

Received April 19, 2020, accepted May 5, 2020, date of publication May 12, 2020, date of current version May 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994077

A Novel Fault Prediction Method of Wind Turbine Gearbox Based on Pair-Copula Construction and BP Neural Network

ZHIHONG LUO¹, CHANGLIANG LIU^{1,2}, AND SHUAI LIU^{1,2}

¹School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

²Hebei Technology Innovation Center of Simulation and Optimized Control for Power Generation, North China Electric Power University, Baoding 071003, China

Corresponding author: Shuai Liu (liushuai@ncepu.edu.cn)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4182061, and in part by the Fundamental Research Funds for the Central Universities under Grant 9161717007, Grant 2018ZD05, and Grant 2020MS117.

ABSTRACT Among the various maintenance technologies of wind turbines, online fault prediction technology is a kind of more cost-effective and reliable method. It may also be the most promising method for wind turbines with potential mechanical faults. SCADA data-based online condition monitoring technology has become a hot spot in current researches. Therefore, a novel fault prediction method based on the Pair-Copula model is proposed in this study. First, the conditional mutual information method is introduced to screen out useful variables from a number of variables. Then aiming at the limitation that the conventional Copula model can only deal with two-dimensional variables, the Pair-Copula model is introduced. In addition, the complexity of the prediction model and the dimension of the input variables are greatly reduced by the Pair-Copula model. So, the BP neural network is selected to complete the prediction model. A combined model based on BP neural network and Pair-Copula model is proposed. In order to solve the problem that the conventional Pair-Copula model cannot process real-time data which must be required in fault prediction, a kind of improved Pair-Copula model combined with the kernel density estimation is used to calculate the real-time data. Finally, the proposed method is validated with real data from a 1.5 MW wind turbine, and the effectiveness is confirmed.

INDEX TERMS Wind turbine, gearbox, fault prediction, Pair-Copula, SCADA.

I. INTRODUCTION

According to the structure of the transmission system, wind turbines (WTs) are mainly divided into two categories: doubly-fed WTGs with gearbox, and direct-drive WTGs without gearbox [1]. Because of the WTGs operate in harsh environments for a long time [2] and are affected by factors such as random wind loads, acceleration and deceleration shocks [3]–[6], the gearbox has become one of the key components with a high failure rate [7]. The maintenance cost of the gearbox is also relatively high, the repair time and cost are at the forefront of all major equipment. An effective way to reduce the cost of breakdown maintenance is to use condition monitoring technology for early detection of faults. Therefore, carrying out research on the condition monitoring and fault early warning of WT gearbox, and rationally adjusting the operation and arranging maintenance according to the healthy

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou¹.

degeneration trend of the gearbox are great significance for improving the reliability and reducing the maintenance costs.

The gearbox condition monitoring methods mainly include oil analysis, vibration analysis and SCADA data analysis. The essence of oil analysis is to analyze the physical and chemical properties of lubricating oil. The health state of the gearbox is monitored by indicators such as metal abrasive particles, which can directly reflect its mechanical deterioration and has a high accuracy [8]–[10]. However, the cost of metal abrasive particle sensors is high and the real-time performance of oil analysis method is poor. Thus, It is necessary to further investigate low-cost solutions [11]–[13].

Vibration analysis is to collect vibration signals from the sensitive parts of the gearbox's fault features. The characteristic factors are extracted and analyzed by time-domain and frequency-domain signal processing methods. Vibration analysis has high sensitivity and good real-time performance, and it can perform full life cycle condition monitoring and fault diagnosis on the gearbox. Li Z.X. *et al.*

proposed a periodic potential underdamping stochastic resonance (PPUSR) in the paper and used PPUSR to extract the characteristics of the gearbox vibration signals. The results showed that the proposed method can detect gear wear faults and tooth broken faults [14]. In the research of Li Y.B. *et al.*, the vibration data was filtered by using the Vold-Kalman filter (VKF), and the fault features were extracted by using the refined composite multi-scale fuzzy entropy (RCMFE). Rolling bearing failure was adopted as an example to prove the effectiveness of the VKF-RCMEF method. Experimental results showed that the performance of the proposed method was better than that of RA-RCMFE and VKF-MFE, and it could accurately diagnose inner race fault, outer race fault and ball fault [15]. However, the early fault signal is weaker than when the fault is heavy, so the early fault signal is difficult to detect. Therefore, the above feature extraction method cannot achieve early fault prediction. In response to this problem, Lu L. *et al.* proposed to use deep belief network (DBN) to extract early fault features in vibration signals. Least squares support vector machine (LSSVM) was used to establish a fault prediction model, and early faults of gearbox were detected [16]. But most of the diagnostic methods based on vibration data are after-the-fact diagnosis. So, it is difficult to provide early warning for more than one day if the gearbox has a potential failure. In addition, additional dedicated sensors must be installed to acquire vibration signals, which will cause significant upfront investment and subsequent maintenance costs.

Among wind farms, the supervisory control and data acquisition (SCADA) system is widely adopted to monitor the operating status of WTs and components. The rich data foundation is provided by SCADA system for the research of fault prediction technology. So, mining fault information contained in SCADA data is gradually becoming a current research hotspot. SCADA data-based condition monitoring methods are mainly divided into three categories: The first category is based on temperature monitoring. The second category is based on power curves. The third category is based on machine learning methods.

In the studies of condition monitoring based on SCADA data, many studies take temperature as the monitoring target [17]. In [18], Gearbox oil temperature, nacelle temperature and rotor rotations were used as inputs to predict output power. In [19], rear bearing temperature, active power output, nacelle temperature, turbine speed were used as inputs. Rear bearing temperature was used as output. A normal behavior model was established based on artificial neural network (ANN). In [20], the bearing temperature was used as the predicted target variable, and Time Delay Neural Network (TDNN) was used to establish the generic models. In reference [21], bearing temperature was also selected as the prediction target variable, abnormal level index (ALI) was defined to quantify the abnormal level of prediction error of each selected model, and a fuzzy synthetic evaluation method was used to integrate the identification results. In [22], A kind of health index from the perspective of temperature-related

parameters was developed by separating the statistics concerning the conformity of the predicted values of key temperature parameters within a certain time window from the measured values.

In recent years, there have been many studies based on SCADA data to achieve condition monitoring by comparing power curves. In [23], a method of evaluating the WT performance based on a kernel method was proposed. First, the kernel method was used to estimate the distribution of power data, and then the similarity index was used to evaluate the abnormality of power. In [24], Pandit used the Gaussian Process (GP) to model the power curve, and monitored the abnormal state of the WT by comparing the difference between the normal power curve and the abnormal power curve. In [25], Pandit used GP to model the power curve and realized the detection of absolute yaw error. The experimental results showed that the performance of the proposed method is better than the online power curve model and probabilistic assessment using binning. In [26], Sun Q.L. found that using the rotor speed-power and rotor speed-pitch angle curves can accurately monitor the abnormality of WT. The fault of the pitch system was accurately identified by the distance from the actual operating point to the theoretical curve. Although the functions of fault prediction and diagnosis can be realized through the above methods, fewer variables are used in the methods. A single variable is easily interfered by the external environment, which result in the consequence that a lot of noise is contained. Consequently, the weak fault information will be overwhelmed. The timeliness and accuracy of fault prediction are affected by above problems.

In order to make full use of SCADA data, mine fault information from multiple variables and achieve more accurate fault prediction, a variety of machine learning algorithms have been applied to the condition monitoring of WT. In [27], a WT fault detection method based on expanded linguistic terms and rules using non-singleton fuzzy logic was proposed. In [28], a multivariable power curve model was constructed with a modified Cholesky decomposition Gaussian process (GP) and validated with Supervisory Control and Data Acquisition (SCADA) data. In [29], through irregular space-division and nonlinear space-mapping, stepwise data cleaning procedure was proposed. On this basis, the optimized least squares support vector machine (LSSVM) was selected to model the WT power curve (WTPC). In [30], a new condition monitoring approach was introduced for extracting fault signatures in WT blades by utilizing the data from a real-time SCADA system, and a hybrid fault detection system based on a combination of Generalized Regression Neural Network Ensemble for Single Imputation (GRNN-ESI) algorithm was proposed. In [31], a first attempt to use Dempster-Shafer (D-S) evidence theory for the fault diagnosis of WT on SCADA data was presented. Due to the traditional parameter methods such as artificial neural networks and mixed Gaussian models, there are shortcomings such as poor generalization ability and limited ability to mine complex correlations among variables, which resulting in low

modeling accuracy. Deep learning can make up for the above deficiencies and effectively use massive SCADA data. In [32], Wan proposed a deep feature learning (DFL) approach for wind speed forecasting because of its advantages at both multi-layer feature extraction and unsupervised learning. In [33], a framework based on deep neural network (DNN) was developed to monitor the conditions of WT gearbox and identified the impending failures. In [34], Jiang proposed a novel fault detection method based on denoising autoencoder (DAE). In [35], Wang established a normal behavior model of WTs based on deep belief networks. The optimized modeling method can capture the sophisticated nonlinear correlations among different monitoring variables, which is helpful to enhance the prediction performance. However, the principles of deep learning models at this stage are poorly interpretable and good parameter tuning skills are required to ensure the performance of algorithm, and further research on practical engineering applications is necessary. The disadvantage of deep learning is that the data samples of full fault types and full operating conditions are necessary to train a high-precision prediction model. There are two reasons why long-term data is difficult to save. On the one hand, the storage capacity of the SCADA system is limited. On the other hand, a large amount of data needs to be collected during operation. Therefore, the data samples of full operating conditions and full failure types are scarce. The actual data are mostly small-scale data sets close to the failure time, which makes the research of condition monitoring methods based on small-scale data sets have great significance and application value.

The innovations and main contributions of the proposed method are listed below:

In this paper, the conditional mutual information and Pair-Copula model are introduced to tackle with WT fault prediction. With the powerful variable filtering capability of conditional mutual information, redundant variables can be removed while retaining useful variables to the greatest extent. This method is used to solve the problem of filtering out the model input variables from multiple variables, thereby modeling accuracy and failure prediction accuracy are improved. The Pair-Copula model can handle multiple variables and mine the correlation between multiple variables, the limitation that conventional Copula model can only handle two dimensional variables is overcome. The Pair-Copula model combines three input variables into one variable, which greatly reduces the complexity of modeling. A complete fault prediction model is established based on the combination of Pair-Copula model and BP neural network. In order to solve the problem that the conventional Pair-Copula model cannot process real-time data which must be required in fault prediction, a kind of improved Pair-Copula model combined with the kernel density estimation is used to calculate the real-time data. Finally, the method in [42] has been modified and applied to the determination of fault alarm threshold. This method can continuously update the threshold with change in operating conditions and has better accuracy and flexibility.

The rest of this paper is organized as follows. In Section 2, the proposed method is introduced, and the basic knowledge of each part in this method is explained in detail. Section 3 describes the modeling process and validates the effectiveness of the proposed method with actual SCADA data, the experimental results are discussed in section 4. At last, the conclusions are described in Section 5.

II. METHODS INTRODUCTION

According to Sklar's theorem, the marginal distributions of the random variable $X = [X_1, X_2, \dots, X_n]$ is denoted as F_1, F_2, \dots, F_n and the corresponding joint distribution function is denoted as F , then there is a Copula probability distribution function C such that the equation (1) holds for any $X \in R^n$.

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (1)$$

In equation (1), $[X_1, X_2, \dots, X_n]$ is the input vector and $[x_1, x_2, \dots, x_n]$ is a data point in the vector. n is the number of variables in the vector.

A. PRINCIPLE OF COPULA

Copula functions include the Normal-Copula, t-Copula, and Archimedean-Copulas function cluster. Only the Archimedean-Copulas function cluster can describe asymmetric correlation. The Archimedean-Copulas function cluster includes three functions: Gumbel-Copula, Clayton-Copula, and Frank-Copula. The principle and characteristic of each function are as follows [36]:

Gumbel-Copula:

$$C(U, \theta) = \exp\left(-\left[\sum_{i=1}^n (-\ln u_i)^{\frac{1}{\theta}}\right]^\theta\right) \quad (2)$$

In equation (2), $U = (u_1, u_2, \dots, u_n)$ is the input vector and n is the number of input variables. $\theta = [1, +\infty]$, when $\theta = 1$, the random variables are independent of each other. When $\theta \rightarrow +\infty$, the random variables are completely related. The distribution of Gumbel-Copula function is relatively sensitive to the upper tail correlation among variables.

Frank-Copula:

$$C(U, \theta) = -\frac{1}{\theta} \ln \left[1 + \frac{\prod_{i=1}^n (\exp(-\theta u_i) - 1)}{(\exp(-\theta) - 1)^{n-1}} \right] \quad (3)$$

In equation (3), $\theta = (-\infty, +\infty) \setminus \{0\}$, When θ is a positive number, the random variables are positively correlated. When θ is a negative number, the random variables are negatively correlated. The distribution of Frank-Copula function is relatively average and can well describe the overall correlation among variables.

Clayton-Copula:

$$C(U, \theta) = \left(\sum_{i=1}^n u_i^{-\theta} - n + 1 \right)^{-\frac{1}{\theta}} \quad (4)$$

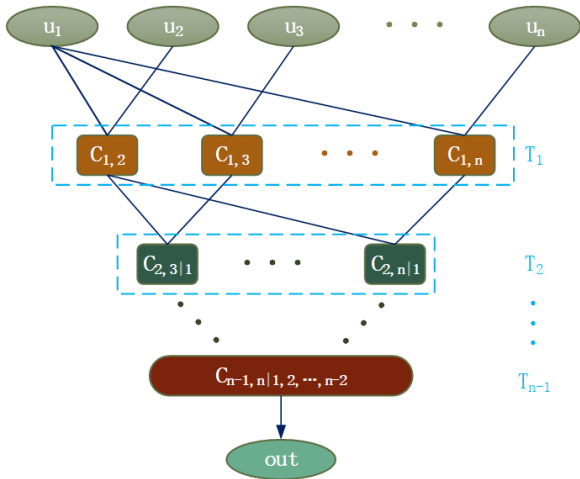


FIGURE 1. The structure of Pair-Copula.

In equation (4), $\theta = (0, +\infty)$, If $\theta \rightarrow 0$, random variables tend to be independent. If $\theta \rightarrow +\infty$, random variables tend to be completely related. Clayton-Copula function has the opposite characteristics of Gumbel-Copula function and is sensitive to the lower tail correlation among variables.

B. PAIR-COPULA CONSTRUCTION

Aiming at the problem that the correlation among WT parameters is complex and difficult to accurately modeled, Pair-Copula is used to model the correlation between multiple parameters.

The Pair-Copula method is first proposed by Aas *et al.* [37]. The n dimensions Pair-Copula model has $n - 1$ layers, and the number of layers is denoted as T_p ($p = 1, 2, \dots, n - 1$). Each layer has a root node which is connected to the other nodes, and the multivariate probability distribution is constructed by merging these nodes hierarchically. The structure is shown in Fig. 1. Each node in the figure is a binary Copula function. u_i is the distribution function value corresponding to the input variable x_i , $u_i = F(x_i)$.

$$c_{1,1+t} = c_{1,1+t}(u_1, u_{1+t})$$

$$c_{j,j+i|1,2,\dots,j-1} \tag{5}$$

$$= c_{j,j+i|1,2,\dots,j-1} \left(F(u_j | u_1, u_2, \dots, u_{j-1}), F(u_{j+i} | u_1, u_2, \dots, u_{j-1}) \right) \tag{6}$$

In equation (5) and (6), $j = 2, 3, \dots, n - 1$ and $t = 1, 2, \dots, n - j$. The structure of Pair-Copula contains multiple nodes corresponding to multiple types of binary Copula functions, which makes the flexibility and fitting accuracy of Pair-Copula model higher than conventional Copula functions [37].

C. FITTING ACCURACY EVALUATION

It is necessary to calculate the fitting accuracy in order to quantitatively evaluate the fitting effect of the probability model. In this paper, Euclidean distance (ED) is used as the evaluation index.

the calculation process of ED is shown in equation (7). One of data samples in the n -dimensional input vector X is denoted as $(x_{1i}, x_{2i}, \dots, x_{ni})$ ($i = 1, 2, \dots, m$). m is the number of data samples contained in X .

$$d_E = \sum_{i=1}^m |C_{em}(F_1(x_{1i}), F_2(x_{2i}), \dots, F_n(x_{ni})) - C(F_1(x_{1i}), F_2(x_{2i}), \dots, F_n(x_{ni}))|^2 \tag{7}$$

C_{em} is the empirical distribution function, and C is the Copula function obtained by training. The ED is smaller, the distribution of the input vector is better fitted by the Copula function.

D. OVERALL FRAMEWORK

The overall framework of this paper is shown in Figure 2. During the training model phase, the Pair-Copula model is used to fit the complex correlation among the three input variables and construct a multivariate joint distribution based on historical SCADA data of normal state. The output value of the Pair-Copula model is called ‘‘state parameter’’. The gearbox bearing temperature at $t - 1$ moment $T_B(t - 1)$ and state parameter are taken as the inputs of BP neural network and the gearbox bearing temperature $T_B(t)$ is used as the prediction target. The BP neural network model is trained with historical SCADA data in normal state.

During the real-time monitoring phase, the Pair-Copula model and the BP model trained in the training model phase are used to predict the gearbox bearing temperature. The real-time SCADA data is used as input. The residual between predicted value and actual value is calculated. Whether the WT has potential faults can be judged by the residual exceeding threshold. Specific steps are as follows:

(1) Variables selection: wind speed (WS), main shaft rotation speed (RS), active power (AP) and gearbox bearing temperature at $t - 1$ moment $T_B(t - 1)$ are selected as input variables, and gearbox bearing temperature $T_B(t)$ is selected as the predicted target.

(2) The Copula function’s types of each node in the Pair-Copula structure are determined, and the optimal model parameters are calculated. The Pair-Copula model of the WT in normal state is trained, and the state parameter is calculated.

(3) BP neural network parameters are determined. State parameter and $T_B(t - 1)$ are taken as the inputs of BP neural network and the gearbox bearing temperature $T_B(t)$ is taken as the prediction target variable, the BP neural network is trained to produce the prediction model of the WT in normal state.

(4) The real-time state parameter can be calculated by inputting the real-time SCADA data into the Pair-Copula model trained in step (2). The predicted value of gearbox bearing temperature can be calculated by inputting the real-time $T_B(t - 1)$ and the real-time state parameter into the BP model trained in step (3).

(5) Calculating the residual of predicted value and actual value.

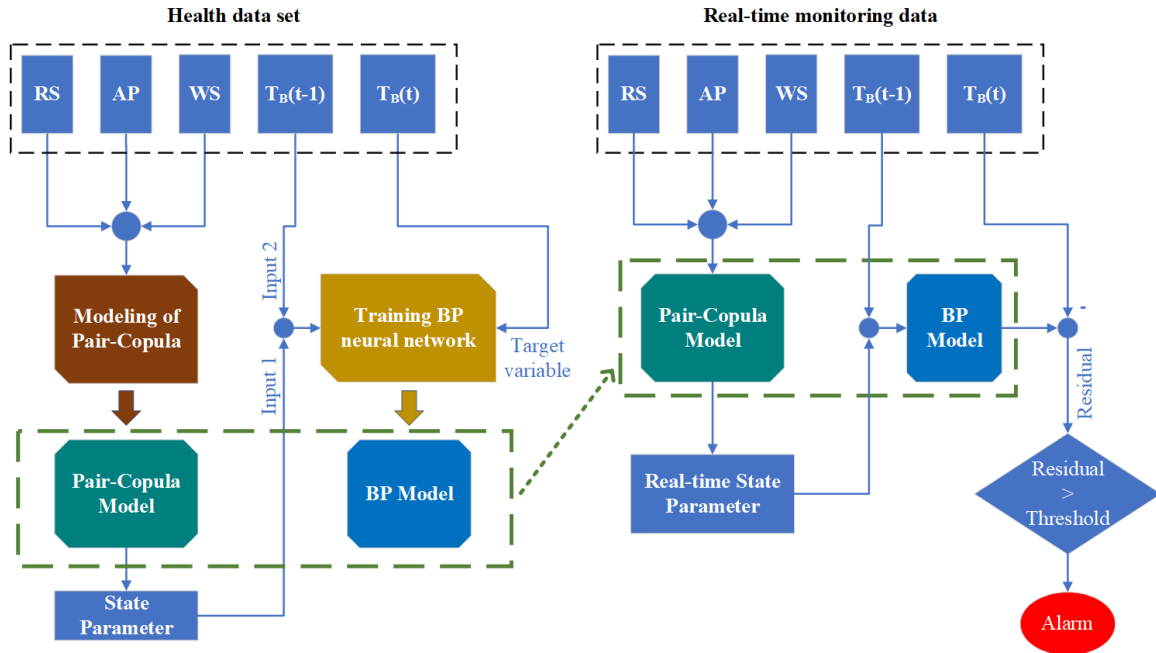


FIGURE 2. Fault prediction framework.

TABLE 1. Introduction of SCADA data.

Type	Total	Used	Training Data	Testing Data	Verifying Data
FDS	14000	9000	3500	1000	4500
HDS	14000	9000	3500	1000	4500

(6) The threshold is calculated from the residual in normal state, and is used to determine whether the gearbox has a potential fault.

III. CASE STUDY

Real SCADA data of a doubly-fed WT in Fujian, China is used to verify the actual effect of the proposed method. The rated power of the WT is 1.5 MW, the cut-in wind speed is 3 m/s, and the cut-out wind speed is 25 m/s. The SCADA system records operational data per 10 minutes. At 10:20 on July 13, the WT was shut down due to a gearbox bearing failure. The SCADA data was divided into two parts. The data of the first part was collected before the shutdown caused by the fault, there are 14,000 data in total, this part is called the fault data set (FDS). The data of the other part was collected after the WT was repaired, a total of 14,000 data, this part is called the health data set (HDS). When the two data sets are used for this research, the samples close to the cut-in wind speed and the samples exceeding the rated wind speed are removed. The number of remaining data samples is 9000. According to the order, 3500 samples are selected as training data samples, 1000 samples as testing data, and 4500 samples as verifying data. The two data sets are allocated in the same way, as shown in Table 1.

A. VARIABLES SELECTION

There are 19 parameters in the SCADA system, as shown in Table 2. The gearbox bearing failure is the actual example in this paper, furthermore, gearbox bearing temperature is the most relevant variable that can directly reflect the health state of the gearbox bearing. Hence, gearbox bearing temperature is selected as the dominant variable in the variable screening process. This is also why gearbox bearing temperature is selected as the predicted target variable in Fig.2. Then the conditional mutual information (CMI) is used to screen out the effective auxiliary variables from the other 18 parameters.

Information entropy is an index to measure the uncertainty of random variables. The information entropy $H(M)$ of random variable M is calculated by follow formula [38]:

$$H(M) = - \sum_{m \in M} f(m) \log f(m) \quad (8)$$

In the equation, $f(m)$ is the probability distribution function of M .

The information entropy of two-dimensional random variable (M, N) is called joint entropy $H(M, N)$, and its calculation formula is as follows:

$$H(M, N) = - \sum_{m \in M} \sum_{n \in N} f(m, n) \log f(m, n) \quad (9)$$

$f(m, n)$ is the joint distribution of (M, N) .

If M is known, the information entropy of N is called conditional entropy $H(N | M)$, and the calculation formula is equation (10):

$$H(N | M) = - \sum_{m \in M} \sum_{n \in N} f(m, n) \log f(n | m) \quad (10)$$

In the equation, $f(n | m)$ is a conditional probability distribution of N .

TABLE 2. Variables in SCADA system.

Variable Types			
Wind Speed	Generator Rotation Speed	Main Shaft Rotation Speed	Wind Direction Angle
Yaw Angle	Gearbox Bearing Temperature	Ambient Temperature	Cabin Temperature
Generator Temperature	Phase A Current	Phase B Current	Phase C Current
Phase A Voltage	Phase B Voltage	Phase C Voltage	Generator Frequency
Reactive Power	Active Power	Power Factor	

The mutual information $I(M, N)$ reflects the amount of information in random variable M containing random variable N . The calculation formula is equation (11):

$$I(M, N) = H(M) - H(N | M) \tag{11}$$

The larger the mutual information, the more N information is contained in M , and the stronger the correlation between M and N .

If the auxiliary variable set is a multidimensional random variable $M = \{M_1, M_2, \dots, M_s\}$, the dominant variable is a one-dimensional random variable N . Then the conditional mutual information between the auxiliary variable M_s and dominant variable N is $I(N; M_s | M_1, M_2, \dots, M_{s-1})$. The calculation formula is as follows:

$$I(N; M_s | M_1, M_2, \dots, M_{s-1}) = H(M_1, \dots, M_s) + H(M_1, \dots, M_{s-1}, N) - H(M_1, \dots, M_s, N) + H(M_1, \dots, M_{s-1}) \tag{12}$$

In this paper, the dominant variable N is determined as gearbox bearing temperature. The auxiliary variables M_1, \dots, M_s correspond to the 18 variables in Table 2 except the dominant variable. According to formula (12), the conditional mutual information between 18 parameters and gearbox bearing temperature can be calculated, $s = 1, 2, \dots, 18$. The variables with conditional mutual information $I > 0.4$ are selected as the effective auxiliary variables. A total of 3 variables are screened, namely main shaft rotation speed, wind speed and active power. In addition, due to the specific heat capacity of the solid, the change of temperature will be affected by the temperature at the previous moment. Therefore, when gearbox bearing temperature $T_B(t)$ is predicted, the influence of the temperature at $t - 1$ moment $T_B(t - 1)$ should be considered.

Summarizing the above, a total of 4 variables (wind speed, active power, main shaft rotation speed, gearbox bearing temperature) are chosen to establish the normal behavior model for WT’s gearbox bearing. Wind speed, active power, main shaft rotation speed, gearbox bearing temperature at $t - 1$ moment $T_B(t - 1)$ are used as input variables, and gearbox bearing temperature $T_B(t)$ is used as the predicted target variable, as shown in Fig.2.

B. ESTABLISHMENT OF PAIR-COPULA MODEL

The modeling process of Pair-Copula model is described in this section. The premise of Pair-Copula model is that the

distribution and probability of input variables are known. However, the distribution of the entire samples cannot be reflected by a small number of data samples. If a large number of data samples are accumulated, the timeliness of fault prediction will be seriously reduced. To solve the above problems, the kernel density estimation (KDE) is proposed to estimate the probability of the real-time input data.

Suppose x_1, x_2, \dots, x_N is a set of normalized wind speed data with N samples. $f(x)$ is the original probability density function of wind speed, and $\hat{f}(x)$ is KDE of $f(x)$, expressed as:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K(h^{-1}(x - x_i)) \tag{13}$$

where, h is the window width; $K(u)$ is the kernel function; x_i is the i -th wind speed data sample.

The Gaussian kernel function is chosen in this paper, that is.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \tag{14}$$

The estimate $\hat{f}(x)$ is expressed by equation (15):

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right] \tag{15}$$

The estimation error is calculated by the root mean square error. The window width h plays a local smoothing role to $\hat{f}(x)$. If h is too small, the estimation bias can be improved but the randomness will be increased, which resulting in an irregular shape of $\hat{f}(x)$. If h is too large, $\hat{f}(x)$ will be too smooth and the detailed features of $f(x)$ cannot be displayed sufficiently [39]. In this paper, the iterative method is used to determine the window width of KDE.

As shown in Fig.3, the probability densities of the three input variables (wind speed, active power, and main shaft rotation speed) are estimated by KDE, and then the three probability densities are input to the Pair-Copula model to calculate the intermediate variable “state parameter” which is a joint distribution of the three input variables and has no practical physical meaning.

If a key component of WT has potential fault, the relationship between the three input variables changes from the normal state, which is reflected in the distance that the three-dimensional joint distribution deviates from the normal state. If the distance is large, it means that the health state of WT

TABLE 3. Correlation coefficients.

Wind Speed - Active Power	Main Shaft Rotation Speed - Wind Speed	Main Shaft Rotation Speed - Active Power
0.8686	0.9915	0.8822

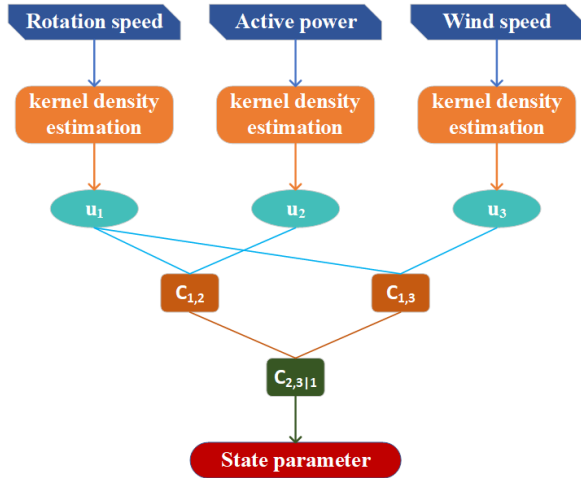


FIGURE 3. The improved structure of Pair-copula.

is worse. If the distance is small, it means that the health state is good. Therefore, although state parameter has no practical physical meaning, it can represent the health state of WT.

Fig.4 is the distribution of state parameter corresponding to HDS and FDS. It can be seen from the figure that when there is potential fault in the WT, the distribution of state parameter will change significantly. It indicates whether the WT is faulty can also be judged by comparing the distribution. However, this method needs to accumulate many data samples for each comparison. The accumulation of data samples takes a long time. The fault may have evolved from a minor fault to a serious fault within this period. Therefore, the distributed comparison method seriously reduces the timeliness of fault prediction and cannot meet the requirements of real-time fault prediction. Aiming at this problem, a combined model of Pair-Copula and BP neural network is proposed, which uses the residual change of predicted value and actual value to realize the real-time fault prediction function. The Pair-Copula model is used to fit the relationship among the three input variables and the three variables are merged into one variable, which is equivalent to that the number of input variables is reduced. The relationship between the inputs and output of the prediction model is simplified. Based on the above analysis, a learning algorithm with strong learning ability and simple structure can meet the modeling need. BP neural network is the basis of all neural networks and has good learning performance. Therefore, BP neural network is selected as the complementary model.

In the rest of this section, the modeling process of Pair-Copula model is depicted in detail. First, wind speed, main shaft rotation speed, and active power are used as inputs to the

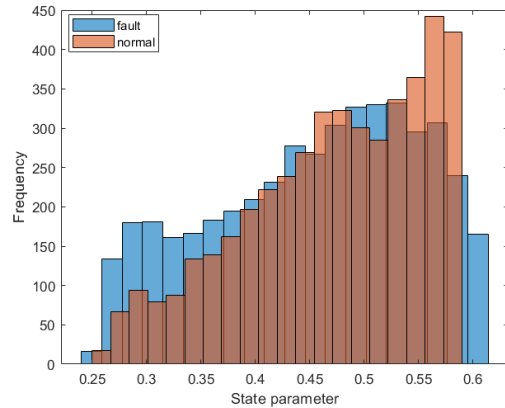


FIGURE 4. The distribution of state parameters in normal and fault states. (The red histogram is translucent, and the darker part is the area where the two histograms overlap.)

TABLE 4. Parameters of the optimized copula function.

Rho	θ_G	θ_F	θ_C
0.9251	13.3552	51.7219	24.7104

Pair-Copula model, and the correlation coefficients between two variables are calculated, as shown in Table 3. Main shaft rotation speed has a strong correlation with wind speed and active power, so main shaft rotation speed is selected as the input variable corresponding to the root node in Fig. 3. The three input variables are combined into one variable (state parameter) by the Pair-Copula model.

To determine the function types of the node Copula functions $c_{1,2}$, $c_{1,3}$, $c_{2,3|1}$ and the corresponding optimal parameters is the key process for establishing the Pair-Copula model.

1) CORRELATION BETWEEN MAIN SHAFT ROTATION SPEED AND ACTIVE POWER

The node $c_{1,2}$ is the connection point between u_1 and u_2 , u_1 corresponds to main shaft rotation speed, and u_2 corresponds to active power. The frequency histogram of main shaft rotation speed and active power is shown in Fig. 5. It can be seen that the correlation is asymmetric. According to the characteristics of the three types of Copula functions in section 2.1, only Gumbel-Copula, Frank-Copula and Clayton-Copula can be used to fit the asymmetric correlation. The correlation between main shaft rotation speed and active power is fitted by the three functions, respectively. Table 4 shows the optimization results of the parameters which are optimized by maximum likelihood estimation.

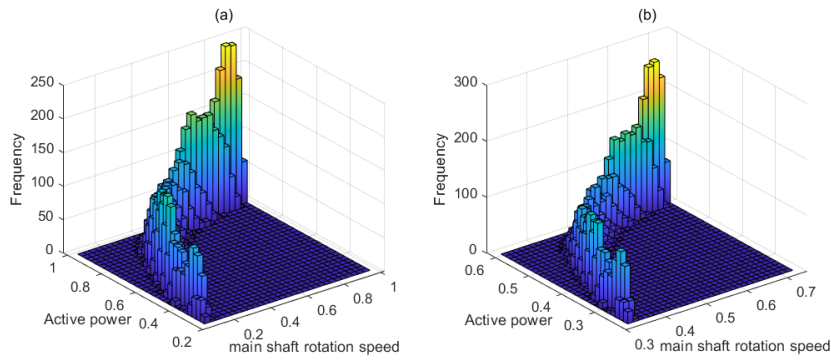


FIGURE 5. Frequency histogram of main shaft rotation speed and active power(a)Raw data (b)Data processed by kernel density estimation.

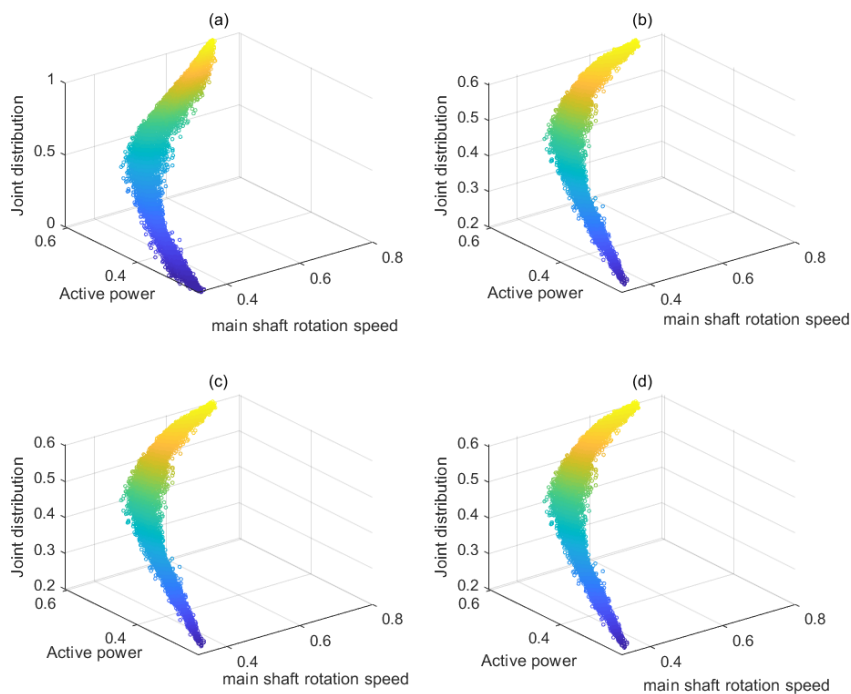


FIGURE 6. Joint distributions(a) Experience Copula; (b) Frank-Copula;(c) Clayton-Copula;(d) Gumbel-Copula.

Before comparing various Copula functions, the definition of empirical Copula function must be introduced. Experience Copula’s definition: Let $(x_i, y_i) (i = 1, 2, \dots, n)$ be a data sample taken from the two-dimensional vector (X, Y) , and the empirical distribution functions of X and Y are $F(x)$ and $G(y)$, respectively. The empirical Copula of u and v is calculated by the following formula [40].

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n I_{[F(x_i) \leq u]} I_{[G(y_i) \leq v]} \quad (16)$$

where $u, v \in [0, 1]$, $I_{[*]}$ is an indicative function; When $F(x_i) \leq u$, $I_{[F(x_i) \leq u]} = 1$, otherwise $I_{[F(x_i) \leq u]} = 0$. The Squared Euclidean Distance (SED) between the above three Copula functions and the empirical Copula function are

calculated separately. The smaller the value of SED, the better the fitting effect. The principle of SED is as follows:

$$d = \sum_{i=1}^n |C_e(u_i, v_i) - C(u_i, v_i)|^2 \quad (17)$$

where n is the number of data samples in the input vector; $C_e(*)$ is empirical Copula function; $C(*)$ is the evaluated Copula function. The value of d is smaller, the distance between fitted distribution and empirical Copula is smaller, the fitting effect is better.

Analysis of Fig. 6 shows that the difference among each figure is small, indicating that the fitting results of the above three functions are very close to the experience Copula. Therefore, it is necessary to further evaluate the fitting effect

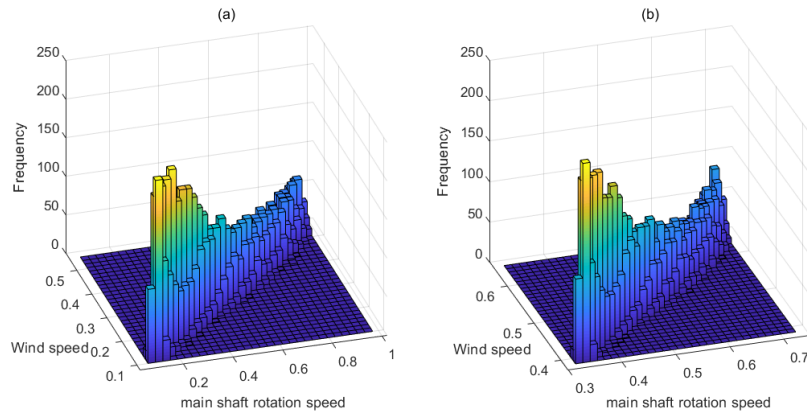


FIGURE 7. Frequency histogram of main shaft rotation speed and wind speed.

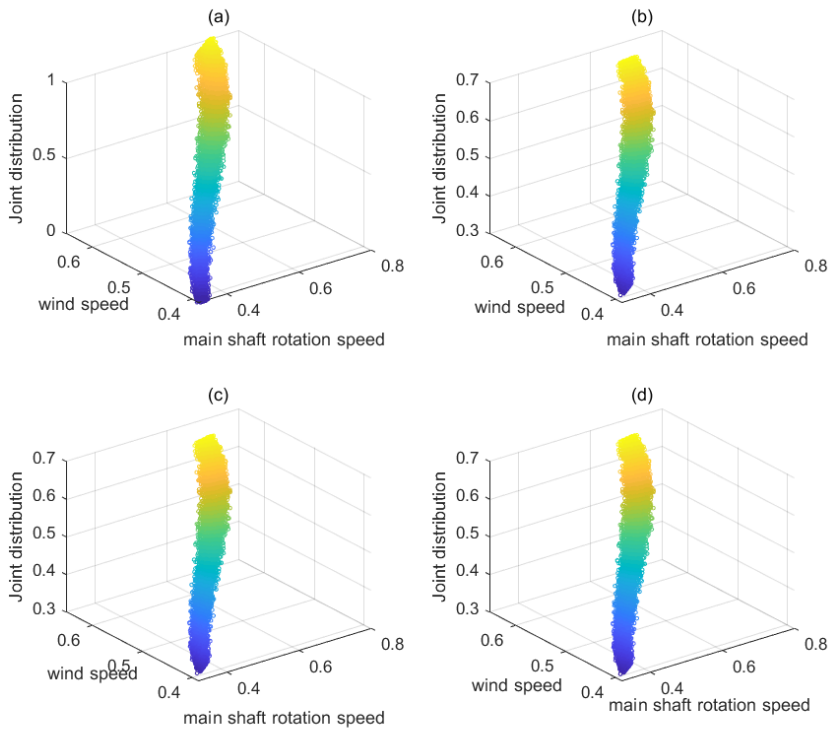


FIGURE 8. Joint distributions(a) Experience Copula; (b) Frank-Copula; (c) Clayton-Copula; (d) Gumbel-Copula.

of the three functions though SED. The corresponding SED values are as follows:

$$\begin{aligned} \text{Gumbel-Copula, } d_G &= 173.4646; \text{ Frank-Copula,} \\ d_F &= 175.6847; \text{ Clayton-Copula,} \\ d_C &= 178.978; \end{aligned}$$

where d_G is the smallest, it can be determined that Gumbel-Copula is the most suitable function to fit the correlation between main shaft rotation speed and active power. It can be seen from Fig. 5 that the distribution of main shaft rotation speed and active power is denser at the upper tail,

and the Gumbel-Copula function is characterized by a better description of the upper tail characteristic. This conclusion is consistent with the analysis result by the SED values.

2) CORRELATION BETWEEN MAIN SHAFT ROTATION SPEED AND WIND SPEED

The Copula function selection process for the node $c_{1,3}$ is similar to part (1). It can be seen from the frequency histogram of Fig. 7 that the correlation between main shaft rotation speed and wind speed at the lower tail is strong, and the correlation is also asymmetric. The fit results of the

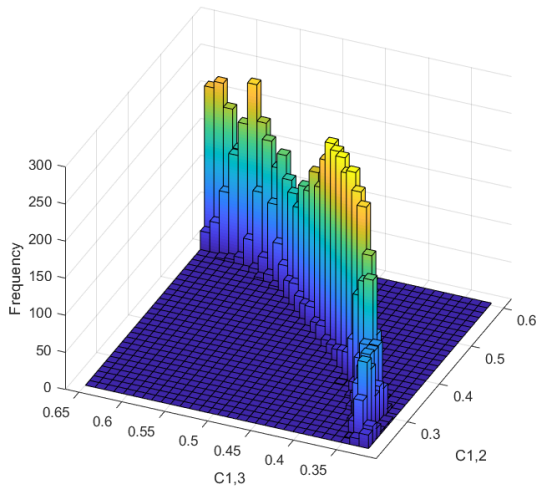


FIGURE 9. Frequency histogram of the first layer nodes $c_{1,2}$ and $c_{1,3}$.

three Copula functions are compared with empirical Copula in Fig.8. The SED values of the three functions are:

$$d_G = 180.807; \quad d_F = 179.4299; \quad d_C = 178.4727$$

The d_C is the smallest, indicating that the Clayton-Copula function has the best fitting effect. According to the characteristic of Clayton-Copula, the lower tail correlation can be described better by the function. The judgment result of the SED is consistent with the characteristic of Fig. 7.

The Copula function of the second layer node $c_{23|1}$ is determined in the same way as in part (1) and part (2). The frequency histogram of the first layer nodes $c_{1,2}$ and $c_{1,3}$ is shown in Fig. 9. The SED values determine that the optimal function of the node $c_{23|1}$ is Frank-Copula.

C. MODEL PERFORMANCE EVALUATION

In this paper, the combination of Pair-Copula model and BP neural network is used as a gearbox fault prediction model for WT. In order to prove the effectiveness of the proposed method, four methods are adopted to establish four models, and their respective performances are compared. The modeling methods and model types are as follows: (1) BP model. Wind speed, main shaft rotation speed, active power and $T_B(t-1)$ are taken as inputs, and gearbox bearing temperature is used as the target variable to train BP neural network and obtain corresponding prediction model. (2) SVM model. Wind speed, main shaft rotation speed, active power and $T_B(t-1)$ are taken as inputs, gearbox bearing temperature is taken as the predicted target variable. The SVM is trained to obtain the corresponding prediction model. (3) C-BP model. The combination model of Pair-Copula model and BP neural network. Wind speed, main shaft rotation speed, and active power are taken as inputs of the Pair-Copula model to calculate the state parameter. Then the state parameter and $T_B(t-1)$ are taken as inputs, and gearbox bearing temperature is taken as output to train BP neural network. (4) C-SVM model. The

TABLE 5. Performance comparison of BP, SVM, C-BP, and C-SVM.

Type of Data	Index	Type of Prediction Model			
		BP	SVM	C-BP	C-SVM
Training Data	RMSE	0.0471	0.0467	0.0426	0.0432
	MAE	0.0360	0.0356	0.0315	0.0327
	R2	0.8783	0.8805	0.9037	0.9016
Test Data	RMSE	0.0577	0.0583	0.0524	0.0509
	MAE	0.0425	0.0429	0.0386	0.0410
	R2	0.8596	0.8568	0.8903	0.8719
Monitoring Data	RMSE	0.0739	0.0730	0.0453	0.0460
	MAE	0.0622	0.0614	0.0406	0.0421
	R2	0.6798	0.6876	0.9051	0.8945

combination model of Pair-Copula model and SVM. The training steps of C-SVM model are same as (3).

In order to evaluate the accuracy of each model, three indexes are introduced, namely RMSE, MAE, and R2. The respective definitions are shown in following formulas [41].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - \hat{o}_i)^2} \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |o_i - \hat{o}_i| \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (o_i - \hat{o}_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (20)$$

where o_i is the i -th actual value, \hat{o}_i is the i -th predicted value, and \bar{o} is the mean of actual values. If the values of RMSE and MAE are smaller, the error between predicted value and actual value is smaller. If the value of R2 is larger, the overall fitting effect is better.

The prediction accuracies of the four models are shown in Table 5. It can be seen from the table, the prediction accuracy of BP is similar to that of SVM, and the accuracy of C-BP is similar to that of C-SVM. Moreover, the prediction accuracies of C-BP and C-SVM are significantly higher than the other two models, which indicates that the addition of Pair-Copula model can improve the prediction accuracy. The main reasons are summarized in two aspects: First, the multi-parameter joint distribution density function model is constructed and the complex correlation between multiple variables is captured by Pair-Copula model. The shortcomings of conventional learning algorithms in capturing complex correlations are compensated, and the prediction accuracy of prediction model is improved. Second, the dimension of the input variables is reduced by Pair-Copula model, which greatly reduces the complexity of the supplementary model. The prediction accuracy is indirectly improved. According to the results in Table 5, the prediction accuracy of C-BP is slightly higher than that of C-SVM, so C-BP is selected as the final prediction model.

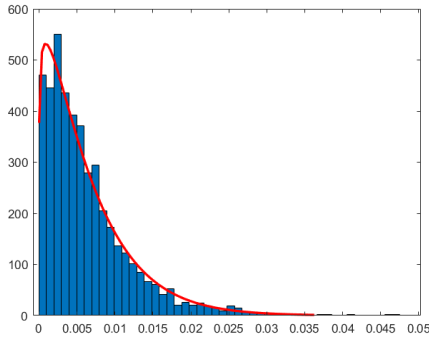


FIGURE 10. Histogram of MD values.

IV. RESULTS

A. THRESHOLD CALCULATION

The MD-Weibull method was proposed in [42] and used for the fault detection in notebook computer. In this paper, the method is modified to make it suitable for the need of this study. The modified MD-Weibull method is used to calculate the threshold.

MD is an unitless distance measurement method that can be used to calculate the degree which a single data sample deviates from the mean of sample set. In this paper, the residual data is used as the measurement object, and the MD value of each residual data sample is calculated. The residual data sequence is denoted as E , and the i -th data sample in the sequence is denoted as e_i , where $i = 1, 2, \dots, n$. n is the number of data samples in the sequence. Before the MD value is calculated, the data needs to be normalized by using formula (21).

$$z_i = \frac{(e_i - \bar{E})}{S_i} \quad i = 1, 2, \dots, n \quad (21)$$

where \bar{E} is the mean of E and S_i is the standard deviation. The calculation method is shown in the following formula:

$$\bar{E} = \frac{1}{n} \sum_{i=1}^n e_i$$

$$S_i = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{E})^2}{n}} \quad (22)$$

Since the residual data is a 1-dimensional vector, the calculation formula of MD can be simplified to following formula:

$$D_i = \sqrt{(e_i - \bar{E})^2} \quad i = 1, 2, \dots, n \quad (23)$$

The MD values of all elements in E are calculated by formula (23), and the MD data set is obtained. It can be seen from the histogram in Fig.10 that the distribution of the MD data set is basically in line with the Weibull distribution, so the 2-parameters Weibull distribution is used to fit the MD data set. The 2-parameters Weibull distribution can be expressed by formula (24).

$$f(D_i) = \beta \eta^{-\beta} (D_i)^{\beta-1} e^{-\left(\frac{D_i}{\eta}\right)^\beta} \quad (24)$$

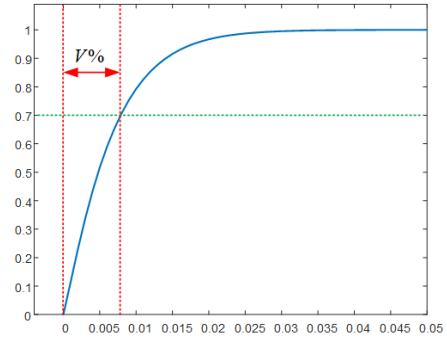


FIGURE 11. CDF curve of Weibull distribution.

where β is the shape parameter and determines the graph form. η is the scale parameter, which determines the spread of distribution. The parameters (β and η) of Weibull distribution can be estimated by using maximum likelihood estimation.

The Weibull distribution's CDF is used to calculate the fault alarm threshold. The CDF is defined as

$$F(D_i) = 1 - e^{-\left(\frac{D_i}{\eta}\right)^\beta} \quad (25)$$

Fig.11 is the CDF curve. $V\%$ in the figure is defined as the normal interval. That is, the residual value within this range is normal, and the residual value outside the range is abnormal. Fig.11 is taken as an example, assuming $V\%$ is equal to 70%, the corresponding MD value D_V can be obtained by the following formula:

$$D_V = \eta[-\ln(1 - V\%)]^{\frac{1}{\beta}}$$

$$v_a = D_V + \bar{E} \quad (26)$$

In formula (26), v_a is the corresponding fault alarm threshold when the normal interval is set to 70%. The normal interval $V\%$ can be set to any value within 0 – 100% according to actual situation.

The three thresholds in Fig.12 correspond to three normal intervals of 70%, 80%, and 90% respectively. Fig.12(a) is the original residual data which is calculated by the C-BP model with HDS. Fig.12(b) is obtained after the original residual data is processed by sliding window method, where the window width $w = 1000$, the number of original residual data $n_o = 4500$, The number of the data after processing $n_p = 3500$, $n_p = n_o - w$. The definition of sliding window method is as follow:

$$e_w(r) = \frac{\sum_{k=i-w}^i e(k)}{w} \quad (27)$$

In formula (27), $i = w, w + 1, \dots, n$, $r = i - w$, e_w is the residual value after sliding window processing. e is the original residual value. w is the window width. n is the number of original residual data. In engineering applications, the system can calculate the normal interval $V\%$ based on the peak value of the residual fluctuation when the WT is in a normal state, and then the alarm threshold is automatically

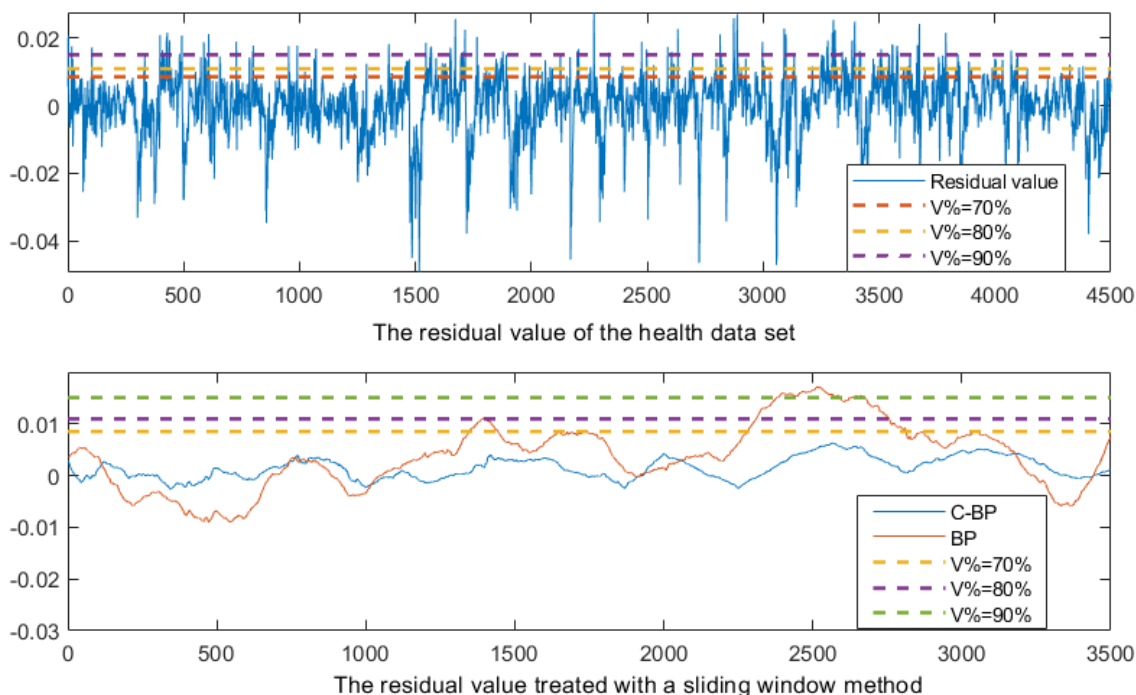


FIGURE 12. Thresholds corresponding to different normal intervals.

calculated. This function can be achieved through simple programming without theoretical difficulty. Since different faults correspond to different monitoring index variables, the corresponding residuals and thresholds are different. In other words, the threshold and the type of failure are related.

In this paper, the normal interval and threshold can be determined by analyzing Fig.12 (b). For the C-BP model, when the normal interval is set to 70%, it can ensure that the residual value of HDS is below the threshold. That is, when the WT is in healthy state, the false alarm will not occur. But the residual fluctuation of the BP model is relatively large. So, a higher threshold needs to be determined. In order to make an unified comparison of the two models, the normal interval is set to 95%, and the corresponding threshold is 0.0187, as shown in Fig.14 and Fig.15.

B. RESULT ANALYSIS

The actual value and predicted value of BP model based on HDS are compared in Fig.13(a). In order to clearly show the fluctuation of curves, 1-500 data points are selected for comparison. It can be seen from Fig.13(a) that although the overall trend of the predicted value and the actual value is consistent, the error between each data point is large and the prediction accuracy is also low. The predicted values and actual values of C-BP model are compared in Fig. 13 (b). By comparing Fig.13 (a) and (b), it can be found that the prediction accuracy of C-BP model is significantly higher than that of BP model, and the prediction error is significantly reduced. This result indicates that the accuracy of prediction model can be effectively improved by Pair-Copula model.

Fig. 14 is the comparison of BP and C-BP prediction residuals based on HDS. The original residual data is processed by the sliding window method. As shown in Fig. 14, there is a large fluctuation in the residual curve of BP model, which seriously affects the result of fault prediction. In comparison, the variation of C-BP is small, and the overall curve is relatively stable. Since the current WT is in normal state, the residual curve should theoretically be stable overall and there is no upward trend. So, the residual curve of C-BP is more authentic and representative.

The residuals of BP and C-BP based on FDS are compared in Fig.15 and the threshold is same as Fig.14. The residual curve of C-BP shows an increasing trend, this is due to the fact that the deviation between predicted value and actual value of the gearbox bearing temperature increases with time. The upward trend of the curve indicates that the gearbox bearing temperature gradually deviates from the normal state, which makes the degradation tendency of the gearbox bearing is intuitively reflected. The residual curve corresponding to C-BP model exceeds the threshold at the 843th data point and then rises continuously, which can be used as an early alarm point for the WT. In contrast, the BP model’s prediction residual curve fluctuates greatly, which cause the fault can not be predicted effectively. Although the curve exceeds the threshold at the 549th point, the hold time is very short and then falls below the threshold. So, this point cannot be used as a valid alarm point. At the position of the 2526th data point, the curve again exceeds the threshold and remains for a longer period of time and then falls below the threshold again. So, the 2,526th data point can barely be used as an alarm point

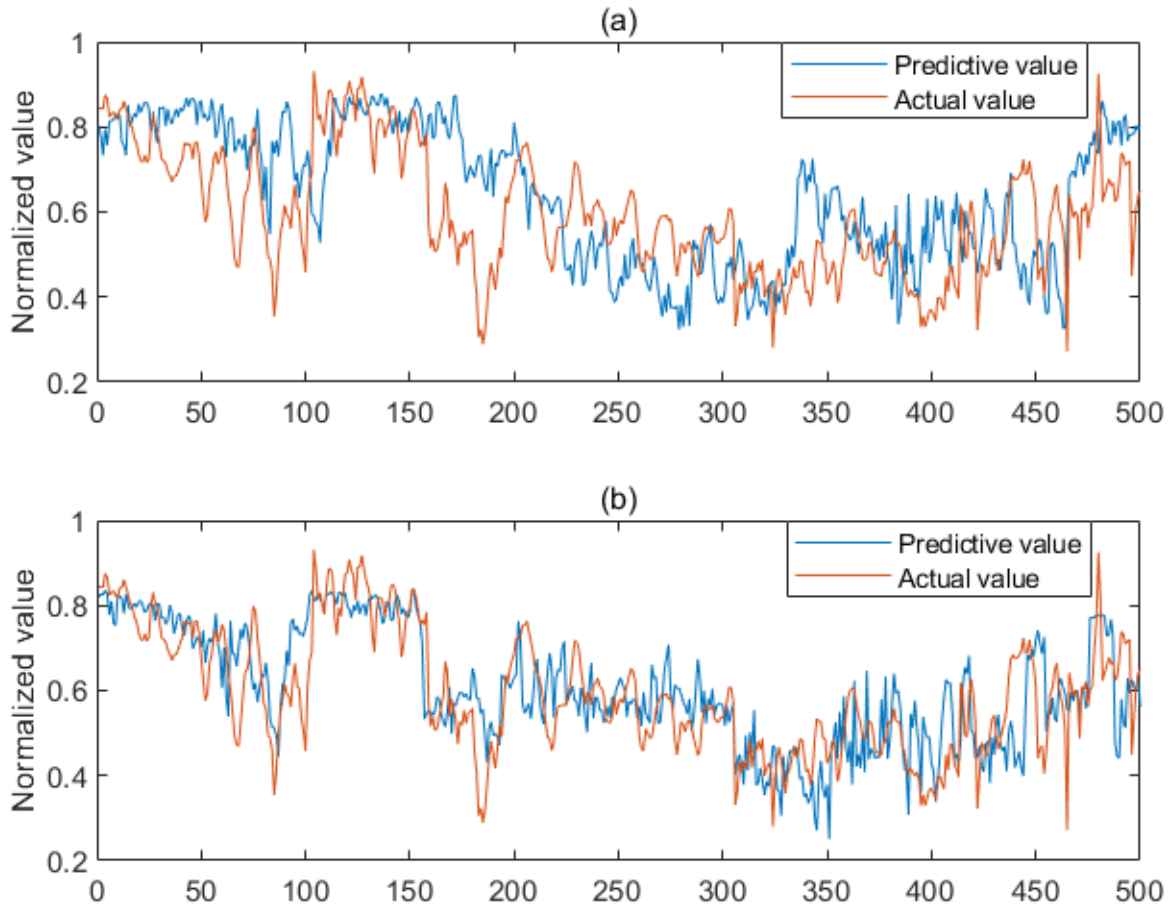


FIGURE 13. Comparison of predicted and actual values:(a)BP model; (b)C-BP model.

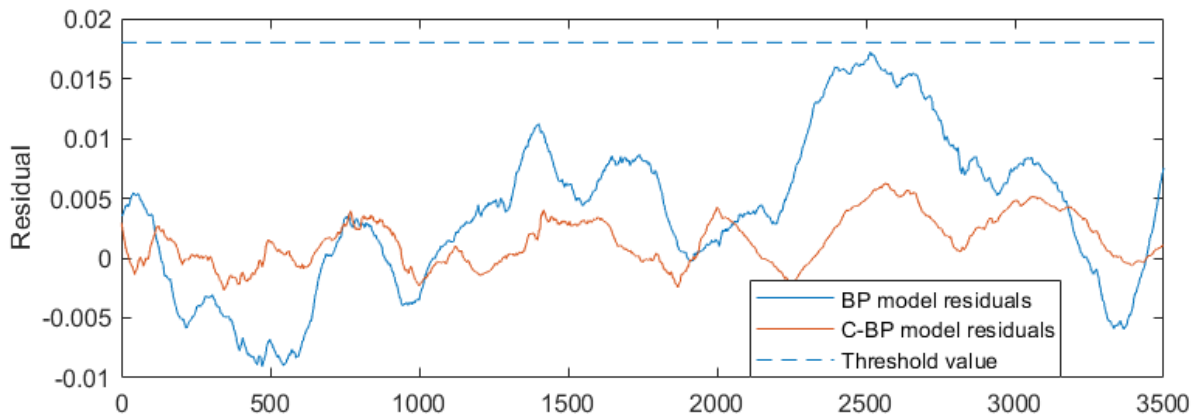


FIGURE 14. Prediction residuals of C-BP and BP corresponding to HDS.

for the BP model. In summary, although both BP model and C-BP model can implement fault prediction, the performance of C-BP model is significantly better than that of BP model in terms of prediction accuracy and timeliness.

The SCADA data sampling interval is 10 minutes. According to the above paragraphs, the alarm point of the C-BP model is the 843th data point. From this point to the WT’s shutdown, there are 2657 sample points, which is converted

into 443 hours, about 18.5 days. The alarm point of the BP model is the 2,526th sample point, there are 974 sample points between the point and the shutdown, which is converted into 162 hours, about 6.8 days. The analysis result in this paragraph shows that the prediction effect of C-BP model is significantly better than that of BP model, which once again proves the effectiveness of the proposed method.

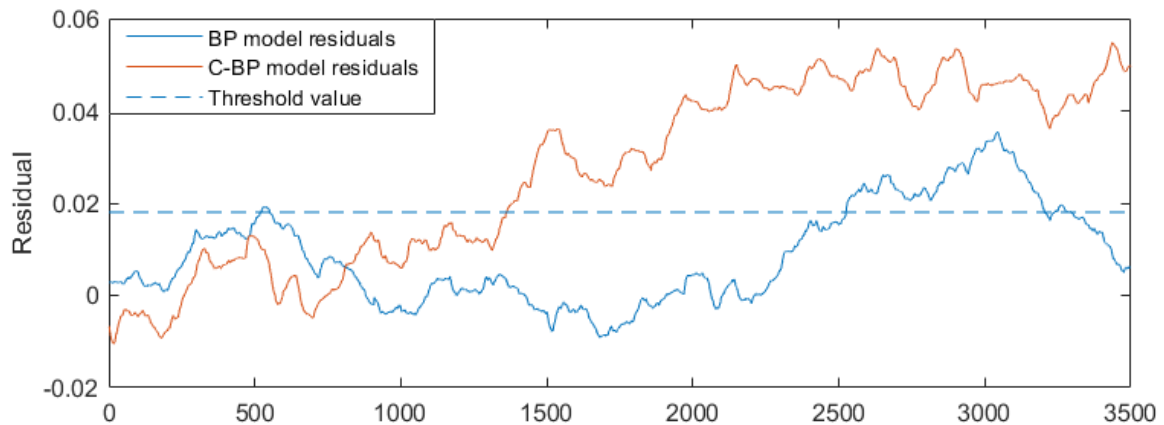


FIGURE 15. Prediction residuals of C-BP and BP corresponding to FDS.

V. CONCLUSION

Based on the Pair-Copula model, a novel WT gearbox fault prediction method is proposed in this paper. The superiority and effectiveness of the proposed method are verified with actual SCADA data. The following is a summary of the paper work:

(1) Conditional mutual information is introduced for variable screening, and 3 variables are selected from 18 variables. Through comparative analysis, it can be known that all 3 variables are effective auxiliary variables. Generator rotation speed is accurately eliminated as a redundant variable corresponding to main shaft rotation speed. The above results show that conditional mutual information can retain useful variables and accurately eliminate redundant variables.

(2) The good performance of Pair-Copula model in mining correlation among multiple variables and establishing multi-dimensional joint distribution is shown. Pair-Copula model is combined with SVM and BP neural network to form two combined models. The experimental results show that the prediction accuracies of the two combined models are significantly higher than the original models, and the key functions of Pair-Copula are fully reflected.

(3) The conventional Pair-Copula model cannot process real-time data, but fault prediction must require the function of real-time calculation. In order to solve this problem, kernel density estimation is adopted to modify the Pair-Copula model, so that it has the function of real-time data calculation.

(4) The proposed method is more suitable for small-scale data samples. Experimental results show that the proposed method has higher prediction accuracy and can identify potential faults earlier than conventional learning algorithms. Finding potential faults early and taking preventive measures can ensure the safe operation of WTs and reduce maintenance costs.

Although the gearbox bearing fault is taken as the case in this paper, the proposed method may also be effective in predicting other faults. It is the follow-up work to further explore the application of the proposed method in other faults. In addition, the research results of reference [43]–[48] show

that air density has a significant effect on the output power of WTs. So, the SCADA data processing method which takes into account the influence of air density is also one of the future research directions.

REFERENCES

- [1] G. A. Skrimpas, C. W. Sweeney, and K. Marhadi, "Detection of wind turbine power performance abnormalities using eigenvalue analysis," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2014, pp. 725–731.
- [2] K. Oh, J. Park, J. Lee, B. Epureanu, and J. Lee, "A novel method and its field tests for monitoring and diagnosing blade health for wind turbines," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 6, pp. 1726–1733, Jun. 2015, doi: [10.1109/TIM.2014.2381791](https://doi.org/10.1109/TIM.2014.2381791).
- [3] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018, doi: [10.1109/ACCESS.2018.2818678](https://doi.org/10.1109/ACCESS.2018.2818678).
- [4] H. Ren, A. Yin, Q. Zhou, J. Li, and Y. Hu, "A wind turbine bearing performance evaluation method based on similarity analysis of fuzzy k-Principal curves in manifold space," *IEEE Access*, vol. 7, pp. 36154–36163, 2019, doi: [10.1109/ACCESS.2019.2903355](https://doi.org/10.1109/ACCESS.2019.2903355).
- [5] J. Fu, J. Chu, P. Guo, and Z. Chen, "Condition monitoring of wind turbine gearbox bearing based on deep learning model," *IEEE Access*, vol. 7, pp. 57078–57087, 2019, doi: [10.1109/ACCESS.2019.2912621](https://doi.org/10.1109/ACCESS.2019.2912621).
- [6] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renew. Energy*, vol. 133, pp. 620–635, Apr. 2019, doi: [10.1016/j.renene.2018.10.047](https://doi.org/10.1016/j.renene.2018.10.047).
- [7] H. Yin, R. Jia, F. Ma, and D. Wang, "Wind turbine condition monitoring based on SCADA data analysis," in *Proc. IEEE 3rd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Oct. 2018, pp. 1101–1105.
- [8] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, and L. Rambault, "Gearbox condition monitoring in wind turbines: A review," *Mech. Syst. Signal Process.*, vol. 111, pp. 251–264, Oct. 2018, doi: [10.1016/j.ymsp.2018.03.052](https://doi.org/10.1016/j.ymsp.2018.03.052).
- [9] G. N. P. Leite, A. M. Araújo, and P. A. C. Rosas, "Prognostic techniques applied to maintenance of wind turbines: A concise and specific review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1917–1925, Jan. 2018, doi: [10.1016/j.rser.2017.06.002](https://doi.org/10.1016/j.rser.2017.06.002).
- [10] S. Liu and C. Liu, "Scale-varying dynamic time warping based on hesitant fuzzy sets for multivariate time series classification," *Measurement*, vol. 130, pp. 290–297, Dec. 2018, doi: [10.1016/j.measurement.2018.07.094](https://doi.org/10.1016/j.measurement.2018.07.094).
- [11] P. Qian, X. Ma, D. Zhang, and J. Wang, "Data-driven condition monitoring approaches to improving power output of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6012–6020, Aug. 2019, doi: [10.1109/TIE.2018.2873519](https://doi.org/10.1109/TIE.2018.2873519).
- [12] A. Agasthian, R. Pamula, and L. A. Kumaraswamidhas, "Fault classification and detection in wind turbine using cuckoo-optimized support vector machine," *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1503–1511, May 2019, doi: [10.1007/s00521-018-3690-z](https://doi.org/10.1007/s00521-018-3690-z).

- [13] Y. Wang, X. Ma, and P. Qian, "Wind turbine fault detection and identification through PCA-based optimal variable selection," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1627–1635, Oct. 2018, doi: [10.1109/TSTE.2018.2801625](https://doi.org/10.1109/TSTE.2018.2801625).
- [14] Z. Li, X. Liu, T. He, and Y. Shan, "A periodic potential underdamped stochastic resonance method and its application for gear fault diagnosis," *IEEE Access*, vol. 7, pp. 141633–141647, 2019, doi: [10.1109/ACCESS.2019.2943191](https://doi.org/10.1109/ACCESS.2019.2943191).
- [15] Y. Li, Y. Wei, K. Feng, X. Wang, and Z. Liu, "Fault diagnosis of rolling bearing under speed fluctuation condition based on vold-Kalman filter and RCMFE," *IEEE Access*, vol. 6, pp. 37349–37360, 2018, doi: [10.1109/ACCESS.2018.2851966](https://doi.org/10.1109/ACCESS.2018.2851966).
- [16] L. Lu, Y. He, T. Wang, T. Shi, and Y. Ruan, "Wind turbine planetary gearbox fault diagnosis based on self-powered wireless sensor and deep learning approach," *IEEE Access*, vol. 7, pp. 119430–119442, 2019, doi: [10.1109/ACCESS.2019.2936228](https://doi.org/10.1109/ACCESS.2019.2936228).
- [17] P. J. Tavner, *Offshore Wind Turbines: Reliability, Availability and Maintenance*. Edison, NJ, USA: IET, 2012, pp. 1–270.
- [18] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 980–987, Mar. 2015, doi: [10.1109/TSG.2014.2386305](https://doi.org/10.1109/TSG.2014.2386305).
- [19] Z.-Y. Zhang and K.-S. Wang, "Wind turbine fault detection based on SCADA data analysis using ANN," *Adv. Manuf.*, vol. 2, no. 1, pp. 70–78, Mar. 2014, doi: [10.1007/s40436-014-0061-6](https://doi.org/10.1007/s40436-014-0061-6).
- [20] M. Rodríguez-López, L. López-González, L. López-Ochoa, and J. Las-Heras-Casas, "Methodology for detecting malfunctions and evaluating the maintenance effectiveness in wind turbine generator bearings using generic versus specific models from SCADA data," *Energies*, vol. 11, no. 4, p. 746, 2018, doi: [10.3390/en11040746](https://doi.org/10.3390/en11040746).
- [21] P. Sun, J. Li, C. Wang, and X. Lei, "A generalized model for wind turbine anomaly identification based on SCADA data," *Appl. Energy*, vol. 168, pp. 550–567, Apr. 2016, doi: [10.1016/j.apenergy.2016.01.133](https://doi.org/10.1016/j.apenergy.2016.01.133).
- [22] X. Wang, Q. Zhao, X. Yang, and B. Zeng, "Condition monitoring of wind turbines based on analysis of temperature-related parameters in supervisory control and data acquisition data," *Meas. Control*, vol. 53, nos. 1–2, pp. 164–180, Jan. 2020, doi: [10.1177/0020294019888239](https://doi.org/10.1177/0020294019888239).
- [23] G. A. Skrimpas, C. W. Sweeney, K. S. Marhadi, B. B. Jensen, N. Mijatovic, and J. Holboll, "Employment of kernel methods on wind turbine power performance assessment," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 698–706, Jul. 2015, doi: [10.1109/TSTE.2015.2405971](https://doi.org/10.1109/TSTE.2015.2405971).
- [24] R. Pandit and D. Infield, "Performance assessment of a wind turbine using SCADA based Gaussian process model," *Int. J. Prognostics Health Manage.*, vol. 9, no. 1, pp. 2153–2168, 2018.
- [25] R. Pandit and D. Infield, "SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes," *IET Renew. Power Gener.*, vol. 12, no. 11, pp. 1249–1255, Aug. 2018, doi: [10.1049/iet-rpg.2018.0156](https://doi.org/10.1049/iet-rpg.2018.0156).
- [26] Q. Sun, C. Liu, and C. Zhen, "Abnormal detection of wind turbine operating conditions based on state curves," *J. Energy Eng.*, vol. 145, no. 5, Oct. 2019, Art. no. 06019001, doi: [10.1061/\(ASCE\)EY.1943-7897.0000612](https://doi.org/10.1061/(ASCE)EY.1943-7897.0000612).
- [27] F. Qu, J. Liu, H. Zhu, and B. Zhou, "Wind turbine fault detection based on expanded linguistic terms and rules using non-singleton fuzzy logic," *Appl. Energy*, vol. 262, Mar. 2020, Art. no. 114469, doi: [10.1016/j.apenergy.2019.114469](https://doi.org/10.1016/j.apenergy.2019.114469).
- [28] P. Guo and D. Infield, "Wind turbine power curve modeling and monitoring with Gaussian process and SPRT," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 107–115, Jan. 2020, doi: [10.1109/TSTE.2018.2884699](https://doi.org/10.1109/TSTE.2018.2884699).
- [29] Y. Hu, Y. Xi, C. Pan, G. Li, and B. Chen, "Daily condition monitoring of grid-connected wind turbine via high-fidelity power curve and its comprehensive rating," *Renew. Energy*, vol. 146, pp. 2095–2111, Feb. 2020, doi: [10.1016/j.renene.2019.08.043](https://doi.org/10.1016/j.renene.2019.08.043).
- [30] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, and M. Saif, "A new hybrid fault detection method for wind turbine blades using recursive PCA and wavelet-based PDF," *IEEE Sensors J.*, vol. 20, no. 4, pp. 2023–2033, Feb. 2020, doi: [10.1109/JSEN.2019.2948997](https://doi.org/10.1109/JSEN.2019.2948997).
- [31] Y. Qiu, Y. Feng, and D. Infield, "Fault diagnosis of wind turbine with SCADA alarms based multidimensional information processing method," *Renew. Energy*, vol. 145, pp. 1923–1931, Jan. 2020, doi: [10.1016/j.renene.2019.07.110](https://doi.org/10.1016/j.renene.2019.07.110).
- [32] J. Wan, J. Liu, G. Ren, Y. Guo, D. Yu, and Q. Hu, "Day-ahead prediction of wind speed with deep feature learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 5, Jun. 2016, Art. no. 1650011, doi: [10.1142/S0218001416500117](https://doi.org/10.1142/S0218001416500117).
- [33] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017, doi: [10.1109/TII.2016.2607179](https://doi.org/10.1109/TII.2016.2607179).
- [34] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 89–100, Feb. 2018, doi: [10.1109/TMECH.2017.2759301](https://doi.org/10.1109/TMECH.2017.2759301).
- [35] H. Wang, H. Wang, G. Jiang, J. Li, and Y. Wang, "Early fault detection of wind turbines based on operational condition clustering and optimized deep belief network modeling," *Energies*, vol. 12, no. 6, p. 984, 2019, doi: [10.3390/en12060984](https://doi.org/10.3390/en12060984).
- [36] Y. Wang, D. G. Infield, B. Stephen, and S. J. Galloway, "Copula-based model for wind turbine power curve outlier rejection," *Wind Energy*, vol. 17, no. 11, pp. 1677–1688, Nov. 2014, doi: [10.1002/we.1661](https://doi.org/10.1002/we.1661).
- [37] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Math. Econ.*, vol. 44, no. 2, pp. 182–198, 2009, doi: [10.1016/j.insmatheco.2007.02.001](https://doi.org/10.1016/j.insmatheco.2007.02.001).
- [38] Y. Wang, X. Yang, and Y. Lu, "Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information," *Insurance: Math. Econ.*, vol. 44, no. 2, pp. 182–198, 2009, doi: [10.1016/j.insmatheco.2007.02.001](https://doi.org/10.1016/j.insmatheco.2007.02.001).
- [39] Q. Han, S. Ma, T. Wang, and F. Chu, "Kernel density estimation model for wind speed probability distribution with applicability to wind energy assessment in China," *Renew. Sustain. Energy Rev.*, vol. 115, Nov. 2019, Art. no. 109387, doi: [10.1016/j.rser.2019.109387](https://doi.org/10.1016/j.rser.2019.109387).
- [40] J. Feng, J. Zhang, and J. Qiu, "Reliability analysis for high-lift device based on copula function and Bayesian updating," *Adv. Mech. Eng.*, vol. 10, no. 12, Dec. 2018, Art. no. 168781401881929, doi: [10.1177/1687814018819297](https://doi.org/10.1177/1687814018819297).
- [41] Z. Liu and L. Zhang, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings," *Measurement*, vol. 149, Jan. 2020, Art. no. 107002, doi: [10.1016/j.measurement.2019.107002](https://doi.org/10.1016/j.measurement.2019.107002).
- [42] G. Niu, S. Singh, S. W. Holland, and M. Pecht, "Health monitoring of electronic products based on mahalanobis distance and weibull decision metrics," *Microelectron. Rel.*, vol. 51, no. 2, pp. 279–284, Feb. 2011, doi: [10.1016/j.microrel.2010.09.009](https://doi.org/10.1016/j.microrel.2010.09.009).
- [43] R. K. Pandit and D. Infield, "Comparative analysis of Gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy," *Renew. Energy*, vol. 140, pp. 190–202, Sep. 2019, doi: [10.1016/j.renene.2019.03.047](https://doi.org/10.1016/j.renene.2019.03.047).
- [44] R. K. Pandit, D. Infield, and J. Carroll, "Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy," *Wind Energy*, vol. 22, no. 2, pp. 302–315, Feb. 2019, doi: [10.1002/we.2285](https://doi.org/10.1002/we.2285).
- [45] F. Guerrero-Villar, R. Dorado-Vicente, M. Fike, and E. Torres-Jiménez, "Influence of ambient conditions on wind speed measurement: Impact on the annual energy production assessment," *Energy Convers. Manage.*, vol. 195, pp. 1111–1123, Sep. 2019, doi: [10.1016/j.enconman.2019.05.067](https://doi.org/10.1016/j.enconman.2019.05.067).
- [46] M. Vahidzadeh and C. D. Markfort, "Modified power curves for prediction of power output of wind farms," *Energies*, vol. 12, no. 9, p. 1805, 2019, doi: [10.3390/en12091805](https://doi.org/10.3390/en12091805).
- [47] J. C. Y. Lee, P. Stuart, A. Clifton, M. J. Fields, J. Perr-Sauer, L. Williams, L. Cameron, T. Geer, and P. Housley, "The power curve working Group's assessment of wind turbine power performance prediction methods," *Wind Energy Sci.*, vol. 5, no. 1, pp. 199–223, 2020, doi: [10.5194/wes-5-199-2020](https://doi.org/10.5194/wes-5-199-2020).
- [48] K. Jung and D. Schindler, "The role of air density in wind energy assessment—A case study from Germany," *Energy*, vol. 171, pp. 385–392, Mar. 2019, doi: [10.1016/j.energy.2019.01.041](https://doi.org/10.1016/j.energy.2019.01.041).



ZHIHONG LUO received the B.S. degree in automation from Inner Mongolia Agricultural University, Hohhot, China, in 2014. He is currently pursuing the Ph.D. degree in control theory and control engineering with North China Electric Power University, Beijing, China.

His current research interests include wind turbine condition monitoring and fault prediction.



CHANGLIANG LIU received the B.S. and Ph.D. degrees in power engineering from North China Electric Power University, Beijing, China, in 1985 and 2002, respectively.

He is currently a Professor with North China Electric Power University. His current research interests include thermal system control in thermal power plants and wind turbine condition monitoring.



SHUAI LIU received the B.S. and Ph.D. degrees in power engineering from North China Electric Power University, Beijing, China, in 2013 and 2019, respectively.

He is currently a Lecturer with North China Electric Power University. His current research interests include wind turbine condition monitoring, fault prediction, and fault diagnosis.

...