Copula Correlation Modeling of Wind Farms Generation and Its Application in Power Dispatching

Ning Nan¹, Wu Xiaokang¹, Shao Qianwen¹, Xie Min², Huang Binbin²*, Li Shujia²

¹.Gui’an power supply bureau, Gui’an, 550003 China
².School of Electric Power, South China University of Technology, Guangzhou, 510640 China

The corresponding author’s e-mail address: minxie@scut.edu.cn

Abstract. With the large-scale integration of wind power into the grid, a model that can accurately describe the randomness of wind farm output and the correlation between them is of great significance to be constructed. A joint distribution function of multiple wind plants output model based on Copula theory is constructed. By introducing correlation and fitting coefficient, the attribute recognition theory is put forward to select the optimal model based on entropy weight method. Finally, the validity of Copula modeling is verified by using the synchronized historical data of California coastal wind farms as a sample. The results show that t-copula can not only describe the correlation between the original variables well, but also fit the empirical distribution function of the original samples accurately.

1. Introduction

The increasing popularity of wind power has greatly enhanced the randomness of power grid operation[1-2]. Due to the similarity of region and climate, the output characteristics of wind farms may have strong correlation in time and region. Therefore, it is necessary to accurately construct the model describing the randomness and correlation of multiple wind farms.

The methods to describe the random nature of wind power mainly include probability distribution modeling[3-4] and stochastic sequence modeling[5]. However, these methods don’t consider the correlation between the wind power farm outputs, which often causes problems such as deviation of power flow calculation, overload of lines, and the difficulty of active power scheduling[6].

Recently, Copula theory has become a popular method to formulate the correlation of random variables. The basic nature of Copula is introduced in book[7] and some of their main applications are discussed including correlation and correlation measures. In the paper[8], the clustering of the wind field near the coastline of the Netherlands is equivalent to two wind fields, and the Gaussian-Copula function is introduced to establish the joint output distribution model of the two wind fields. In the paper[9], a probabilistic power flow algorithm considering multi-correlation wind sources in the power grid is proposed. Based on the copula method, a bivariate model is established to simulate the relationship between wind farms. However, the copula functions used in the above literatures are simply selected only by certain correlation coefficients, while the fitness between the model and the actual sample model is not evaluated. So, the accuracy of the correlation formulation between multiple wind farms output is not high enough.

In this paper, a set of correlation description index and fitting effect index are proposed for the establishment of multi-wind farms Copula correlation model. In order to verify the effectiveness of the
model, the synchronous history data of California coastal wind farms is applied to IEEE118 system as a simulation example.

2. The evaluation index of random variable correlation based on Copula

The historical output data of a typical coastal wind farm are processed as follows: drawing wind power plant output frequency histogram (as shown in figure 1), its skewness value is 1.4751 and its kurtosis value is 6.1178.

![Figure 1. Output frequency histogram of a coastal wind farm.](image1)

It can be seen that the output distribution of this wind farm presents the characteristic of peak and thick tail compared with the normal distribution. The wind farm does not follow a normal distribution by kolmogorov-smirnov test[10]. The Q-Q distribution map of two wind farms along the coast is shown in figure 2. In the figure, the distribution points of output of the two wind farms are basically concentrated on the diagonal, and there is a certain correlation.

To measure the correlation between wind farms, a general correlation measurement criterion that is not affected by the type of edge distribution needs to be introduced, and this index can be obtained from the Copula model parameters. Kendall rank correlation coefficient and Spearman rank correlation coefficient[11] are chosen as the correlation evaluation index of the original sample in this paper.

3. The joint distribution modeling of wind farm output based on Copula theory

3.1. Multi-wind farm Copula correlation modeling

Copula function can not only characterize the non-normal properties of a single random variable but also describe the complex correlation between different variables. The Copula function mainly contains five types: Gaussian Copula, t Copula, Gumbel Copula, Clayton Copula and Frank Copula. The specific expressions and properties of the function are shown in article [12].

The kernel density estimation method is adopted to determine the cumulative distribution function of wind farm output in this paper. Unknown parameters in five types of Copula functions are obtained by using the method of stepwise maximum likelihood function. The joint distribution model of five types of wind farms can be generated by fitting the five types of Copula functions.

3.2. Optimization method of Copula model based on entropy weight attribute recognition theory

After the establishment of five types of Copula models, the most appropriate type is necessary to be selected as the wind farm joint distribution model according to the evaluation index, including: (1) Correlation index: as for the difference between Kendall correlation coefficient and the difference between Spearman correlation coefficient represented as \( d_\tau \) and \( d_\rho \). (2) Fitting index: the smaller the Euclidean distance \( d_{eu} \) and the maximum distance \( d_z \), the better the model can fit the original empirical data distribution.

The entropy weight attribute identification theory is applied to optimize the Copula model of joint distribution of wind farms. The specific operation steps are as follows:

![Figure 2. Q-Q diagram of two wind farm output.](image2)
(1) The attribute space based on the four evaluation indexes was established, and the classification standard matrix $A$ was obtained.

$$
A = \begin{bmatrix}
P_1 & P_2 & P_3 & P_4 & P_5 \\
1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 2 \\
3 & 3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4 & 4 \\
\end{bmatrix}
$$

(1)

$x_i (i = 1, 2, 3, 4)$ is the evaluation index. $P$ is the attribute space of the index $x_i$. $P_j (j = 1, 2, ..., 5)$ is quality grade. $a_{ij}$ is the $j$-th grading standard value of indicator $i$ on attribute space $P$: $a_{ij} = X_{k_{\text{min}}} + j \times (X_{k_{\text{max}}} - X_{k_{\text{min}}}) / 5$ (2)

(2) The weight of each index is determined by entropy theory to get the attribute measure value of Copula model. $\omega_i$ is the weight of evaluation index $x_i$, which is determined by entropy method. According to the confidence criterion, the grade $m$ of Copula model is obtained:

$$
\lambda(P_m) = \sum_{i=1}^{4} \omega_i \times \lambda_i(P_m) \quad m = 1, 2, \cdots, 5
$$

(3)

$$
m = \min \{m | \sum_{i=1}^{4} \lambda_i(P_m) \geq \delta, 1 \leq m \leq 5\}
$$

(4)

The confidence ($\delta \in (0.5 - 1)$) of Copula model which between 0.6 and 0.7 and the model rated $P_1$ can be regarded as the wind farm output joint distribution model.

4. Application of multi-wind farms correlation model in dynamic economic dispatching model of power system

4.1. Model formulation

The stochastic scenario method to solve the stochastic programming model of dynamic economic dispatching with multi-wind power is used in this paper.

(1) Objective function: minimize power generation cost function $f$:

$$
\min f = f_y + f_w = \sum_{i=1}^{T} \sum_{s=1}^{S} a_i \times P_{Gi}(t)^2 + b_i \times P_{Gi}(t) + c_i + \min \left(\sum_{i=1}^{T} \sum_{s=1}^{S} b_i \beta_i (P_{Gi}(t) - P_{Gi}(t))^2 \right)
$$

(5)

where $f_y$ is the coal consumption cost of conventional units under the predicted scenario; $f_w$ is the rescheduling cost caused by the difference between the scheduling scheme of error scenarios and prediction scenario formulated. $T$ is the total number of time periods of the scheduling cycle; $N$ is the number of conventional generating sets; $P_{Gi}(t)$ is the generating power of unit $i$ at time period $t$; $a_i, b_i, c_i$ are the coal consumption cost coefficients of conventional units. $S$ is the number of random error scenes; $h_i$ is the probability of random error scene $s$; $\beta_i$ is the fuel cost coefficient of the redispatching of unit $i$; $P_{Gi}(t)$ is the active power output of conventional unit $i$ in scene $s$ at time period $t$.

(2) Prediction of scenario constraints.
\[
\sum_{i=1}^{N} P_{Gi}(t) + \sum_{k=1}^{N_w} P_{wk}(t) = P_{load}(t)
\]
\[
P_{Gi,min} \leq P_{Gi}(t) \leq P_{Gi,max}
\]
\[
P_{Gi}(t-1) - P_{Gi}(t) \leq r_{di} \times T
\]
\[
P_{Gi}(t) - P_{Gi}(t-1) \leq r_{ui} \times T
\]
\[
|P_{mn}(t)| \leq \overline{P}_{mn}, P_{mn}(t) \in R^l
\]

where \(N_w\) is the number of wind turbines, \(k = 1,2,\ldots,N_w\); \(P_{wk}(t)\) is the predicted output of wind turbine \(k\) at time period \(t\); \(P_{Gi,max}\) and \(P_{Gi,min}\) are the upper and lower limits of output of conventional units; \(r_{di}\) and \(r_{ui}\) are the upward and downward climbing rates of conventional units. \(P_{mn}(t)\) is the active power transmission power of branch \(mn\) in time period \(t\), which can be obtained according to the dc power flow method. \(\overline{P}_{mn}\) is the maximum active power transmission upper limit of branch \(mn\).

(3) Random error scene constraint. Equation (8) is the power speed regulation constraint between the prediction scene and the random error scene of conventional units.
\[
|P_{Gi}(t) - P_{Gi}^t(t)| \leq \Delta_i
\]
where \(\Delta_i\) is the adjustable power of conventional unit at time period \(t\).

4.2. Solution procedure
The scenario considering the correlation of wind farm output is generated by random variable sampling in the known Copula joint distribution with Monte Carlo (MC) sampling. The solution process of the model is shown as figure 3.

5. Application of multi wind power plant correlation model in dynamic economic dispatching model of power system
The IEEE118 node system is taken as an example to verify the effectiveness of the proposed algorithm.

5.1. The joint output distribution of multiple wind farms
The joint Copula distribution function between the output of the two wind farms is constructed with the historical synchronous data of the two wind farms\(^{[13]}\).
1. Model evaluation and optimization.

The comparison between the empirical distribution and nuclear distribution estimation of the output power of W5 wind farm is shown in figure 4. The original data and evaluation indexes of the five types of Copula models are obtained, as shown in table 1.

| Function type       | $\rho_0$ | $\rho_1$ | $d_{40}$ | $d_{r}$ |
|---------------------|----------|----------|-----------|---------|
| sample data         | 0.6657   | 0.8452   | ---       | ---     |
| Gaussian-Copula     | 0.6550   | 0.8539   | 2.3379    | 0.0177  |
| t-Copula            | 0.6653   | 0.8542   | 2.1785    | 0.0172  |
| Gumbel-Copula       | 0.6651   | 0.8476   | 2.2436    | 0.0184  |
| Clayton-Copula      | 0.5104   | 0.6939   | 52.8739   | 0.0654  |
| Frank-Copula        | 0.6537   | 0.8501   | 7.3781    | 0.0279  |

The t-copula model has the best rating. So t-copula model is selected as the joint probability distribution of the output of two wind farms.

2. Model validation

By analyzing the binary frequency histogram of the edge distribution of the original sample (Figure 5), it can be seen that the distribution of the wind farm is U-symmetric and has the correlation between the upper and lower tails, which is also consistent with the probability density distribution map of t-Copula (Figure 6).

5.2. The solution of dynamic economic dispatch for multi-wind power grid based on Copula scenario

1. Dispatch considering multi-wind farms correlation

1000 wind farm output scenes are generated randomly from the wind farms Copula correlation distribution for 24 hours. In order to increase the computation efficiency, the scenes are reduced to 20 by using the scene elimination method\[14\]. Three meteorology cases are considered: the windy case, the less-windy case and the ordinary case.

The corresponding results are shown in table 2. In the windy and windless seasons, the wind power output increases the challenge to the system operation flexibility compared with the ordinary scenario. Therefore, the rescheduling cost of windy and windless cases is both higher than ordinary case. It can be seen that the smaller the actual rescheduling cost is, the closer the scheduling scheme corresponding to the dispatching model is to the real scheduling result. It means that the more accurate the model is.

| Period | Consider correlation? | Power costs | Coal consumption costs | Rescheduling costs | Actual rescheduling |
|--------|-----------------------|-------------|------------------------|--------------------|--------------------|
| Yes    | 3473199               | 3455174     | 18024                  | 199054             |
Normal period | Yes | 3558380 | 3491889 | 66491 | 70914 | No | 3538640 | 3492050 | 46590 | 103663 |

Windy period | Yes | 3501114 | 3471122 | 29992 | 50479 | No | 3487404 | 3471108 | 16296 | 52404 |

less windy period | Yes | 3501114 | 3471122 | 29992 | 50479 | No | 3487404 | 3471108 | 16296 | 52404 |

Compare the objective costs \((f_c, f_y, f_w)\) of the dispatching model considering wind farm correlation are generally larger than the model without wind farm correlation. This is because there is an up-down tail correlation between the two wind farms. It is verified that the dispatching model considering wind farm correlation can higher accuracy to simulate the system scheduling than non-considering wind farms correlation.

6. Conclusion
The following characteristics in copula-based multi-wind farm output correlation modeling and its application are shown in this paper:

1) The general evaluation index of the model is put forward based on five types of Copula function are given. Then, the optimization method of Copula model based on entropy weight attribute recognition theory is proposed, so that the Copula model selected finally fits the actual distribution better.

2) By sampling the Copula distribution based on monte carlo sampling to form a multi-scene, the application of dynamic economic scheduling of multi-wind power plants considering correlation is analyzed and the validity of Copula correlation modeling and optimization method of multi-wind field output is verified in this paper.

Acknowledgements
This work is supported by the Science and Technology Project of Guizhou Power Grid(No. 061000KK52180003) and the major industry university research cooperation support projects for the Fundamental Research Funds of the Central Universities (No. x2dlD2201280).

References
[1] Tang C., Xu, J., Sun Y., et al. Look-ahead economic dispatch with adjustable confidence interval based on a truncated versatile distribution model for wind power [J]. IEEE Trans. On Power System, 2018, 33(2): 1755-1767.

[2] H.L. Liao, Q.H. Wu, Y.Z. Li, L. Jiang, Economic emission dispatching with variations of wind power and loads using multi-objective optimization by learning automata [J]. Energy Conversion & Management, 2014, 87: 990-999.

[3] Dao R N, Meng K Q L, Zhang Z Q. Improvement and modeling simulation of combined wind speed in the wind power system by Rayleigh probability distribution[J]. Renewable Energy, 2014, 32(4):461-465.

[4] Chiodo E , Lauria D . Analytical study of different probability distributions for wind speed related to power statistics[C]// Clean Electrical Power, 2009 International Conference on. IEEE, 2009.

[5] Chen P , Pedersen T , Bak-Jensen B , et al. ARIMA-Based Time Series Model of Stochastic Wind Power Generation[J]. IEEE Transactions on Power Systems, 2010, 25(2):667-676.

[6] Xie K , Billinton R . Considering wind speed correlation of WECS in reliability evaluation using the time-shifting technique[J]. Electric Power Systems Research, 2009, 79(4):687-693.

[7] Nelsen R B. An Introduction to Copulas. New York: Elsevier Science,1983.

[8] Pagaefthymiou G , Kurowicka D. Using Copulas for modeling stochastic dependence in power system uncertainty analysis[J]. IEEE Trans on Power Systems, 2009, (24): 40-49.

[9] Qiang Fu, Yu D, Ghorai J. Probabilistic load flow analysis for power systems with multi-correlated wind sources[C]. IEEE Power and Energy Society General Meeting, San Diego, USA, 2011: 1-6.
[10] Davidson B T. A Kolmogorov-Smirnov Test for Censored Samples[J]. Technometrics, 1973, 15(4):739-757.
[11] GaiBer S, Schmid F. On testing equality of pairwise rank correlations in a multivariate random vector[J]. Journal of Multivariate Analysis, 2010, 101(10): 2598-2615.
[12] Lehmann E L. Nonparametrics: Statistic methods based on ranks[M]. San Francisco: Holdem-Day Inc, 1975: 43-45.
[13] Debbie Lew. Western wind dataset[DB/OL]. http://www.nrel.gov/.
[14] Heitsch H, Römisch W. Scenario reduction algorithms in stochastic programming[J]. Computational Optimization and Applications, 2003, 24(2-3): 187-206.