Correlation Analysis of Transformer Parameters Based on Pair-Copula

Ting Yu¹,a, Junjie Yang¹,², JinChen Lin¹ and Chuanping Geng¹

¹School of Electronic Information Engineering, Shanghai DianJi University, Shanghai, Shanghai, 201306, China
²College of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai, Shanghai, 201306, China

¹86003010117@st.sdju.edu.cn

Abstract. Transformer temperature is an important indicator to detect the operating status of equipment. State evaluation of transformer temperature can analyze whether the equipment is in good operating condition. However, the internal structure of the transformer is complicated and collaborative failures are prone to occur. In addition, different ambient temperatures will also affect the temperature measurement of the equipment itself. Therefore, there are a large number of uncertain factors inside and outside the transformer, resulting in inaccurate temperature analysis. The article introduces the Pair-Copula function model to analyze the correlation of various variables in the transformer system, and combines the rank correlation coefficient to filter the variables, and finally constructs the correlation model of the transformer temperature, which provides an important analysis basis for subsequent state evaluation.

1. Introduction

The temperature of the transformer is an important indicator reflecting the status of the equipment, but when the equipment fails, it will be accompanied by temperature rises of different magnitudes. Therefore, the health of the equipment can be judged by observing the temperature of the transformer. However, due to the complicated internal structure of the transformer, it is prone to inaccurate temperature judgment caused by various fault factors. Therefore, one of the hot issues in the research of the power industry at this stage is how to effectively determine the operating status of the transformer based on the equipment temperature data.

Literature 1 proposes a finite element-based simulation analysis method for transformer flow field and temperature field, which solves the problem that the temperature difference of the transformer heat collection point is too large, which causes the simulation flow field and temperature field to be missing[1]. Literature 2 introduces the Bayesian evaluation model and studies the use of the maximum membership function to determine the operating state of electrical equipment. The method adds a correction model to correct the transformer temperature, but the basis for temperature correction is not described in detail[2]. Literature 3 proposed a soft compensation method to correct the temperature to eliminate the difference between the two tables[3]. Literature 4 proposes to calculate the top oil temperature and winding hot spot temperature of the transformer based on the modified thermal circuit model method[4].

This paper proposes the Pair-Copula model. The Copula function has high flexibility and can construct the correlation between variables [5]. Through the Pair-Copula model, the correlation analysis
of some state parameters of the transformer is carried out, and a structural model that is more in line with the actual operating state of the equipment is established to realize the analysis and evaluation of the operating state of the transformer.

2. Pair-Copula Model

2.1. Copula Function

In 1959, Sklar proposed the Copula theory, which can combine the marginal distribution of d variables with a Copula function describing the correlation of variables to form an overall multivariate joint distribution function [6]. If \( \mathbf{X} = (X_1, X_2, \cdots, X_d) \) is a d-dimensional random variable, \( F_1(x_1), F_2(x_2), \cdots, F_d(x_d) \) is the marginal distribution function on the interval \([0,1]\). The joint distribution function of \( \mathbf{X} \) is \( F(x_1, x_2, \cdots, x_d) \), and the parameter of the Copula function is \( \kappa \), then there is a d-dimensional Copula function \( C \), which satisfies the following formula:

\[
F(x_1, x_2, \cdots, x_d) = C( F_1(x_1), F_2(x_2), \cdots, F_d(x_d) )
\]

The Copula function contains many distribution families [7], and the distribution function of the typical Copula function is shown in Table 1.

| Distribution family | Typical function | Distribution function |
|---------------------|------------------|----------------------|
| Ellipse             | Gaussian         | \( C(u,v;\kappa) = \phi^\ast(\kappa)(\phi^{-1}(u), \phi^{-1}(v)) \) |
|                     | t                | \( C(u,v;\kappa,\mu) = T_{\mu}^{-1}(u), T_{\mu}^{-1}(v) \) |
|                     | Gumbel           | \( C(u,v;\omega) = \exp\{-(\ln u)^{\alpha} - (\ln v)^{\alpha}\} \) |
|                     | Clayton          | \( C(u,v;\omega) = \exp\{-(\ln u)^{\alpha} + (\ln v)^{\alpha}\} \) |
|                     | Frank            | \( C(u,v;\omega) = -\frac{1}{\omega}\ln[1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}] \) |

Among them, the elliptic family Copula is more convenient for simulation experiments than the Archimedean family, and has a symmetrical tail correlation [8], so this paper uses the binary normal Copula function and the binary t-Copula function in the elliptic family Copula function to operate.

When the random vector is a continuous distribution function, it is also true if there is an inverse function of the marginal distribution function. The Copula function of the inverse function is as follow:

\[
C(g_1, \cdots, g_d) = F\{F_1^{-1}(g_1), \cdots, F_d^{-1}(g_d)\}, (g_1, \cdots, g_d) \in R^d
\]

The Copula density function is:

\[
f_{\kappa}(g_1, \cdots, g_d) = \frac{\partial^d C(g_1, \cdots, g_d)}{\partial g_1 \cdots \partial g_d}
\]

When the edge distribution is continuous, the joint distribution density of the joint distribution function \( F(\cdot) \) is \( f(\cdot) \), the edge density is \( f_j \), and the joint distribution density formula is as follow:

\[
f(x_1, \cdots, x_d) = c_{\kappa}(F_1(x_1), \cdots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j)
\]

2.1.1. Binary normal copula function. The distribution function of the binary normal Copula is as follow:

\[
C(u,v;\kappa) = \phi[\phi^{-1}(u), \phi^{-1}(v);\kappa] = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} (2\pi)^{-1/2} (1 - \kappa^2)^{-1/2} \exp\left\{-(r^2 + s^2 - 2\kappa rs)\right\} dr ds
\]

Among them, \( \kappa \in [-1,1] \) is the correlation parameter, \( \phi(\cdot) \) is the one-variable standard normal distribution function, and \( \phi^{-1}(\cdot) \) is the inverse function.
2.1.2. Binary t-Copula function. The distribution function of the binary t-Copula function is as follow:

\[ C(u, v; \kappa, \mu) = T[T^{-1}_\mu(u), T^{-1}_\mu(v); \kappa, \mu] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\kappa^2}} \left[ 1 + \frac{s^2 + t^2 - 2\kappa st}{\mu(1-\kappa^2)} \right] \frac{1}{2\pi} ds dt \]  

(5)

Among them, \( \kappa \) and \( \mu \) denote the correlation parameter and the degree of freedom parameter respectively, \( T_\mu(\cdot) \) is the one-variable t distribution function, and \( T^{-1}_\mu(\cdot) \) is the inverse function.

2.2. Rank correlation coefficient

The Copula function connects two or more correlation marginal distribution functions, and the correlation between variables needs to be measured by introducing a correlation coefficient [9]. In this paper, Kendall rank correlation coefficient and Spearman correlation coefficient are used to evaluate the correlation structure between variables.

The calculation formula of Kendall rank correlation coefficient is as follow:

\[ \tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \]  

(6)

Let \( F(x) \) and \( G(y) \) be the marginal distribution functions of continuous random variables \( X \) and \( Y \) respectively, \( C(u, v) \) is a Copula function, and let \( u = F(x) \), \( v = G(y) \) According to the Copula function, \( \tau \) can be expressed as:

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

(7)

According to the Copula function, the Spearman rank correlation coefficient \( \rho_s \) can be expressed as:

\[ \rho_s = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3 = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv \]  

(8)

3. Research on correlation of transformer parameters based on Pair-Copula

3.1. Choose the best Copula

In the actual operation of the transformer, the temperature of the transformer is one of the indicators for testing the health of the transformer, which can intuitively reflect whether the operation of the transformer is in a normal state. In the process of condition assessment, the temperature of the transformer will be affected by the components of the equipment or the surrounding environmental factors, resulting in deviations in the measured temperature. In order to analyze the relationship between transformer temperature and status detection more accurately, this paper collected the voltage, current, active power, reactive power, transformer node temperature, and ambient temperature of the transformer model SF1 025 000/110 from the SCADA system. A feature quantity. The data set in this data sample expresses real-time data generated every 10 minutes in a month.

According to the transformer principle and actual situation analysis, the fault of the equipment itself can be judged by reference values such as transformer voltage, current, and power. Due to the length of space, the current and device temperature are taken as examples for data analysis. The current nuclear density distribution function is shown in Figure 1. The result of the empirical distribution function is close to the kernel distribution estimate. The difference between the two is very small, indicating that the current value tends to be stable. When the current is in the order of 70, there is a small fluctuation, which means that the current value here deviates from the predicted value. The analysis of several other variables is the same as that of current, and it is more efficient to check out abnormal values through this method.
3.2. Parameter estimation
Kendall reflects whether the trend of change between random variables is consistent or not, and Spearman reflects the multiple of the difference between the probability of consistent and inconsistent changes between random variables [10]. Based on the historical data of the transformer temperature and the other five parameters, the maximum likelihood estimation method is used to estimate the five sets of data respectively, and the parameter estimates of the binary normal Copula and the binary t-Copula function are shown in Table 2. In addition, to reduce the length of the article, this paper selects the binary normal Copula and the binary t-Copula function model of transformer temperature and current as shown in Figure 2 and Figure 3.

| Variable             | Copula     | Kendall ($\tau_k$) | Spearman ($\rho_s$) |
|----------------------|------------|--------------------|---------------------|
| transformer temperature-voltage | normal     | 0.6124             | 0.7894              |
| transformer temperature-current | t          | 0.7201             | 0.8315              |
| transformer temperature-active power | normal     | 0.6105             | 0.8112              |
| transformer temperature-reactive power | t          | 0.7317             | 0.8796              |
| transformer temperature-ambient temperature | normal     | 0.4918             | 0.6108              |
| transformer temperature-reactive power | t          | 0.5152             | 0.6913              |
| transformer temperature-ambient temperature | t          | 0.1915             | 0.2493              |
| transformer temperature-ambient temperature | normal     | 0.2240             | 0.3428              |
| transformer temperature-ambient temperature | t          | 0.6899             | 0.7617              |
| transformer temperature-ambient temperature | t          | 0.7127             | 0.8142              |
According to the rank correlation coefficient results in Table 2, the binary normal Copula is better than the binary t-Copula. By following the principle of \( \frac{(\tau_k+\rho_s)}{2} \geq 0.6 \), the relevant variables that do not meet the conditions are removed, and several sets of Copula functions with higher correlation are retained. Therefore, except for reactive power, all other variables meet the conditions. And the degree of correlation with transformer temperature, sorting them in descending order is: current, voltage, active power, and ambient temperature.

4. Conclusion
When the transformer temperature evaluation method can analyze the state of the transformer, the equipment temperature is greatly affected by the operation of the internal components of the transformer or the external environment temperature, and the state of the transformer cannot be accurately evaluated by the threshold method. In addition, the internal structure of the transformer is complex, and a variety of factors are likely to cause the temperature to be abnormally high or low. Therefore, the research on the degree of influence of equipment temperature on equipment status is a part of the current electrical field that cannot be ignored.

For this kind of system with a large number of uncertain factors, this paper takes the advantage of the high flexibility of the Copula function into account, and uses the Pair-Copula function model to analyze the correlation of various parameters in the system. The research results show that the Pair-Copula function model can intuitively reflect the correlation of transformer parameters, and determine the influence factors of transformer temperature and the degree of correlation. It provides a good theoretical basis for subsequent correction of equipment temperature.

Acknowledgment
In the process of studying the evaluation of transformer temperature status, I was fortunate to find this method to assist in temperature correction and evaluation. This research and thesis were completed under the patient guidance of my supervisor, Professor Junjie Yang. The Mr. Yang’s profound professional knowledge and rigorous academic attitude have a great influence on me. Here, I would like to solemnly extend my sincere thanks to supervisor Junjie Yang. I also want to thank my colleagues for accompanying me in my study life. When I encounter problems in the research process, you can always help me find the answers together. Finally, I would like to thank the two major funding projects for their support: the Shanghai Technology Innovation Project (17020500900), and “Shuguang Program” sponsored by Shanghai Education Development Foundation and Shanghai Municipal Education Commission (17SG51) for their support.
References

[1] LI Tao Sun Xuewu Du Xiaoping Song Yuanyuan.(2020) Simulation Analysis of Transformer Flow Field and Temperature Field Based on Finite,35(05):1-5+10.

[2] Chen Lin (China).(2013) The research of bayes assessment model and system development of electrical equipment state based on temperature.https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD&filename=1014015886.nh&v=MDI4OTJxWkViUElSOGVYMUx1eFlTN0RoMVQzcVRyV00xRnJDVV13cWZiK1pzRnlobFVM01WRj2R3JPNUc5bkU=.

[3] Li Tailin Luo Sulin.(2018) Study on the Method of the Differential Correction of the Main Transformer Temperature Control System in 500kV Substation,56(02):17-19.

[4] Wei Bengang, Huang Hua, Fu Chenzhao, Li Honglei, Liu Jiayu.(2012) Algorithm for Transformer Top-Oil Temperature and Winding Hot-Point Temperature Based on Modified Thermal Circuit Model,40(03):444-447.

[5] Yang Xiyun, MAIHEMUTI Mierzhati, Liu Siqu, Wang Qile.(2020) State assessment method of capacity reduction at high temperature for wind turbine based on vine-copula bayesian network model.Proceedings of the CSEE,40(11):3583-3592.

[6] SKLAR, M Paris.(1959) Fonctions de répartition à n dimensions et leurs marges.1959.https://xueshu.baidu.com/usercenter/paper/show?paperid=2466fd8443a0c5014a5bf97b2c7abdf5&site=xueshu_se.

[7] Xie H, Wang K China.(2013) Joint-probability methods for precipitation and flood frequencies analysis.https://xueshu.baidu.com/usercenter/paper/show?paperid=182fc14360bcf cad6c0a2bbf78d82833&site=xueshu_se.

[8] Xu Ke, Liu Yanan, Zhang Feng.(2017) Research on integrated risk measurement of commercial banks based on hybrid copula.Market Modernization,(07):146-147.

[9] Yang Mao, Ma Jian.(2018) Study on wind power data completion method based on copula theory.Electrical Measurement & Instrumentation,55(03):13-19.

[10] Li Jinghua, Wen Jinyu, Cheng Shijie, Wei Hua.(2013) A scene generation method considering copula correlation relationship of multi-wind farms power.Proceedings of the CSEE,33(16):30-36+21.