Multi-wind Farm Output Correlation Model Based on Clayton-Copula Function

Shan Jinning¹, Chen Xiaodong², Wang Chenqi¹, Xing Guiyang³*, Wang Baoshi³
¹State Grid Fuxin Electric Power Supply Company, Fuxin, Liaoning, 123000, China
²State Grid Liaoning Electric Power Supply Co., Ltd, Shenyang, Liaoning, 110004, China
³Shenyang Institute of Engineering, Shenyang, Liaoning, 110136, China
*Corresponding author’s e-mail: 574452296@qq.com

Abstract. Because multiple wind farms are connected to the grid at the same time and the total amount of energy in the same wind zone is limited, there is a strong correlation between wind farms with similar geographical locations. Neglecting this correlation can lead to a large difference between wind power analysis and actual operation, which in turn leads to a series of adverse consequences. In this paper, we use nuclear density estimation to establish the edge distribution of wind power output, compare and analyze various Copula functions based on correlation parameters and entropy weight optimization theory. The simulation analysis results show that the Clayton-Copula function is the best correlation function, which can describe the tail part of the random time series more accurately.

1. Introduction
According to the existing researches, due to the similarity of regions and climates, wind farms in the same wind zone and geographically similar in the same region often have strong correlation characteristics between wind powers. Ignoring this correlation in wind power analysis often creates a greater risk of a difference between wind power analysis and actuality. Therefore, in the context of large-scale wind farm integration, it is extremely necessary to consider the correlation of wind power output.

Constructing a model that considers the correlation between multiple wind farms is necessary and effective for developing a reasonable dispatch plan and realizing the safe and economic operation of the power system. The study of the relationship between two random variables establishes a link between two random variables that was originated in the field of finance in order to find the correlation between two variables. Later, it was used by scholars in other fields. For the output correlation of two wind farms with the same wind source, the research results can also be used to analyze and establish two wind farm output correlation models to solve the problem of wind farm output correlation.

Traditional research has focused on the random nature of wind farms. When studying the randomness of wind power output, there are mainly two methods of probability distribution modeling and random sequence modeling, but these two methods have higher precision under the premise that the random sequence obeys the normal distribution, but later scholars proved that the wind power output is largely nonlinear and non-normally distributed. Therefore, the traditional method has a large deviation from the actual operation in the description of wind power output correlation. Therefore,
constructing a mathematical model that can accurately describe the correlation of wind power output is of great significance to the economic, safe and stable operation of the power grid.

Reference [4], the idea of using normal-copula distribution to describe the temporal correlation of prediction error is proposed. The conditional normal-copula model is used to fit the prediction error distribution under different prediction conditions, and the joint probability density distribution of high-dimensional variables is realized. In addition, the conservative degree of wind power distribution range is reduced. This paper applies wind power correlation theory to wind power load forecasting. It is also mentioned that the conditional normal-copula model is still multivariate normal distribution in nature, so there is systematic error in fitting the high-dimensional random variables of the actual asymmetric distribution. Reference [5] proposed the versatile-copula distribution and used it to model the wind power cross-time correlation. Based on this, a joint distribution function considering the time correlation of wind power sequence is designed, which realizes the wind power correlation considering the time scale. However, this article only studies the simulation of a single copula. It is not compared with other methods and can not prove the pros and cons of the algorithm. Reference [6] focuses on the study of tail-related properties between wind farms, and uses the Gumbel-Copula function to establish a probability density model for wind farms output. In addition to focusing on the output data curve, this method also analyzes the tail correlation between data. The above literature does not verify the fitness of the model and the actual operation of the wind farm. When selecting the optimal function, the indicators considered are relatively simple and the above research pays too much attention to the application of correlation, and the relevance modeling process is not paid enough attention.

In this paper, the relationship between wind speed and output is obtained by selecting wind farms with similar areas. The edge distribution is established by nuclear density estimation, and the correlation between wind farm output is calculated and analyzed by Copula function. Through the correlation index and weight optimization theory, the optimal evaluation is carried out among the five Copula functions of Normal-Copula, t-Copula, Gumbel-Copula, Clayton-Copula and Frank-Copula, and the correlation between the two wind farms can be described[7]. Therefore, the best function model that can describe the correlation between the two wind farms is selected.

2. Copula function and entropy weight optimization theory

2.1. Copula function

The correlation study of two random variables was mentioned very early in economics. The Copula function was born here. It was proposed by SKlar in 1959[8]. He believed that a random variable can be composed of a Copula function and multiple edges. The cumulative distribution function composition, which means that the Copula function is actually a link edge distribution function and a variable joint distribution function, so some people call it a link function.

Sklar’s theorem: Let the joint fraction of random variables, \( X_1, X_2, \ldots, X_n \), be \( H \). the edge distribution be \( F_i(X_i), F_2(X_2), \ldots, F_n(X_n) \), then there is a Copula function \( C \), so that \( H(x_1, x_2, \ldots, x_n) = C[F_1(x_1), F_2(x_2), \ldots, F_n(x_n)] \)

If the function \( H \) is a continuous function, the Copula function \( C \) is uniquely determined. It can be seen from this theorem that the multi-wind farm combined output distribution can be split into a form in which the edge distribution is multiplied by the Copula function. Its advantage is that variables are not required to have the same edge distribution, any edge distribution can be connected into a joint distribution by Copula function, and the information of random sequence is distributed in the edge, so in the process of conversion through Copula function, almost there is no data distortion. When using the Copula function, the following steps are mainly included to determine the edge distribution of the variable, and then the parameters of the Copula function need to be calculated, later, the appropriate Copula function is selected according to the appropriate selection index, and the distribution is finally established according to the correlation distribution obtained.
There are five commonly used types of Copula functions, namely Normal-Copula, t-Copula, Gumbel-Copula, Clayton-Copula, and Frank-Copula. They have their own differences and advantages in describing the symmetry and tail distribution characteristics of random variables.

2.2. Entropy weight optimization theory
The entropy weight optimization theory is needed to select the appropriate Copula function for the power correlation study. The concept of entropy was proposed by the famous German physicist Clausius in 1864, and then the concept was gradually generalized over time. It was not until 1948 that Shennong once again proposed the concept of information entropy to make the meaning of entropy more explicit. Understanding the information entropy from the physical meaning is actually the relative rate of change of some related data describing the sample. The closer the coefficient is to 1, the closer the coefficient is. The closer the entropy is to 1, the closer it is to the target; conversely, the closer the entropy is to 0, the further away from the target. In the same way, the actual value is compared with the ideal value, the faster the ratio changes, the smaller the entropy is, the greater the weight it occupies, and the greater the utility is; and vice versa.

2.3. Correlation coefficient evaluation theory
The non-normally distributed random number series index not only needs to require various parameters of the Copula model, but also introduces a correlation relationship metric that is not interfered by the edge characteristic distribution type. The Kendall correlation coefficient and the Spearman correlation coefficient have the characteristics of being unaffected by the edge distribution, and the related variables can be represented by the Copula model and always exist. Therefore, these two coefficients are the indicators for measuring the output correlation model. The main considerations in the analysis of the most suitable Copula function include fitting indicators: Euclidean distance (distance sum of sample empirical distribution function values and Copula joint distribution function values) and maximum distance (sample empirical distribution function value and maximum value of Copula joint distribution function value). The smaller the difference is, the closer the model is to the empirical distribution; for correlation indicators: the smaller the difference between the kendall correlation coefficient $\tau$ and the Spearman correlation coefficient $\rho$ is, the closer the empirical data model is to the selected model.

Definition: Sampling $(x, y)$ from two population samples $(X, Y)$, records $F(x)$ and $G(y)$ as $X, Y$ empirical distribution functions, $u$ and $v$ are uniform distributions after $X$ and $Y$ transformation, so the empirical distribution of the sample can be defined as:

$$\hat{C}(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[F(x_i) \leq u] \mathbb{I}[G(y_i) \leq v]$$

(1)

Where $\mathbb{I}[G(y_i) \leq v]$ is an indicative function, when $F(x) \leq u$, $\mathbb{I}[F(x) \leq u] = 1$; when $F(x) > u$, $\mathbb{I}[F(x) \leq u] = 0$. Let $C(u, v)$ be the Copula joint distribution function value, then the Euclidean distance definition is:

$$d_G = \sqrt{\sum_{i=1}^{n} (\hat{C}(u, v_i) - C(u, v_i))^2}$$

(2)

The maximum distance is defined as:

$$d = \max \{|\hat{C}(u, v_i) - C(u, v_i)|\}$$

(3)

According to the above analysis, it can be concluded that the preferred Copula function can be classified into a decision problem, which can be carried out by using the entropy weight optimization theory described above. First, it is necessary to establish the attribute space of the four evaluation criteria to generate a standard hierarchical matrix:
$$\begin{bmatrix}
    P_1 & P_2 & P_3 & P_4 & P_5 \\
    x_1 & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
    A = x_2 & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
    x_3 & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
    x_4 & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} 
\end{bmatrix}$$

\[ a_{ij} = X_{i, \min} + j \cdot (X_{i, \max} - X_{i, \min}) / 5 \] (4)

\[ a_{ij} = X_{i, \min} + j \cdot (X_{i, \max} - X_{i, \min}) / 5 \] (5)

Where \( x(i = 1,2,3,4) \) is the evaluation criterion; \( P \) is the attribute space of index \( x \), \( P_j = (j = 1,2,\ldots,5) \) is the superiority and bad grade, \( P_1 \) is the best, and \( P_5 \) is the worst; \( a_{ij} \) is the \( j \) th grading standard of index \( i \) in the attribute space \( P \).

In order to obtain the attribute test value of the Copula model, it is required to evaluate the attribute measure matrix of the index, and then use the entropy ownership consciousness theory to determine various indicators. Indicator \( x \) has a \( P_m \) attribute, which is recorded as “\( x \propto P_m \)”. The attribute degree can be expressed by \( \lambda_m(P_m) \), which is the attribute measure value. The calculation formula of the model measure value is:

\[ \lambda(P_m) = \sum_{i=1}^{4} \omega_i \lambda_i(P_m), m = 1,2,\ldots,5 \] (6)

Where \( \omega_i \) is the weight of the evaluation criterion \( x_i \), which can be determined by the entropy method. Then according to the confidence criteria, you can find the pros and cons of the Copula model \( m \):

\[ m = \min \left\{ m \mid \sum_{i=1}^{4} \lambda_i(P_m) \geq \delta, 1 \leq m \leq 5 \right\} \] (7)

Where the confidence \( \delta \in (0.5 \sim 1) \) is generally between 0.6 and 0.7, and the highest-rated \( P_1 \)-class Copula model can be used as the optimal joint distribution model for the wind farm output.

3. Simulation study analysis

3.1. Nuclear density estimation

Here, two wind farms 1 and a wind farm 2, which are relatively close to each other, are located in Fuxin City, Liaoning Province. Take the real data of the 2018 year of active output as a simulation example. The data sampling interval is 15 minutes. First determine the edge distribution function of wind power output.

Through the statistics of the output frequency of the two wind farms, as shown in Figure 1, Figure 2, and the joint output distribution map, as shown in Figure 3. It can be seen that the output of the two wind farms is extremely low at the top and the tail is prominent. Therefore, the two wind farms have a tail asymmetric distribution, that is, when the two wind farms have a small output at the same time, they occupy a large proportion.
Figure 1. Statistics of the frequency of wind farm 1.

Figure 2. Statistics of the frequency of wind farm 2.

Figure 3. Statistics of the frequency of joint output of two wind farms.

Because the wind farm output has strong randomness and uncertainty, and verify that the two wind farms are subject to a normal distribution, the results show that they do not obey the normal distribution and cannot be simulated by common distributions such as t distribution and normal distribution. Therefore, the more versatile Copula function is used for correlation modeling. The non-parametric estimation method, i.e., the nuclear density estimation method, is used to simulate the edge distribution of wind power output, as shown in Figure 4 and Figure 5.

Figure 4. Nuclear density estimation map of wind farm 1.

Figure 5. Nuclear density estimation map of wind farm 2.

As shown in the figure, the solid blue line is the empirical cumulative probability distribution function. This line is an approximation of the actual cumulative probability distribution. It can be used to judge whether the fitting result is close to the actual; and the black dotted line indicates the nuclear density. The cumulative probability distribution function estimated by the estimation method is verified by the fitting index, and the close relationship between the nuclear density estimation distribution and the actual distribution is in accordance with the analysis standard, so it can be used as the edge distribution for correlation modeling.
3.2. Choice of optimal Copula function

According to the characteristics of the five Copula functions, the three functions describing the asymmetric distribution and the characteristics of the tail part are selected for the correlation evaluation, such as t-Copula function, Clayton-Copula function and Normal-Copula function, as shown in Table 1. The optimal function is selected by the entropy weight optimization theory to complete the modeling of wind farm correlation.

Table 1. Evaluation index parameter values.

| Model type | Pros and cons |
|------------|---------------|
| Frank      | $P_5$         |
| t          | $P_2$         |
| Normal     | $P_2$         |
| Clayton    | $P_1$         |
| Gumbel     | $P_2$         |

According to the data in the chart, the final result is that the Clayton-Copula function is the closest. The maximum likelihood distance of the Clayton-Copula function can be obtained by using the maximum likelihood estimation method to verify the results. Therefore, the Clayton-Copula function is chosen to describe the correlation between wind farm 1 and wind farm 2.

3.3. Copula function modeling

The density distribution and distribution function of two function models can be obtained by bringing the nuclear density estimation edge distribution of two wind farms into the Clayton-Copula function and the binary t-Copula function. Figure 6-10 shows the function.

![Figure 6. Density function of Clayton-Copula.](image1)

![Figure 7. Cumulative function diagram of Clayton-Copula.](image2)

![Figure 8. Cumulative function diagram of t-Copula](image3)

![Figure 9. Density function of t-Copula.](image4)
Figure 10. Empirical cumulative distribution function.

The t-Copula and Clayton-Copula functions are compared with the empirical Copula function respectively. The calculation results show that the Clayton-Copula function is closer to the empirical Copula function. And according to the characteristics of Clayton-Copula function: asymmetric distribution, lower tail correlation, upper tail progressive independence, just in line with the characteristics of data distribution. Therefore, the Clayton-Copula function is the best correlation function for the two wind farms.

4. Conclusions
In this paper, for the multi-wind farms with similar geographical locations, the nuclear density estimation method is used to solve the edge distribution, and the output correlation function between multiple wind farms is established. In a variety of Copula functions, Kendall correlation coefficient, Spearman correlation coefficient, Euclidean distance and maximum distance are analyzed and compared. Combined with the entropy weight optimization theory, the best correlation model of wind farm 1 and wind farm 2 is Clayton-Copula function. It can be seen from the function that the joint distribution has a strong tail-tail correlation. In wind power analysis applications, ignoring this correlation may lead to a series of situations that are inconsistent with the actual operation and increase the risk of safe and stable operation of the grid.

References
[1] Qiu Yibin, Meng Xiang, Ouyang Yubo, et al. (2017) Dependence Modeling of Multidimensional Wind Farm Output Based on Mixture Vine Copula Structure. J. Acta energiae solaris sinica., 38: 2512-2519.
[2] Ding Ming, Wu Yichun, Zhang Lijun. (2005) Study on the Algorithm to the Probabilistic Distribution Parameters of Wind Speed in Wind Farms. J. Proceedings of the CSEE., 25: 107-110.
[3] Zhao Wenmeng, Liu Mingbo, Zhu Jianquan. (2015) A Bi-level Decomposition And Coordination Economic Dispatch Method for PowerPlants/Network Considering Stochastic Wind Generation. J. Power System Technology., 39: 1847-1854.
[4] Wu Wei, Wang Keyou, Li Guoje, et al. (2017) The Consideration Electric Service Condition Relevance Ryukyu Indeterminate Set Construction. J. Proceedings of The Chinese Society for Electrical Engineering., 37: 2500-2507.
[5] Liu Ji, Xu Jian, Sun Yuanzhang, et al. (2019) Dynamic Economic Dispatch of Power System Considering Temporal Correlation of Wind Power Sequence. J. Automation of Electric Power Systems. 43: 43-91.
[6] Wang Shuang. (2011) High-energy and Low-power Research. D. Changsha University of Technology, Changsha.
[7] Zhou Hui, Zhang Xinsong, Guo Xiaoli, et al. (2018) Modeling of wind power correlation based on Copula theory and its application in transmission network expansion planning. J. Science Technology and Engineering., 18: 273—278.
[8] Nelsen R B. (1999) An Introduction to Copulas. S. Berlin: Springer.