IOP Conf. Series: Materials Science and Engineering **366** (2018) 012019 doi:10.1088/1757-899X/366/1/012019

Wind farm's clustering equivalent method based on mixed-copula function considering wind speed correlation of complex terrain

Wang Chenggen[1], Zhou Qian[1] and Bu Jing[2]

[1]Power Grid Technology Department State Grid Jiangsu Electric Power Company Research Institute, Nanjing, 211100, China
[2]School of Automation, Nanjing University of Science and Technology, Nanjing, 210094, China

Email: 535462103@qq.com

Abstract. Large-scale grid connected wind farms are the main development trend of wind power in China. In the planning and design stage of wind farms, a simplified equivalent model is often used in wind farm equivalents. It makes the wind farm simply equivalent to a single wind turbine, neglecting the change of wind speed correlation caused by the terrain effect. This makes a large error in estimating wind power. In order to solve this problem, a method based on wind speed correlation is proposed to cluster the wind farm of complex terrain. By using the mixed Copula function with tail characteristics, a wind field correlation model with complex terrain features is built. Based on that, EM function is used to identify the Copula function parameters, and the equivalent clustering method of wind farms is built. Finally, the simulation analysis shows that the complex terrain wind farm clustering equivalent method considering wind speed correlation is verified.

1. Introduction

As an intermittent energy, the randomness and volatility of wind power lead to large errors in its prediction. Meanwhile, because of the existence of Meteorology inertia within a certain area, the wind speed of different wind turbines has a significant correlation. This correlation will be converted to the correlation between wind power[1], which makes the overall characteristics of the wind farm different from the characteristic of the single wind turbine[2], which will lead to a large error in the prediction of the output of the wind farm.

In the equivalent modeling of wind farm, the wind speed of the same type of wind turbine is also different because of the difference of the spatial factors inside the wind farm. The wind speed correlation caused by the wake effect and the topography and geomorphology leads to a large degree of correlation between wind farms. Based on the correlation characteristics of wind speed, the wind turbines are equivalent to simplify the calculation. Many researches have been studied on the multi-machine equivalence or single-machine equivalence in wind farms. References [3, 4] improve the prediction accuracy by improving the prediction method of wind power, but the prediction error can only be controlled at 10%-20%, and the prediction accuracy cannot meet the power balance accurately. Reference [5] optimizes the same row of wind turbines equivalents as one machine on the basis of this. Although the optimization method simplifies the wind farm model, the difference of wind speed distribution of the actual wind farm has great influence on the output of the wind turbine, and there is a large computational error for large wind farms. According to Reference [6], a general statistical rule of cross correlation between adjacent units is obtained based on actual wind farm's data for common rectangular and one wind turbine arrangement modes in wind farms. Other traditional probabilistic clustering methods are based on the K-MEANS algorithm, and the state variable matrix of the wind turbine is used as a classification criterion for clustering. Due to the fluctuation of wind speed, the change of classification and partition is greatly increased due to the fluctuation of wind speed. At the same time, the correlation of wind speed is not considered.

In order to solve this problem, this paper presents a method of clustering equal value of wind speed correlation of wind farm considering the influence of complex terrain. It uses a Mixed-Copula function considering the tail effect. Then based on the wind speed sample data, it uses EM method of mixed Copula function parameter identification. Finally, it uses WAsP software to construct simulation model for algorithm verification.

2. Wind speed correlation model based on Copula theory

2.1 Wind speed correlation and distribution characteristics

The probability distribution of wind speed can be described by the Weibull distribution [7], which expression is shown as (1):

$$P(v)=\frac{k}{v}(\frac{v}{c})^{k-1}\exp[-(\frac{v}{c})^{k}] \tag{1}$$

Where v represents wind speed, k and c represents the shape and scale parameters, which could be calculated by measured data of wind speed.

Tail characteristics are used to describe the degree of interaction between variables when extremely small probability events occur, and the upper and lower tail variables are used to indicate the correlation characteristics of variables tend to be minimum and maximum. Thus the tail characteristics can be used to analyze the correlation between the samples when the wind speed is minimization and maximum.

When the wind speed is close to the minimum and maximum, it is possible to analyze whether the two variables are asymptotically independent or independent by introducing tail characteristics, and it is also a theoretical support for wind turbine convergence based on wind speed. The tail characteristics are common in the field of financial analysis to describe the degree of interaction between variables when extremely small probability events occur[8]. The correlation features of the minimum and maximum variables are represented by the upper and lower tails respectively.

The joint distribution function of random vector (X, Y) is F, and the edge distribution functions are F_1 and F_2 respectively, and the tail correlation is expressed by the limit. The characteristics of the upper and lower tails are as follows:

$$\lim_{u\to 1^{-1}} P(F_1(X)>u\,|\,F_2(Y)>u)=\lambda_U \tag{2}$$

$$\lim_{u\to 0^{+}} P(F_1(X)<u\,|\,F_2(Y)<u)=\lambda_L \tag{3}$$

Where, λ_U and λ_L represent the upper and tail correlation coefficients and the lower tail correlation coefficients respectively, respectively, which are represented by the generating element of the Copula function, which are expressed as follows:

$$\lambda_L=\lim_{u\to 0^{+}}\frac{\varphi^{[-1]}(2\varphi(u))}{u}=\lim_{x\to\infty}\frac{\varphi^{[-1]}(2x)}{\varphi^{[-1]}(x)} \tag{4}$$

$$\lambda_U = \lim_{u \to 1^-} \frac{1-\varphi^{[-1]}(2\varphi(u))}{1-u} = 2 - \lim_{x \to \infty} \frac{1-\varphi^{[-1]}(2x)}{1-\varphi^{[-1]}(x)} \tag{5}$$

2.2 *Copula correlation evaluation index considering wind speed correlation*

In order to facilitate describing nonlinear and non normal distribution of random variables, and considering the tail characteristics of correlation distribution, two rank correlation coefficient expressions are introduced. They are Kendall rank correlation coefficient and Spearman rank correlation coefficient.

1) Kendall rank correlation coefficient

In order to represent the nonlinear parameters and non normal distribution, we use τ to represent the Kendall rank correlation coefficient, u, V represent two random variables to determine the correlation. The expression of Kendall rank correlation coefficient is as follows:

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 \tag{6}$$

Where, in the range of τ is [-1,1]. Less than zero of negative correlation degree, greater than zero to represent the positive correlation.

2) Spearman rank correlation coefficient

The expression of the Spearman rank correlation coefficient is similar to the linear correlation coefficient. The specific difference is that one variable is the rank of the sample and the other is the coordinate of the sample. As to the correlation of the expression variables, it is clear that the Spearman rank correlation is more widely applicable and is no longer limited to the nonparametric test of the overall distribution. The expression of the Spearman rank correlation coefficient is as follows:

$$\rho = 12\int_0^1\int_0^1 uvdC(u,v) - 3 \tag{7}$$

Where, ρ is the same Spearman rank correlation coefficient, u and v represent two random variables to determine the correlation.

2.3 *Mixed-Copula function model of wind speed correlation with tail characteristics*

According to the difference of the fluctuation size, the wind speed fluctuation range is divided into three levels which are high, medium and low [9]. Among them, the high and low wind speed level are shown by the tail characteristics, and the wind speed level in the middle is expressed with the symmetry. In the Copula function, the Gumbel-Copula function and the Clayton-Copula function can be used to characterize the tail features, and the Frank-Copula function is selected to characterize the symmetry. Symmetry can well reflect the distribution characteristics of the middle level of the wind speed, but there is a great difference between the wind speed's tail characteristics. So single Copula function cannot reflect this characteristic synthetically. Therefore, the Mixed-Copula function is selected to fit the wind speed. The three Copula functions are defined as follows [10]:

1) Gumbel-Copula function

The Gumbel-Copula function is more sensitive to the change of the upper end to deviate from the farther and larger data of the center. The function expression is shown as follows:

$$C_2(u,v,\theta_i) = \exp\left\{-\left[(-lnu)^{\theta_i} + (-lnv)^{\theta_i}\right]^{\frac{1}{\theta_i}}\right\} \tag{8}$$

The corresponding generating element is $(-\ln u)^{\vartheta}$

2) Frank-Copula function

The Frank-Copula function can reflect the characteristics of the symmetrical tail. That is, the fitting effect is better for the speed segment of the stroke. The function expression is shown as follows:

$$C_1\left(u,v,\theta_i\right) = -\frac{1}{\theta_i}\left[1+\frac{\left(e^{-\theta_i u}-1\right)\left(e^{-\theta_i v}-1\right)}{e^{-\theta_i}-1}\right] \tag{9}$$

The corresponding generating element is $-ln\frac{\left(e^{-\theta u}-1\right)}{\left(e^{-\theta}-1\right)}$.

3) Clayton-Copula function

It is very sensitive to the change of the lower tail. So it is better to deviate from the farther and larger data from the center. The function expression is shown as follows.

$$C_3\left(u,v,\theta_i\right) = \max\left[\left(u^{-\theta_i}+v^{-\theta_i}-1\right)^{-\frac{1}{\theta_i}},0\right] \tag{10}$$

The corresponding generating element is $\frac{1}{\theta}\left(t^{-\theta}-1\right)$.

The Mixed-Copula function expression to construct wind speed model considering the tail characteristics is as shown in (11):

$$\begin{aligned} C_M\left(u,v\right) &= \sum_i^K \lambda_i C_i\left(u,v,\theta_i\right) \\ &= \lambda_1 C_F\left(u,v,\theta_1\right)+\lambda_2 C_G\left(u,v,\theta_2\right)+\lambda_3 C_C\left(u,v,\theta_3\right) \end{aligned} \tag{11}$$

Where, C_i represents different Copula function, λ_i represents weight coefficient of Frank-Copula function, Gumbel-Copula and Clayton-Copula function in the Mixed-Copula function; θ_i represents the correlation coefficient between random variables; u and v represent uniform distribution between the two random variables within [0,1].

3. Wind turbine clustering method considering wind speed correlation

The specific flow of the wind turbine clustering method considering wind speed correlation is shown as Figure 1. Firstly, we use historical wind speed data collected by wind-measuring points in the wind farm to construct Mixed-Copula function as Equ.(11). Secondly, we use EM (expectation maximization) method to estimate parameters of Mixed-Copula function. And then we use inverse transformation of its probability density function to fit wind speed. Finally, we use their rank correlation coefficient to judge the degree of relevance and construct wind farm clustering equivalent model.

In this Mixed-Copula function, λ_i and θ_i are the parameters to be estimated. We use EM method which is the non-hierarchical clustering algorithm to estimate. The correlation coefficient of weight coefficient reflects the characteristics that each tail of Copula function and reflect the relationship of the node velocity, joint probability the distribution of final node velocity of two adjacent columns.

And we can use the rank correlation test on the wind speed of any two wind turbine installation points in the wind farm, the strong correlation is determined by the confidence level. If the correlation of two machines is strong correlation, they could be equivalent. The output of the equivalent unit can be obtained by the conversion of the capacity and wind speed of these two machines.

4. Simulation analysis

The wind farm model was established by WAsP software as an example, the wind farm has 24 Bonus 2MW wind turbines of SIEMENS with height of 60m which are the same types. The arrangement of these 24 wind machines is as shown in Figure 2. Where, the red line represents the contour, blue line and the gray line represents the roughness change. And its wind's rose map is shown in Figure 3.

We use WAsP software to generate wind speed prediction data considering wind speed correlation, which fully considers the wake effect caused by wind turbine arrangement, the surface roughness and the influence of shelter. Table 1 lists a row of four matches' related data.

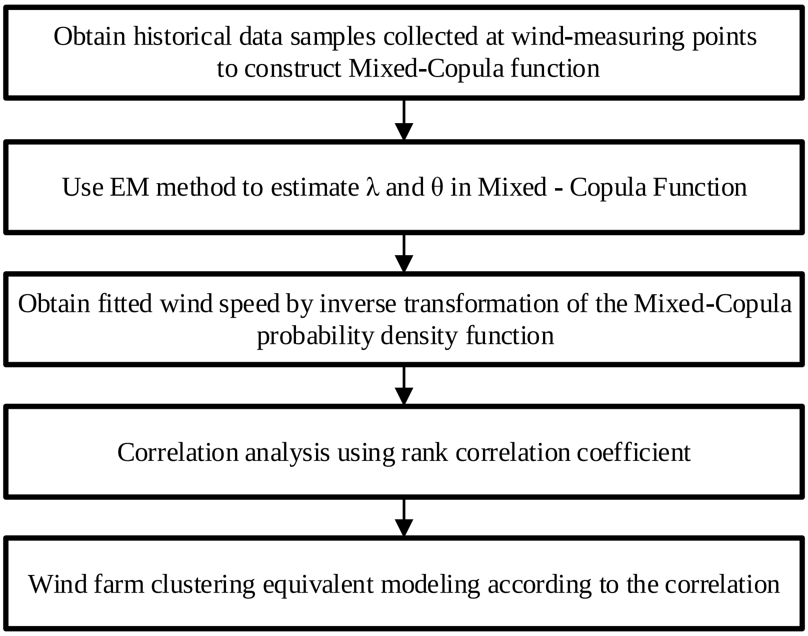

Figure 1. Flow chart of wind turbine clustering method considering wind speed correlation

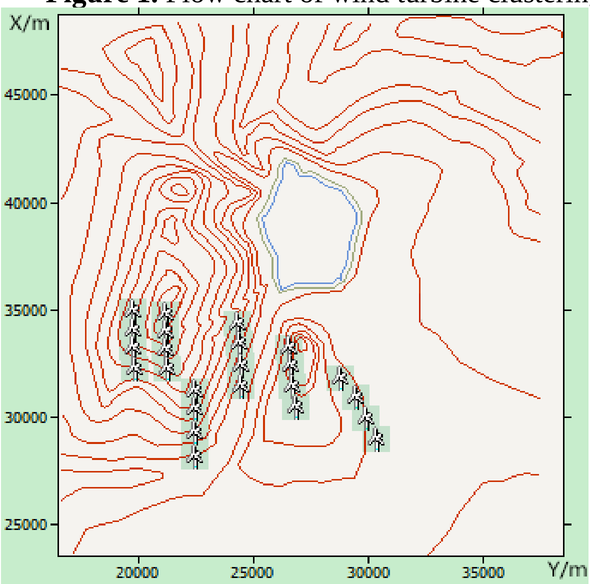

Figure 2. Distribution of Wind Turbines in a complex terrain

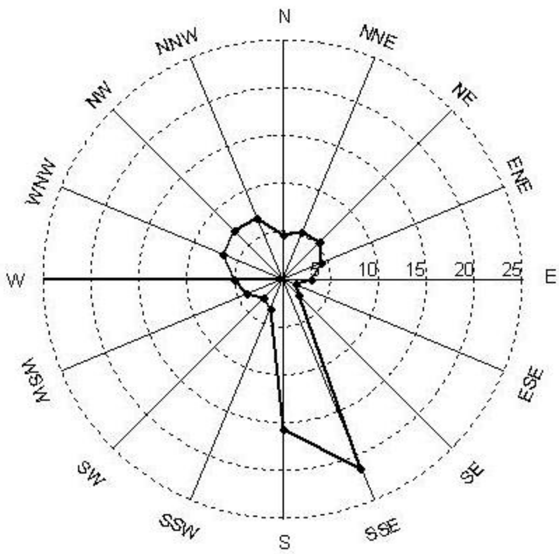

Figure 3. Wind direction distribution roses of the wind farm

TABLE 1. Wind speed prediction data considering wind speed correlation

Position	Coordinate[m]	Wake loss[%]	Wind speed[m/s]	
			Average	Extreme
1	(24298.8,33871.5)	1.60	7.36	10.88
2	(24342.9,32943.9)	1.87	7.25	10.71
3	(24387.1,31972.2)	2.23	6.88	10.26
4	(24431.3,31000.5)	1.82	7.07	10.50

First, according to the historical data of wind speed, we estimate the two parameter of Weibull, which are the shape and the scale parameter. Through these two parameters, we can generate the wind speed distribution scatter diagram fitting Weibull distribution, as shown in Figure 4. And then according to the historical data of wind speed, we use Mixed-Copula function to construct their distribution, which is shown in Figure 5. Through fitting results, we can see that by constructing the

Mixed-Copula function to fit the joint probability distribution of wind speed, the general distribution characteristic is similar to the Weibull shown as Figure 4. The Weibull fitting curve is fitted by the Mixed-Copula function as shown in Figure 6.

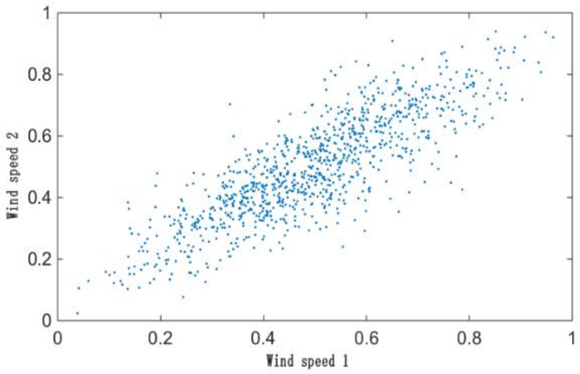

Figure 4. Scatter plot of Binary wind velocity meets Weibull distribution

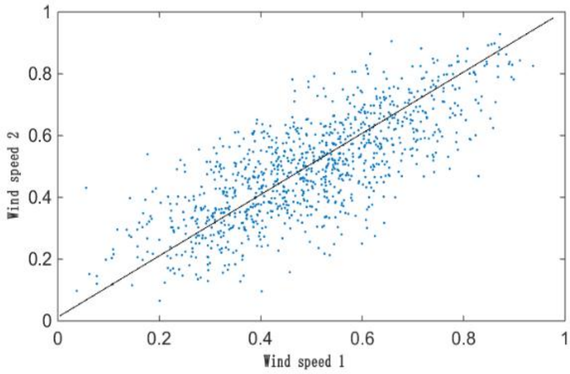

Figure 5. Scatter plot of Binary wind velocity meets Mixed-Copula distribution

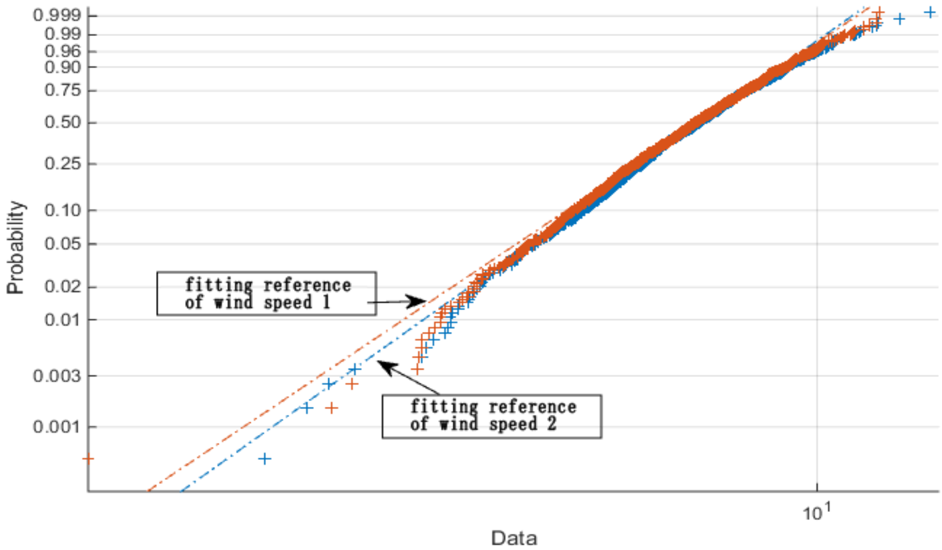

Figure 6. Mixed-Copula function meets the wind velocity

Using the above method to estimate the parameters of λ and θ, the results of which are shown in Table 2. Taking the relevant parameters of the first and second wind turbines as an example, according to the expression, we can make correlation analysis on the wind speed samples of any two wind turbines. A total of 276 sets, and then the equivalent level of wind farms is over 0.99, and the results are aggregated as shown in Table 3.

According to the correlation results, the wind farm is polymerized. After that, the number of the machines changed from 24 to 12 equivalent units. In the above experimental environment, based on the wind speed data obtained from the example and the results of the correlation polymerization, the output of the wind farm considering the correlation of wind speed is calculated. According to each machine using the MPPT control, the traditional single row equivalent simulation and equivalent method considering wind speed correlation clustering results were compared, found that the active power output size differences, through the comparison the cluster correlation of wind speed equivalence than single row equivalent more detailed model.

Figure 7 shows the comparison of the output of wind farm calculated by different equivalent methods. Table 4 shows the accurate value in Figure 7. Through this figure, we can clearly see that the output of which the non-equivalence state is additive, and the clustering equivalent represents 12 machines. The single row equivalent represents 6 machines. The Figure 7 can be seen both in average

IOP Conf. Series: Materials Science and Engineering **366** (2018) 012019 doi:10.1088/1757-899X/366/1/012019

and extreme wind speed, relative aggregated than single row equivalent active power output of the wind farm output closer to the MPPT model, so the clustering method according to the correlation of wind speed is more effective than the traditional single row equivalent.

TABLE 2. Mixed-Copula function related parameters

Paraments	Gumbel-Copula	Clayton-Copula	Frank-Copula
λ	0.339	0.357	0.304
θ	1.524	7.236	9.146

TABLE 3. The wind turbines polymerization results by the rank correlation coefficient

	Wind turbines' combination	Correlation level
Group 1	（4,14,19,23）	0.996
Group 2	（11,13,20）	0.994
Group 3	（7,10,17）	0.995
Group 4	（1,22）	0.997
Group 5	（3,5）	0.998
Group 6	（6,12）	0.997
Group 7	（8,9）	0.991
Group 8	（18,21）	0.992
Not clustering	16，24，2，15	

TABLE 4. Output comparison between average wind and extreme wind speed

Wind Speed	Active power (MW)		
	Single row equivalence （6 machines）	Clustering equivalence （12 machines）	Non-equivalence （24 machines）
Average	11.343	13.950	16.326
Extreme	32.649	41.333	46.888

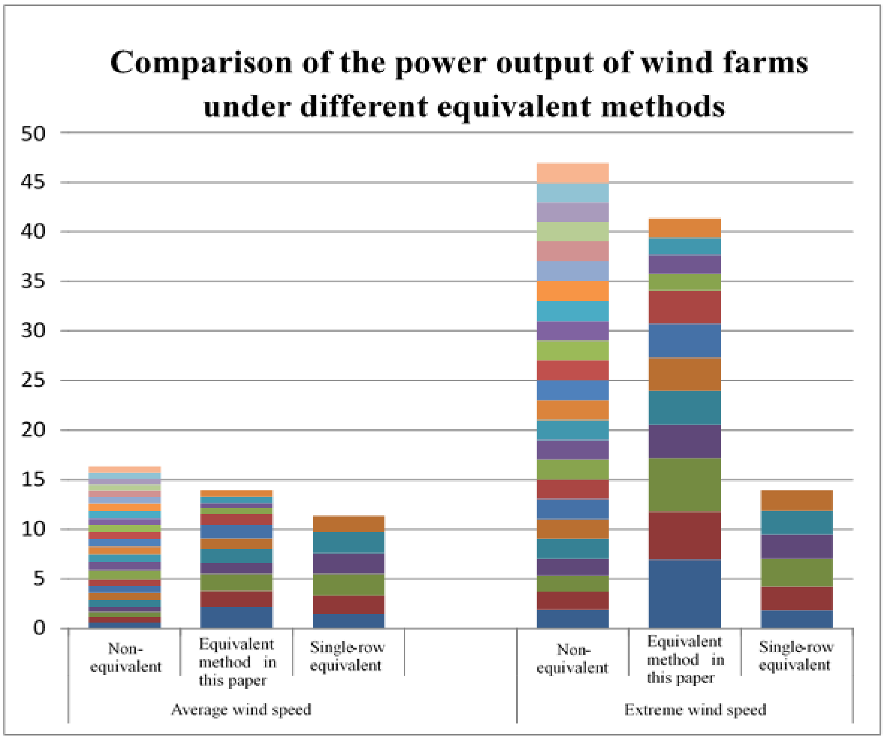

Figure 7. Output comparison between different ways of equivalences

5. Conclusion

In this paper, a method based on wind speed correlation is proposed to cluster the wind farm of complex terrain. The Copula theory is introduced to analyze the correlation of wind farms. The Mixed-Copula function is used to achieve equivalent polymerization by using the correlation of wind speed.

This method is used to test the hypothesis by rank correlation coefficient, to judge the correlation degree between them, and to make equivalent output with higher value, so as to achieve equivalent simplification of wind farms. The wind speed and extreme wind speed under the same climate environment are compared with the traditional single row equivalent method, the unequal value detailed model and the wind farm output of this method. Simulation validates the effectiveness of this method.

Acknowledgment

This work is supported by the Science and Technology Project of State Grid Jiangsu Electric Power Company (J2017046).

References

[1] Papaefthymiou G, Schavemaker P H, Sluis L V D, et al. Integration of stochastic generation in power systems[J]. *International Journal of Electrical Power & Energy Systems*, 2006, 28(9):655-667.
[2] Wu Yuwei, Shi Bin, Zhu Xueqiong, et al. Research on the correlation of wind farm output power based on M-Copula theory [J] .*Electronic Applications*, 2015 (6): 72-78.
[3] Ji Feng, Cai Xingguo, Wang Jun.Analysis of Wind Power Correlation Based on Hybrid Copula Function [J]. *Automation of Electric Power Systems*, 2014, 38 (2): 1-5..
[4] Wang Wei. Analysis on Factors and Routes of Financing of New Energy Industry in China [D]. *China University of Mining and Technology*, 2014.
[5] Zhang Li, Liu Junyong, Liu Youbo, et al. An optimal rotational reserve capacity under wind speed dependence [J]. *Power System Technology*, 2014, 38 (12): 3412-3417.
[6] Yang Jingyi, Li Zheng. Analysis on output correlation of fan with different arrangement [J]. *China Industrial Control Computer*, 2014 (8): 153-155.
[7] Ge Jiangbei, Zhou Ming, Li Gengyin, et al. A review on modeling of large-scale wind farms [J]. *Power System Protection and Control*, 2013 (17): 146-153 (in Chinese).
[8] Louie H. Evaluation of bivariate Archimedean and elliptical copulas to model wind power dependency structures[J]. *Wind Energy*, 2014, 17(2): 225–240.
[9] Liu Zhuo, Li Yanqing, Qi Tingting. The study on power system stability and stable operation under wind convergence effect [C] .*Yunan Electric Power Technology Forum*. 2015.
[10] Cai Fei. Establishment of Wind Speed Dependent Structure Modeling in Wind Farms with Correlation and Its Impact on Transmission and Distribution System Reliability [D]. *Shanghai Jiaotong University*, 2014.