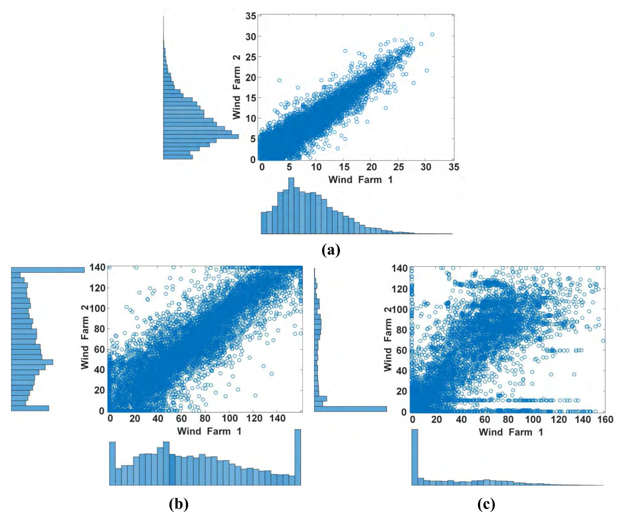

FIGURE 3. Scatter diagrams of (a) wind speed, (b) calculated wind power, and (c) actual wind power of two wind farms.

Fig.3(b) is plotted based on calculated wind power dataset, due to the nonlinearity of (1), the correlation changes, especially in the upper right corner of the figure that there are a large number of data points concentrated around the rated power.

Fig.3(c) shows the relationship based on actual wind power dataset that is collected from the same wind farms with the same time mark in Fig.3(b). Comparatively, there are no data points near the rated power, while a large number of points are distributed along the left and bottom edge of the figure, respectively. This shows abundant evidence that both of the wind farms have great curtailment and cannot operate at full capacity.

Although the linear correlation of the data points in Fig.3(b) and Fig.3(c) is not as obvious as wind speed data, there is still a certain correlation of the wind power. To further investigate the non-linear correlation of wind farms by the three dataset, copula is involved in the next section.

III. THE THEORY OF COPULA FUNCTION

A. COPULA FUNCTION

Based on intuitive analysis and comparison from Section II, in this paper, copula theory is employed to provide a more accurate correlation description to help capture correlation features based on different datasets and better understand whether the probability of a joint wind farm output scenario is underestimated or overestimated.

The greatest advantage of the copula function theory is that it can separate the joint distribution of multivariable into univariate edge distributions and a connection function that connect these edge distributions. The connection function mentioned here is the copula function. It describes the dependence structure between variables that is separated from the univariate distribution rules.

Sklar's theorem [29] states that every multivariate cumulative distribution function $F(x_1, x_2, \cdots, x_N)$ can be expressed in terms of its marginal cumulative distribution function $F_1(x_1), F_2(x_2), \cdots, F_N(x_N)$ and a copula function C that satisfies [30]:

$$F(x_1, x_2, \cdots, x_N) = C(F_1(x_1), F_2(x_2), \cdots, F_N(x_N)) \tag{2}$$

Only when $F_1(x_1), F_2(x_2), \cdots, F_N(x_N)$ are continuous functions, copula function C is uniquely determined, where $F(x_1, x_2, \cdots, x_N)$ is the joint distribution of variables $x_1, x_2, \cdots, x_N$.

B. MEASURES OF COPULA FUNCTIONS

According to Nelson's theorem [30], when the random variables $x_1, x_2, \cdots, x_N$ are strictly monotonously changed, the corresponding copula function remains unchanged, which makes the correlation structure based on copula function reflect the correlation under strict monotone transformation. This characteristic makes copula functions be able to reflect not only linear and symmetric correlations but also nonlinear and asymmetric correlations between variables and have more extensively applicable scope.

Related measures of copula functions mainly include Kendall's rank correlation coefficient τ, Spearman's rank correlation coefficient ρ, and tail correlation coefficient λ^{up} and λ^{lo}. Correlation coefficient mentioned above can be calculated via the following equation:

$$\tau = 4\int_0^1\int_0^1 C(u,v)\,dC(u,v) - 1 \tag{3}$$

$$\rho = 12\int_0^1\int_0^1 C(u,v)\,duv - 3 \tag{4}$$

$$\lambda^{\mathrm{up}} = \lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u} \tag{5}$$

$$\lambda^{\mathrm{lo}} = \lim_{u\to 0}\frac{C(u,u)}{u} \tag{6}$$

where $C(u, v)$ is a two-dimensional copula function.

Since situations that both wind farms having no output or generating rated power at the same time have the greatest impact on the power grid, these two types of situation gain more concern in the research when exploring risks brought about by wind integration to power gird. Hence, it is also necessary to observe the correlation in two extreme situations through the tail correlation coefficient.

In this paper, Euclidean distance criterion, Kendall index, Spearman index and Akaike information criterion (AIC) are used to evaluate the models. However, when several indices are employed at the same time, they may not conclude a unique suggestion for the selection of copula. As a result, it is difficult to select the optimal model clearly.

A two-stage filtration method is proposed to help evaluate copulas in this paper. First, Euclidean, Kendall index, Spearman index and AIC are involved and calculated. For each measure criteria, the worst or best value is given a "×" or a "√" to its corresponding copula model, respectively, and the value that takes the second place is given a "◯". Any copula model with a "×" is dropped. Secondly, the upper tail, lower tail and symmetry feature of the best two models are calculated and examined from a graphical view. Finally, the final optimal copula model selection can be safely concluded.

C. MIXED COPULA FUNCTION AND ITS PARAMETER ESTIMATION

Due to the different characteristics of different types of copula function, the ability to describe asymmetric correlation, upper and lower tails is also different. Using only one type of copula function sometimes is not a practical way to describe the correlation structure, especially when the correlation of variables has apparent upper and lower tails at the same time. Mixed copula can be prospectively employed to highlight the comprehensive characteristics of correlation.

Archimedean copula is a group of copula with good properties, including Gumbel copula and Clayton copula with good ability to describe asymmetric correlation, upper-tail and lower-tail characteristics, respectively, and Frank copula that can describe negative and symmetric correlation between variables. The cumulative distribution functions of the three types of Archimedean copula are expressed as below, respectively:

$$C_{Cl}(u, v; \theta_{Cl}) = \left(\mu^{-\theta_{Cl}} + v^{-\theta_{Cl}} - 1\right)^{-1/\theta_{Cl}} \quad (7)$$

$$C_{Gu}(u, v; \theta_{Gu}) = exp\left(-\left[(-lnu)^{\frac{1}{u}} + (-lnv)^{\frac{1}{v}}\right]^{\theta_{Gu}}\right) \quad (8)$$

$$C_{Fr}(u, v; \theta_{Fr}) = -\frac{1}{\theta_{Fr}} ln\left(1 + \frac{\left(e^{-\theta_{Fr}u} - 1\right)\left(e^{-\theta_{Fr}v} - 1\right)}{\left(e^{-\theta_{Fr}} - 1\right)}\right) \quad (9)$$

where C_{Cl}, C_{Gu}, and C_{Fr} represent Clayton copula function, Gumbel copula function and Frank copula function, respectively, θ_{Cl}, θ_{Gu}, and θ_{Fr} are the parameters of different copula, respectively; $(u, v) \in [0, 1]^2$, and in this paper, u and v are simplified form of marginal cumulative distribution function $F_1(x_1)$ and $F_2(x_2)$, respectively.

The above three copula functions are used synthetically to form a mixed copula to study wind power correlation characteristics using three different datasets in this paper. The mixed copula based on Archimedean copulas can be expressed as

$$C_{mix}(u, v) = \omega_{Cl} C_{Cl}(u, v; \theta_{Cl}) + \omega_{Gu} C_{Gu}(u, v; \theta_{Gu}) + \omega_{Fr} C_{Fr}(u, v; \theta_{Fr}) \quad (10)$$

where ω_{Cl}, ω_{Gu} and ω_{Fr} are weight parameters of Clayton, Gumbel, and Frank copula function, respectively, satisfying $\omega_{Cl} + \omega_{Gu} + \omega_{Fr} = 1$; $C_{mix}(u, v)$ represents mixed copula function.

From (10), there are 6 parameters to be evaluated including 3 weight coefficients and 3 Archimedean copula's parameters in the mixed copula. It is complicated to estimate 6 unknown parameters through maximum likelihood estimation (MLE), so expectation maximization (EM) method is prospectively involved for parameter estimation.

IV. CASE STUDY

In this study, wind speed data and its corresponding actual power output data from two geographically close wind farms in North China in July 2011 is used. The number of data sample for each data type is 10000, respectively. The three datasets are collected with the same time stamp from the same two wind farms.

A. MODELING AND PARAMETER ESTIMATION OF COPULA FUNCTION

1) CALCULATION OF RELATED MEASURES

The scatter plots in Fig.3 used the same dataset with the case study, it intuitively shows the correlation of the three data types from two adjacent wind farms. Pearson coefficient, Kendall coefficient, and Spearman coefficient are used to measure data relevance. Table 1 gives the quantitative comparison of the correlation using three types of dataset. Obviously, the wind speed dataset displays a substantial correlation, while the actual wind power dataset has the weakest correlation.

TABLE 1. Calculation of correlation coefficients.

Dataset	Pearson	Kendall	Spearman
Wind speed	0.9240	0.7429	0.9062
Calculated wind power	0.8885	0.7210	0.8700
Actual wind power	0.7328	0.5689	0.7315

2) PARAMETER ESTIMATION OF MIXED COPULA

Mixed-copula based wind farm correlation analysis mainly consists of 2 steps:

a) Determine the marginal distribution of wind speed (or wind power) from single wind farm. Plot the joint frequency histogram of distributed marginal values.
b) Use mixed-copula function to fit the correlation between marginal distributions. The goal of this step is to determine the weights and parameters for each copula through EM method.

In Fig.3(a), the marginal distribution of wind speed data can be described by Weibull distribution with two parameters, but the marginal distributions in Fig.3(b) and Fig.3(c) are difficult to describe using a simple distribution function. For subsequent analysis and comparison purpose, a nonparametric method to construct the marginal distribution is adopted in this paper.

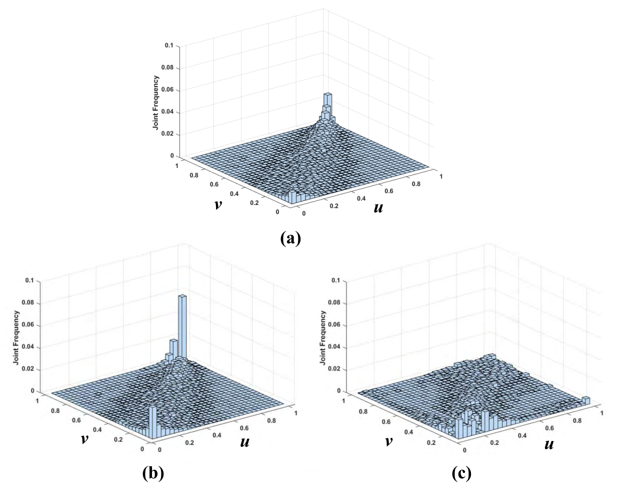

FIGURE 4. Binary frequency histogram of marginal distribution based on (a) wind speed dataset, (b) calculated wind power dataset, and (c) actual wind power dataset.

Fig.4 shows the binary frequency histogram obtained by using the empirical distribution to describe the marginal function and can reach the following conclusions:

a) From Fig.4(a), it can be observed that the correlation structure based on wind speed is basically symmetric, and shows significant upper and lower tail correlation. In addition, the upper-tail correlation is slightly stronger than the lower- tail correlation.
b) Fig.4(b) is based on calculated wind power dataset by (1). Since the wind speed below the cut-in value and

above the rated value cause the wind power output to be 0 or the rated power, respectively, calculated wind power correlation structure has more pronounced upper and lower tail.

c) Fig.4(c) is plotted using actual active power dataset, it shows the actual correlation between wind farms in a certain period. Compared with Fig.4(a) and Fig.4(b), Fig.4(c) are more dispersed, and the overall histogram is shorter. In addition, besides data points concentrated on the upper and lower tails, there are a number of points concentrated near the two coordinate axes, that means only one of the wind farms was generating while the other one is out of service.

TABLE 2. Weights and parameters of mixed copula.

Parameter	Wind speed	Calculated wind power	Actual wind power
ω_{Cl}	0.0015	0.0021	0.1296
ω_{Gu}	**0.8292**	**0.4716**	0.0323
ω_{Fr}	0.1693	**0.5263**	**0.8381**
θ_{Cl}	0.2438	14.3939	16.1233
θ_{Gu}	3.6076	8.4851	1.0001
θ_{Fr}	26.7874	8.8633	11.0879

Table 2 shows the optimal weights and parameters established by EM algorithm according to different types of dataset based on mixed copula established by Clayton, Gumbel, and Frank copula from (10). It can be found from the mixed copula evaluation that:

a) Different dataset resulted in different correlation of wind farms. Mixed copula based on wind speed shows strong upper tail characteristic since Gumbel copula contributes the heaviest weight, mixed copula based on calculated wind power shows upper tail and asymmetry characteristics since Gumbel copula and Frank copula apportion the weights, and mixed copula based on actual wind power only shows asymmetry characteristic since Frank copula has the heaviest weight.

b) In this case study, based on mixed copula, the upper tail characteristic indicated by wind speed and the asymmetry characteristic highlighted by actual wind power are both revealed by calculated wind power. It is prudentially to conclude that, when actual wind power is difficult to obtain, correlation analysis based on calculated wind power can provide more reliable reference for practice than wind speed, and may show conservative evaluation on upper tail to some degree.

Mixed copula is used in this part to analyze the correlation of wind farms for all the data types. To make sure mixed copula can provide credible conclusion, it is necessary to carry out copula modeling verification for different data types.

B. COPULA MODELING ANALYSIS OF DIFFERENT DATASETS

Five types of commonly used copula are considered for copula modeling for the three data types, respectively. They are normal copula, t copula, Clayton copula, Gumbel copula and Frank copula. In this part, the proposed two-stage filtration method is applied and mixed copula is verified.

1) COPULA MODELING ANALYSIS BASED ON WIND SPEED DATASET

In the first stage of the two-stage filtration method, it concludes that Gumbel copula is most suitable for describing the correlation among five commonly used single copula based on the four indices from Table 3, however, mixed copula performs even better, since it gets the most " ○" and " √". So, the figure of Gumbel copula is compared with the figure of mixed copula in the second stage of the proposed method. From Fig.5, it is difficult to figure out which is better because the correlation using Gumbel copula can already describe weaker lower tail and stronger upper tail of dependency structure as mixed copula. This reflected in the mixed copula in Table 2 based on wind speed dataset that Gumbel plays the heaviest weight. Overall, mixed copula outperforms all of the single copulas based on comprehensive criteria estimation.

TABLE 3. Comprehensive evaluation of copula models based on wind speed dataset.

Criterion	normal-Copula	t-Copula	Clayton Copula	Gumbel Copula	Frank Copula	Mixed Copula
Euclidean	1.4305	1.2406	4.8197 ×	0.6611 ○	1.1228	0.6288 √
Kendall	0.6983	0.7184	0.5165 ×	0.7350	0.7401 √	0.7495 ○
Spearman	0.8805	0.8838	0.7007 ×	0.9019 √	0.9154	0.9131 ○
AIC ($\times 10^5$)	-0.8494	-0.8779	-0.6065 ×	-1.0038 ○	-0.8979	-1.0137 √

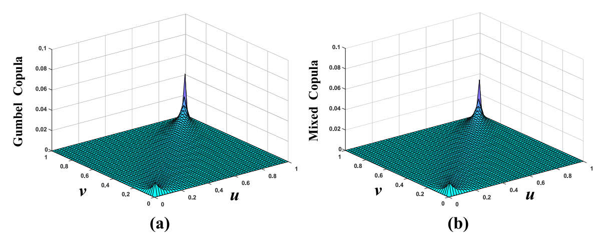

FIGURE 5. Comparison of (a) Gumbel copula model and (b) Mixed copula model based on wind speed dataset.

2) COPULA MODELING ANALYSIS BASED ON CALCULATED WIND POWER DATASET

Based on calculated wind power dataset, Table 4 shows that Frank copula performs best among all other single copulas and mixed copula performs even better by comprehensive parameter comparison. From graphical analysis, Fig.6(a) cannot well express either lower or upper tail correlation, while mixed copula can.

This comparison answered why Frank copula plays the heaviest weight in mixed copula based on calculated wind

TABLE 4. Comprehensive evaluation of copula modeling based on calculated wind power dataset.

Criterion	normal-Copula	t-Copula	Clayton Copula	Gumbel Copula	Frank Copula	Mixed Copula
Euclidean	2.0804	1.4383	3.8602	8.5778 ×	1.0096 ○	0.8659 √
Kendall	0.6432	0.6962	0.5830	0.2315 ×	0.7403 √	0.7360 ○
Spearman	0.8352	0.869	0.7701	0.3377 ×	0.9155 ○	0.8996 √
AIC ($\times 10^5$)	-0.7745	-0.8483	-0.6509	-0.4912 ×	-0.9191 ○	-0.9497 √

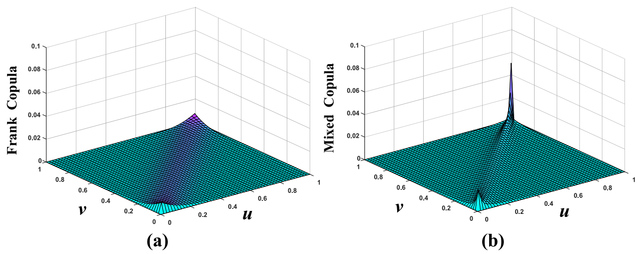

FIGURE 6. Comparison of (a) Frank Copula model and (b) Mixed Copula model based on calculated wind power dataset.

power dataset in Table 2, at the same time, since single copula is not able to describe hybrid characteristics of correlation, mixed copula turns out to be the best and most favorable choice for correlation analysis.

3) COPULA MODELING ANALYSIS BASED ON ACTUAL WIND POWER DATASET

Frank copula outperforms from single copulas in Table 5, while it takes the heaviest weight in mixed copula in Table 2 based on actual wind power dataset.

TABLE 5. Comprehensive evaluation of copula modeling based on actual wind power dataset.

Criterion	normal-Copula	t-Copula	Clayton Copula	Gumbel Copula	Frank Copula	Mixed Copula
Euclidean	3.5554	2.6512	3.9780 ×	3.1029	1.8572 √	1.9514 ○
Kendall	0.4567 ×	0.5653 √	0.4798	0.4890	0.5993 ○	0.6002
Spearman	0.6397 ×	0.7459 √	0.6596	0.6700	0.8003	0.7767 ○
AIC ($\times 10^5$)	-0.6673	-0.7260	-0.6448 ×	-0.6945	-0.7971 √	-0.7868 ○

At the second step of model filtration process, the tail feature of Frank copula and mixed copula are examined as shown in Fig.7. Both of them are weak to reflect the lower tail correlation suggested in Fig.4(c). Comparatively, mixed copula is best since it has the 0.1296 weight of Clayton which contribute the lower tail coefficient $\lambda^{lo} = 0.1241$.

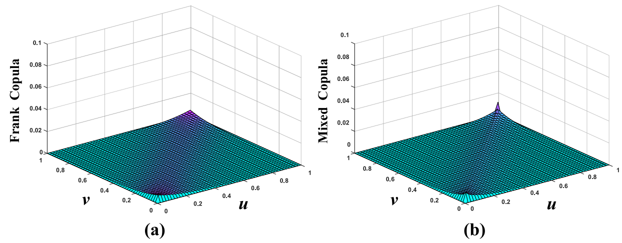

FIGURE 7. Comparison of (a) Frank Copula model and (b) Mixed Copula model based on actual wind power output dataset.

However, the fitting effect is far worse than that of the other two date types through the comparison of Euclidean criterion of mixed copula. Copula function is weak to describe correlation based on actual active power dataset.

In order to figure out the reason of bad fitting effect, details needed to be observed. Fig.8 is the same with Fig.4(c) with another observation angle and a smaller Y-axis range. From the results of correlation analysis based on actual wind power, it can be seen that:

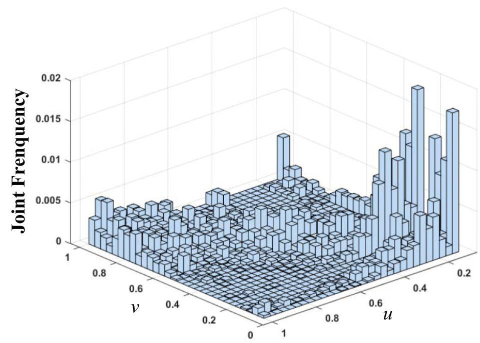

FIGURE 8. Binary frequency histogram of actual wind power output.

a) The frequency histogram based on actual wind power dataset, as shown in Fig.4(c), is shorter than that based on the other two datasets, and the data points are more dispersed.
b) Fig.8 has a thicker lower tail and a less pronounced upper tail. Because of wind curtailment, a large amount of data that should originally been distributed near the upper tail now become smaller and distributed in the middle part, lower tail, and marginal areas.
c) The amount of wind curtailment and power dispatch strategy are main causes of the significant changes in wind farm output.

4) VERIFICATION ANALYSIS OF MIXED COPULA APPLICATION

From the copula modeling selection analysis above, some conclusions can be drawn as follows:

a) It is safe to conclude that mixed copula shows best in describing the correlation structure for all the 3 different datasets.

b) Correlation analysis using wind speed and calculated wind power based on mixed copula is evident, while using actual wind power dataset cannot reflect the correlation of wind farms very well mainly because of equipment maintenance and wind curtailment. To make good use of the correlation effect of wind farms, such as to make better guide to power system operation and dispatch, it is meaningful to establish a more practical correlation structure.

C. FURTHER ANALYSIS OF CORRELATION STRUCTURE CONSIDERING PRACTICAL OPERATION

From the above analysis, the correlation based on actual wind power dataset by mixed copula is weaker than that of calculated wind power dataset although some considerations are applied in Section II for calculated wind power dataset. Therefore, when actual wind power dataset is not available, the correlation study based on calculated wind power should take into account the practical conditions, such as to provide better reference of power system operation.

As a result, in this sub-section, the practical condition is combined to the calculated wind power dataset to better evaluate the correlation for wind farms in practice. After examination of actual wind power dataset, the practical conditions are summarized, and the correlation research based on the calculated wind power dataset with practical adjustment is further explored.

1) EXAMINATION OF ACTUAL WIND POWER DATASET

Fig.2 shows a 72-hour time series of actual wind power of the two wind farms. In the first 40 hours, wind farm 2 is completely out of service while wind farm 1 has the priority to generate power. After that, both wind farms have the same trend as their calculated power output curve, respectively.

According to Fig.2, two extreme ways of wind power dispatch strategy based on historical actual wind power are examined: a) both wind farms generate wind power in strict proportion (Case A) by (11), and b) one of the wind farms (say, wind farm 2) has the priority to generate power (Case B).

$$P_i = (P_{act1} + P_{act2}) \times P_{cali}/(P_{cal1} + P_{cal2}) \tag{11}$$

where P_{act1} (P_{act2}) represents actual power output of wind farm 1 (wind farm 2); respectively;P_{cal1}(P_{cal2}) is calculated power output of wind farm 1 (wind farm 2); P_i (i = 1,2) represents power output of wind farm i under strictly proportional wind dispatch.

Fig.9 shows the results under the two extreme ways of wind dispatch. In Fig.9(a), strictly proportional wind dispatch strategy makes the data concentrate on diagonal section, and this type of wind dispatch shows strong correlation of wind farms, parameters of mixed copula correlation structure is shown in Table 6. The preferential use of wind power from wind farm 2, as shown in Fig.9(b), makes a large number of data points concentrate on the axes, and this type of dispatch is the main cause of weakening the correlation of wind farms.

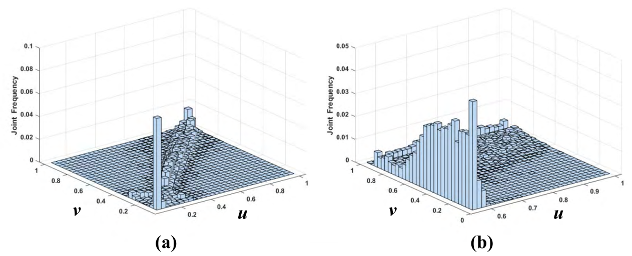

FIGURE 9. Binary frequency histogram of wind power output under (a) strictly proportional wind dispatch strategy and (b) wind farm 2 has the priority to generate power.

The two extreme dispatch strategy indicate two factors in practical conditions: a) Wind dispatch cannot promise strictly proportional generation in practice, so Case A overestimate the output correlation of wind farms. b) Wind dispatch of Case B is one of the reason that weak the actual output correlation and this condition should not be neglected.

2) IMPROVED CORRELATION ANALYSIS CONSIDERING THE PRACTICAL CONDITIONS

To make better use of the intrinsic wind resource and take into account the practical conditions in power systems, the calculated wind power is adjusted based on practical condition, and the correlation analysis is carried out with the intention of providing confidential reference for power system operation and plan.

The power output ratios of two wind farms under different wind speed is analyzed as shown in Fig.10. Power output ratio of wind farm 1 can be calcualted by R_1 in (12):

$$R_1 = \frac{P_{act1}}{P_{act1} + P_{act2}} \tag{12}$$

Points on the upper edges of Fig.10 represent the situation that one of the wind farms has the priority to generate power. By observing the upper edges of Fig.10(a) and Fig.10(b),

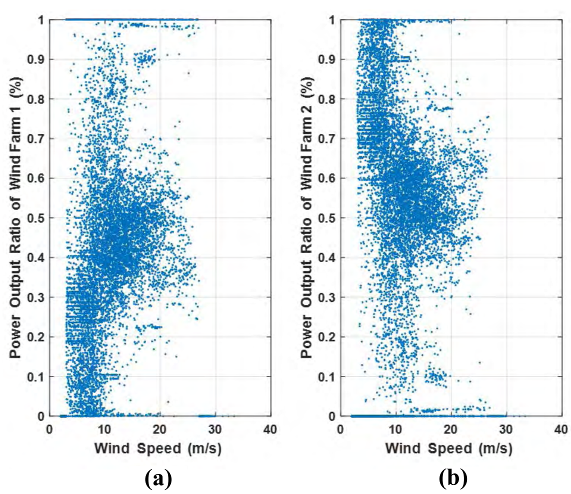

FIGURE 10. Scatter diagram of power output ratio under different wind speed.

it can be found that under Case B dispatch strategy, wind farm 1 is dispatched priorer than wind farm 2. However, under united power generation strategy, when wind speed is below 10m/s, power from wind farm 2 is more likely to be used. In addition, when wind speed is over 10m/s, two wind farms trend to be both dispatched and each take almost half of the required demand, while wind farm 2 takes a little more for most of the time.

In this paper, empirical power dispatch strategy (Case C) is carried out. The value of R_1 is calculated based on actual wind power using (12), for each wind speed interval of 1m/s, in each 10% range of output ratio, the number of R_1 is counted, the occurrence probability is caculated and the average value is used as its typical value. Then for every wind speed interval, the calculated wind power data is statistically adjusted by 10 compents of typical wind power output ratios.

Table 6 shows the correlation structure based on adjusted calculated wind power dataset using mixed copula, and the binary frequency histogram and its mixed copula model are shown in Fig.11.

TABLE 6. Comparison of parameters of mixed copula under different strategies.

Parameter	Calculated wind power	Case A	Case C	Actual wind power
ω_{Cl}	0.0021	**0.4632**	**0.3853**	0.1296
ω_{Gu}	**0.4716**	0.0343	0.0165	0.0323
ω_{Fr}	**0.5263**	**0.5024**	**0.5982**	**0.8381**
θ_{Cl}	14.3939	39.5753	9.7861	16.1233
θ_{Gu}	8.4851	3.8569	1.0001	1.0001
θ_{Fr}	8.8633	18.9409	5.8250	11.0879

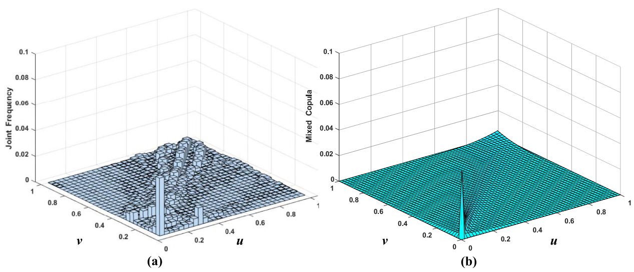

FIGURE 11. (a) Binary frequency histogram and its (b) mixed copula model under empirical power output distribution strategy.

3) COMPARISON OF CORRELATION STRUCTURE UNDER STRICTLY PROPORTIONAL AND EMPIRICAL POWER DISTRIBUTION STRATEGY

Based on mixed copula theory, Table 6 shows parameters estimation results of Case A and Case C by EM method. To make the comparison clear, the parameter estimation results based on calculated wind power and actual wind power are listed again.

From Table 6, Case C shows the confidential correlation of wind farms with practical condition considerations. The weight of Clayton copula in Table 6 and λ^{lo} in Table 7 shows

TABLE 7. Comparison of criteria under different strategies.

Parameter	Calculated wind power	Case A	Case C	Actual wind power
Euclidean	2.8468	4.1042	2.0424	1.9514
λ^{up}	0.4315	0.0275	0	0
λ^{lo}	0.0020	0.4552	0.3590	0.1241

that Case C can dipict the lower tail effect with a proper weight, which is more significant than that of calculated wind power datset, and more conservative than that of Case A.

In addition, the Euclidean distance from Case C is smaller than that of Case A and Calculated wind power dataset and λ^{up} from Case C is near to the characteristics revealed by actual wind power dataset. Overall, study results show Case C indicates a promising copula correlation.

V. CONCLUSION

This paper analyzes and compares the correlation structure between two geographically distributed wind farms using three types of datasets based on mixed copula, and the correlation research based on the calculated wind power dataset with practical adjustment is further explored. Finally, the following conclusions can be safely reached:

a) There are consistency in the trend of wind speed, calculated wind power, and actual wind power from the two geographically close wind farms. However, because of nonlinearity relationship, wind curtailment, and practical dispatch conditions, different datasets lead to different correlation conclusions.

b) Wind speed data has the strongest correlation for two geographically close wind farms, while actual wind power data result to the weakest correlation.

c) Mixed copula behaves better than single copula, and it can improve the description of the correlation structure for all three data types with the proposed two-stage filtration method. As a result, the correlation comparison among the three data types is confidential based on mixed copula analysis.

d) Based on the actual wind power dataset, the practical wind dispatch strategy is examined, and practical conditions are summarized and further applied to calculated wind power dataset. Correlation study based on the adjusted calculated wind power dataset shows it can provide confidential and practical results that is verge to actual condition, such as to further help provide reference for scenario divisions in power system operation with large-scale wind power integration.

REFERENCES

[1] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.

[2] P. Li, X. Guan, J. Wu, and X. Zhou, "Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1594–1605, Oct. 2015.

[3] N. S. Pearre and L. G. Swan, "Spatial and geographic heterogeneity of wind turbine farms for temporally decoupled power output," *Energy*, vol. 145, pp. 417–429, Feb. 2018.
[4] S. Liu, G. Li, H. Xie, and X. Wang, "Correlation characteristic analysis for wind speed in different geographical hierarchies," *Energies*, vol. 10, no. 2, pp. 1–20, 2017.
[5] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *IEEE Trans. Sustain. Energy*, vol. 2, no. 4, pp. 433–442, Oct. 2011.
[6] S. Wen, H. Lan, Q. Fu, D. C. Yu, Y.-Y. Hong, and P. Cheng, "Optimal allocation of energy storage system considering multi-correlated wind farms," *Energies*, vol. 10, no. 5, p. 625, May 2017.
[7] J. Luo, L. Shi, and Y. Ni, "A solution of optimal power flow incorporating wind generation and power grid uncertainties," *IEEE Access*, vol. 6, pp. 19681–19690, Apr. 2018.
[8] M. H. Albadi and E. F. El-Saadany, "Overview of wind power intermittency impacts on power systems," *Electr. Power Syst. Res.*, vol. 80, pp. 627–632, Jun. 2010.
[9] J. B. Jian and S. D. Liu, "Wind speed spatial correlation and optimal power allocation," *Power Syst. Protection Control*, vol. 41, no. 19, pp. 110–117, Oct. 2013.
[10] N. Zhang, C. Kang, Q. Xu, C. Jiang, Z. Chen, and J. Liu, "Modelling and simulating the spatio-temporal correlations of clustered wind power using copula," *J. Elect. Eng. Technol.*, vol. 8, no. 6, pp. 1615–1625, 2013.
[11] B. Li, M. Shahzad, Q. Bing, M. U. Shoukat, M. Shakeel, and E. K. Mohammedsaeed, "Probabilistic computational model for correlated wind farms using copula theory," *IEEE Access*, vol. 6, pp. 14179–14187, Mar. 2018.
[12] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, Feb. 2009.
[13] N. Zhang, C. Kang, C. Singh, and Q. Xia, "Copula based dependent discrete convolution for power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 5204–5205, Nov. 2016.
[14] 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.
[15] C. Sun, Z. Bie, M. Xie, and J. Jiang, "Fuzzy copula model for wind speed correlation and its application in wind curtailment evaluation," *Renew. Energy*, vol. 93, pp. 68–76, Aug. 2016.
[16] Q. Lu, W. Hu, Y. Min, F. Yuan, and Z. Gao, "Wind power uncertainty modeling considering spatial dependence based on pair-copula theory," in *Proc. IEEE PES Gen. Meeting*, National Harbor, MD, USA, Jul. 2014, pp. 1–5.
[17] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 578–589, Jan. 2018.
[18] R. Becker, "Generation of time-coupled wind power infeed scenarios using pair-copula construction," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1298–1306, Jul. 2018.
[19] 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.
[20] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, "Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2239–2247, Aug. 2017.
[21] K. Xie, S. Miao, Y. Xia, Y. Ma, and Y. Li, "A two-stage wind speed model for multiple wind farms considering autocorrelations and cross-correlations," in *Proc. Probabilistic Methods Appl. Power Syst. Conf.*, Beijing, China, Oct. 2016, pp. 1–6.
[22] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1316–1324, May 2014.
[23] M. Xie, J. Xiong, S. Ke, and M. Liu, "Two-stage compensation algorithm for dynamic economic dispatching considering copula correlation of multiwind farms generation," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 763–771, Apr. 2017.
[24] Y. R. Wang, Y. N. Luo, and Z. Y. Dai, "Study and comparison of wind power correlation using two types of dataset," in *Proc. IEEE PES Gen. Meeting*, Portland, OR, USA, 2018.
[25] Y. Zhao, L. Ye, W. Wang, H. Sun, Y. Ju, and Y. Tang, "Data-driven correction approach to refine power curve of wind farm under wind curtailment," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 95–105, Jan. 2018.
[26] A. B. Asghar and X. Liu, "Estimation of wind turbine power coefficient by adaptive neuro-fuzzy methodology," *Neurocomputing*, vol. 238, pp. 227–233, May 2017.
[27] H. A. Heikal, O. S. M. Abu-Elyazeed, M. A. A. Nawar, Y. A. Attai, and M. M. S. Mohamed, "On the actual power coefficient by theoretical developing of the diffuser flange of wind-lens turbine," *Renew. Energy*, vol. 125, pp. 295–305, Sep. 2018.
[28] J. Dai, D. Liu, L. Wen, and X. Long, "Research on power coefficient of wind turbines based on SCADA data," *Renew. Energy*, vol. 86, pp. 206–215, Feb. 2016.
[29] A. Sklar, *Fonctions de Répartition an Dimensions et Leurs Marges*. Paris, France: Univ. Paris Press, 1959, pp. 229–231.
[30] R. B. Nelson, *An Introduction to Copulas*. New York, NY, USA: Springer, 2006, pp. 109–115.

YURONG WANG (S'08–M'12) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, Jiangsu, China, in 2005, 2008, and 2012, respectively, and the joint Ph.D. degree from the University of Tennessee, Knoxville, TN, USA.

She is a Lecturer with the School of Electrical Engineering, Southeast University. Her research interests include time series analysis, renewable energy, power system optimization, and transient stability analysis.

YANNAN LUO was born in Nanjing, China, in 1994. She received the B.S. degree in electrical engineering from Northeast Dianli University, Jilin, China, in 2016. She is currently pursuing the M.S. degree in electrical engineering with Southeast University, Nanjing.

Her main research interests include the uncertainty analysis and optimization of power system incorporating wind farms.