Wind-electricity-heat Correlation and Potential Uncertainty Analysis Based on Copula Function

Tie Li¹*, Jinjing Hu¹, Feng Jiang¹, Haiyang Cong¹, Xiaoyang He¹, Weichun Ge¹ and Yi Li²

¹State Grid Liaoning Electric Power Supply Co. Ltd, Shenyang City, Liaoning Province, 110004, China
²State Grid Shenyang Electric Power Supply Company, Shenyang City, Liaoning Province, 110003, China

*Corresponding author: litie2019@163.com

Abstract. In the economic dispatch and optimal operation of power system, it is necessary and effective to consider the correlation among heat load, power load and wind power output to formulate a reasonable dispatch plan or to evaluate the reliability of the system. Therefore, based on the theory of Copula correlation analysis, a multivariate Copula analysis toolbox (MvCAT) is proposed to infer Copula parameters and estimate potential uncertainties. Firstly, the mixed evolution Markov Chain Monte Carlo (MCMC) method in Bayes framework calculates the posterior distribution of Copula parameters, evaluate their uncertainties relative to fitting, and then select the appropriate Copula function by goodness-of-fit test. Finally, a the typical daily data of a province as samples for analysis, proposed method solves the limitation that the local optimization method often falls into the local minimum, and quantitative evaluation of the correlation between the response variables and the uncertainty associated with the length of the recorded data is essential for multivariate frequency analysis.

Keywords: Correlation, Uncertainty, Copula Function, Markov Chain Monte Carlo

1. Introduction
The economic dispatching of power system is a power generation dispatching mode which takes the minimum total cost of power generation as the goal under the premise of considering the correlation between power generation and load. There are correlations among power load, heat load and wind power output in a certain area [1]. Therefore, it is necessary to accurately model the correlations among the indicators in the process of economic dispatch and optimal operation to quantify the impact of uncertainty among the indicators on the power grid [2]. If this relationship is neglected, it may cause errors in calculation results, and even directly affect the safe and economic operation of power system [3-4]. Therefore, it is of great practical significance to accurately evaluate the correlation among power load, heat load and wind power output.
Correctly deal with the correlation between non-normal random variables is the key to study the correlation between variables [5]. However, the linear correlation coefficient cannot accurately describe the correlation of non-normal distribution of random variables with some errors in the calculation results. Copula theory is a multivariate analysis method, which can accurately describe the structural relevance of multivariate variables. It is widely used in the modeling of the dependent structure of two (or more) random variables [6-8]. Compared with other methods, the Copula function has several advantages in describing the correlation of random variables [9-12]: marginal distribution is not restricted; the modeling steps are simple and use flexible; it can describe the non-linear correlation between random variables. The above research methods adopt Copula theory to analyze the correlation among variables, but did not analyze its potential uncertainty, so that it could not accurately describe the correlation among variables.

Given this, this paper proposes to infer Copula parameters and estimate potential uncertainties by Multivariate Copula Analysis Toolbox (MvCAT). As the data samples of thermal load, electric load and wind power output in a certain area, MCMC is used to constrain the parameters of different copula families, and the posterior distribution of parameters is given to match the uncertainty information of Copula probability profile. The results show that the proposed method can not only completely reflect the correlation between variables, but also quantitatively describe the potential uncertainty of Copula modeling.

2. Copula Function Basic Theory and Relevant Index

2.1 Copula Theory

Copula is a mathematical function that "connects" or "couples" two or more variables independent of time regardless of their univariate distribution, which they are a systematic method to study the basic dependency structure and provide a basis for constructing binary (multivariable) distribution families. Copula function is an n-dimensional distribution function, whose definition range is [0,1]. This function is a mapping tool, which realizes the transformation from edge distribution to joint distribution. The practical significance of Copula function is to establish the correlation between the edge distribution and joint distribution of random variables. Usually, it studies the edge distribution of multiple random variables and the structural characteristics associated with each other. If a variable H is assumed, it is expressed as follows, where h is the joint cumulative distribution function, and its independent variables are the marginal univariate distribution F and G. At the same time, suppose X and y are two continuous two-dimensional random variables. Sklar theorem indicates that when F and G are continuous, there exists a unique and definite Copula function $C(\cdot)$ that satisfies:

$$H(x, y) = C[F(x), G(y)]$$  \hspace{1cm} (1)

We can also extend Sklar's theorem to joint distribution function considering multivariate distribution.

2.2 Correlation Index

Pearson correlation coefficient is usually used to measure the correlation between random variables, which can be used to describe the correlation between variables in normal distribution. Generally, it can only reflect the linear relationship among random variables, but it cannot be used to describe the correlation of variables in non-normal distribution. It is necessary to introduce an index that can measure the correlation of random variables, a general correlation metric that is not affected by the type of marginal distribution, and get the index from the parameters of copula model. Therefore, Next, Kendall correlation coefficient and Spearman correlation coefficient are introduced. Kendall correlation coefficient is used to indicate whether there is consistency in the change trend of multiple random variables. Spearman correlation coefficient is used to measure the degree of linear correlation between multiple random variables. For example, suppose that there are two random variables X and Y in a binary copula function. If $u = f(x)$ and $v = g(y)$, there must be a copula function $C(U, V)$,
which defines the correlation between X and Y by Kendall correlation coefficient, and measures the correlation by Spearman correlation coefficient.

\[
\tau = 4\int_0^1 \int_0^1 C(u, v) \, dC(u, v) - 1
\]

(2)

\[
\rho = 12\int_0^1 \int_0^1 uvdC(u, v) - 3
\]

(3)

But Copula function has a great variety. Obtained and are not the same for different Copula functions, so the accuracy of correlation measure depends on the selection of Copula function fitting degree.

2.3 Copula Family

The Copula function is mainly divided into two types: elliptic type and Archimedean type. According to the different ways of construction, it can be divided into symmetric type and asymmetric type. Different types of Copula functions have different mathematical expressions and are suitable for describing the correlation of different types of variables. In this paper, 26 binary Copulas are built into the multivariate Copula analysis toolbox (MvCAT) to select the Copula family with a simple closed form mathematical formula, namely: Gaussian, t, Clayton, Frank, Gumbel, Independence, ali-mikhail Haq (AMH), Joe, gumbel-barnett, farlie-gumbel-morgenstern (FGM), Plackett-Cuadras-auge, Raftery, shih-louis, Cubic, Burr, Linear Spearman, Nelsen, Galambos, BB1, BB5, and Tawn. They can be used for model inference and represent different forms of dependency structures.

3. Correlation Analysis and Modeling Based On Copula Function Variables

3.1 Bayes Analysis

At present, Bayes model has been widely used in various fields of model reasoning process and uncertainty quantification. Mvcat usually uses Bayes framework based on residual Gauss likelihood function. Firstly, it judges whether there is uncertainty in copula parameter, and then estimates it. According to Bayes theorem, if there is new information input, the priori probability of the hypothesis needs to be updated. By using Bayes law, all modeling uncertainties can be easily transformed into a certain parameter, and the posterior distribution of model parameters can be estimated by the specific methods mentioned below:

\[
p(\theta | \tilde{Y}) = \frac{p(\theta) p(\tilde{Y} | \theta)}{p(\tilde{Y})}
\]

(4)

Where \( p(\theta) \) and \( p(\theta | \tilde{Y}) \) respectively represent the prior distribution and posterior distribution of parameters, \( \theta \) is estimated parameters, \( \tilde{Y} \) represents the joint probability of observed variables, and \( Y \) represents the probability value predicted by Copula. \( p(\tilde{Y} | \theta) \) represents a likelihood function. As a constant in each modeling practice, \( p(\tilde{Y}) \) can be removed from the analysis. The estimated posterior parameter distribution can be obtained by:

\[
p(\theta | \tilde{Y}) \propto p(\theta) p(\tilde{Y} | \theta)
\]

(5)
Bayes equation is a little difficult to calculate. In order to solve this problem, this paper adopts the numerical analysis method of MCMC simulation to extract samples from the posterior distribution.

### 3.2 Markov Chain Monte Carlo Simulation

As a special statistical method, Markov chain Monte Carlo algorithm is generally used to extract relevant sample data from high-dimensional complex distribution. In this paper, a Markov chain Monte Carlo algorithm is proposed, which is based on adaptive suggestion distribution and can be used to describe the posterior parameter region in Bayes environment. The biggest advantage of Markov chain Monte Carlo algorithm based on hybrid evolution is that it can intelligently select the starting point, and use adaptive algorithm to search the solution space, and finally get the result.

### 3.3 Measures for Goodness Of Fit

In this paper, we select a number of different goodness of fit to evaluate the characteristics of each copula model. The goodness of fit here includes likelihood, AIC, BIC, RMSE and NSE. In the calculation process, formula (6) is used to solve the likelihood value, and the maximum likelihood parameter set is found according to the solution result. Then, the minimum model is selected to realize the simulation and observation of the residual. From this point of view, the model contains the most appropriate observation data. The complexity of the model is simplified by AIC, and the error residual is minimized. This calculation provides a stable measurement method and improves the prediction quality of the model. AIC can be expressed as:

$$ AIC = 2D + n \ln n + \frac{\sum_{i=1}^{n} \left( \hat{y}_i - y_i(\theta) \right)^2}{n} - 2CS $$

Where $D$ is the parameter of the statistical model, constant CS. The lower the AIC value, the better the model fitting. Similar to AIC, BIC can be expressed as:

$$ BIC = D \ln n + n \ln n + \frac{\sum_{i=1}^{n} \left( \hat{y}_i - y_i(\theta) \right)^2}{n} - 2CS $$

The smaller the BIC result is, the more consistent the established model is with the real model, which is the same as AIC. Similarly, NSE and RMSE are two kinds of measurement methods, which are widely used to evaluate the fitting degree of the model, but they usually only focus on the indicator of minimizing the residual. Their expressions are as follows:

$$ RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left( \hat{y}_i - y_i(\theta) \right)^2}{n}} $$

$$ NSE = \frac{\sum_{i=1}^{n} \left( \hat{y}_i - y_i(\theta) \right)^2}{\sum_{i=1}^{n} \left( y_i - \bar{y} \right)^2} $$

$$ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i $$

\[
\text{NSEE} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i(\theta))^2}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2}
\]

The perfect model is fitted as \( \text{RMSE} = 0 \) and \( \text{NSEE} = 1 \). All these measures evaluate the performance of Copula in different ways based on the proximity of the simulated bivariate probability \( Y \) to its empirical observation \( \hat{Y} \).

4. Conclusion

Thermal load, electrical load and wind power output are interrelated, and their correlation has certain significance for power system scheduling and risk analysis. This paper uses Multivariate Copula Analysis Toolbox (MvCAT) to describe dependencies and potential uncertainties by Bayes framework. Local optimization algorithm may fall into local optimum and cause deviation. In contrast, Markov Monte Carlo simulation method can randomly select the starting point, and then search the posterior region of interest from the starting point, which effectively explores the whole feasible space and produces results when the target value is reasonably retrieved. In this study, we also focus on the uncertainty of model parameters, which will affect the uncertainty of the results when the information in the constrained data is limited. We are efforts to conduct a more comprehensive analysis of different uncertainties.

Acknowledgements

This work was supported by science and technology project of State Grid Liaoning Electricity Power Limited Company (SGLNSY00FZJS1801113) "research and implementation scheme of the thermoelectric coordination intelligent control assistant decision technology based on big data".

References

[1] Xie Min, Xiong Jing, Liu Mingbo, et al. Modeling of multi wind farm output correlation based on copula and its application in power system economic dispatch[J]. Power System Technology, 40(4):1100-1106(2016)(in Chinese).
[2] Ding Jiaman, Tang Yan, Wang Qingxin, et al. Copula function based correlation analysis and modeling of power grid planning indexes[J]. Modern Electronics Technique, 41(17):95-101 (2018) (in Chinese).
[3] Wang Xiaohong, Zhou Buxiang, Zhang Le, et al. Wind power correlation analysis based on time-variant copula function[J]. Proceedings of the CSU-EPSA,27(1): 43-48(2015)(in Chinese).
[4] Liu H H, Jiang W, Yang G H, et al. Single failure mode reliability analysis of mechanical components under dependent random variables with non-normal distribution[J]. Journal of Dalian University of Technology, 55(1):22-30(2015).
[5] Bin L, Shahzad M, Bing Q, et al. Probabilistic computational model for correlated wind farms using copula theory[J].IEEE Access, 6:14179 – 14187(2018).
[6] Jongman B, Hochrainer-Stigler S, Feyen L, et al. Increasing stress on disaster-risk finance due to large floods[J].Nature Climate Change, 4(4):264-268(2014).
[7] Éric Parent, Favre A C, Bernier J, et al. Copula models for frequency analysis what can be learned from a bayesian perspective?[J]. Advances in Water Resources, 63(2):91-103(2014).
[8] Fei C, Zheng Y, Jingbo Z, et al. Dependence structure models for wind speed and wind power among different wind farms based on copula theory[J].Automation of Electric Power Systems, 37(17):9-16(2013)
[9] Yang P, Libao S, Yixin N. Modelling of multiple wind farms output correlation based on copula theory[J]. The Journal of Engineering, 2017(13):2303-2308(2017).
[10] Grimaldi S, Petroselli A, Salvadori G, et al. Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses[J]. Advances in Water Resources, 90:116-133(2016).

[11] Xu H, Jiang M, Xu K. Archimedean copula estimation of distribution algorithm based on artificial bee colony algorithm[J]. Journal of Systems Engineering and Electronics, 26(2):388-396(2015).

[12] Sadegh M, Ragno E, Aghakouchak A. Multivariate copula analysis toolbox (MvCAT): Describing dependence and underlying uncertainty using a bayesian framework[J]. Water Resources Research, 53(6), 5166-5183(2017).