

Condition Monitoring of Wind Turbine Based on Copula Function and Autoregressive Neural Network

Zhongshan Huang^{1,2}, Ling Tian^{1,2}, Dong Xiang^{1,2}, Sichao Liu^{1,2} and Yaozhong Wei^{1,2}

¹Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

²Beijing Key Lab of Precision/Ultra-precision Manufacturing Equipment and Control, Beijing 100084, China

Abstract. The traditional wind turbine fault monitoring is often based on a single monitoring signal without considering the overall correlation between signals. A global condition monitoring method based on Copula function and autoregressive neural network is proposed for this problem. Firstly, the Copula function was used to construct the binary joint probability density function of the power and wind speed in the fault-free state of the wind turbine. The function was used as the data fusion model to output the fusion data, and a fault-free condition monitoring model based on the auto-regressive neural network in the faultless state was established. The monitoring model makes a single-step prediction of wind speed and power, and statistical analysis of the residual values of the prediction determines whether the value is abnormal, and then establishes a fault warning mechanism. The experimental results show that this method can provide early warning and effectively realize the monitoring of wind turbine condition.

1 Introduction

Wind energy has played an increasingly important role in the world energy structure in recent years, which results in the rapid development of wind power equipment. Statistically, the installed capacity of global wind turbines has increased from 24 GW in 2001 to 433 GW in 2015. According to the current growth trend, it is expected that the capacity will reach 800 GW by 2020[1]. Wind farms are generally located in remote areas where the working environment is quite complex; therefore, the wind turbines have a high probability of failure. If the key components of the turbine fail, the equipment will be damaged, and even the whole unit may be shut down, causing huge economic losses. Condition monitoring can effectively diagnose the internal failure of wind turbines, which makes it possible to optimize maintenance strategy, reduce unplanned shutdowns, and save operation and repair costs. It can be seen that it is necessary to monitor the operating status of the gearbox and realize fault early warning.

As technology advances, there are diverse statuses monitoring methods for different parts of wind turbines [2]. BP neural network was applied to turbine fault forecasting, and a multi-agent method was proposed to give comprehensive evaluation of the overall operating status [3]. Wavelet analysis was adopted to process vibration signals for fault detection [4]. Nonlinear State Estimate Technology (NEST) was deployed to detect and early-warn the temperature of gearbox bearings in wind turbines [5]. An artificial neural network approach was applied to the gearbox bearing and cooling oil

temperature condition monitoring modelling [6]. Condition monitoring technology has developed greatly, but it is still based on single parameters (vibration, temperature, etc.) at present to detect turbine status and carry out fault early warning. Therefore, it is significant to use the overall correlation between the monitoring parameters of wind turbines to conduct state monitoring research.

There is a close relationship between the power curve of the wind turbine and the faults. At the same time, the power curve is greatly influenced by the wind speed. Therefore, a condition monitoring model based on power and wind speed can be considered as a whole. This paper proposes a wind turbine condition monitoring method based on Copula function and auto-regressive neural network, and uses Copula function to construct a joint probability density function data fusion model of wind speed and power binary monitoring parameters.

2 Theoretical basis

2.1 Copula function

Copula is a mathematical tool that connects the edge distribution and joint distribution of random variables. Copula function described in Eq.1 is defined as a connection function of standard uniform random variables [7].

$$C(u_1, L, u_n) = P_r \{U_1 \leq u_1, L, U_n \leq u_n\} \quad (1)$$

where $U_i : U(0,1), i=1,2,L,n$ is a standard uniform distribution.

The Copula function obeys Sklar's theorem [8]: Let F be the joint cumulative distribution function of an n -dimensional variable. For the edge cumulative distribution F_1, F_2, L, F_n , there is an n -dimensional Copula function C such that

$$F(x_1, x_2, L, x_n) = C(F_1(x_1), F_2(x_2), L, F_n(x_n)) \quad (2)$$

If F_1, F_2, L, F_n are continuous functions, then function C exists uniquely. Otherwise, the Copula function C is uniquely determined only within the range of the edge cumulative distribution function. The density function in Eq.3 can be obtained through the derivation of Eq.2

$$f(x_1, L, x_n) = c(F_1(x_1), L, F_n(x_n)) \prod_{i=1}^n f_i(x_i) \quad (3)$$

where $c(u_1, u_2, L, u_n) = \frac{\partial^n C(u_1, u_2, L, u_n)}{\partial u_1 \partial u_2 L \partial u_n}$, $u_i = F_i(x_i)$, and $f_i(x_i)$ is the edge probability density function of x_i .

Archimedes Copula is widely used because of its simple structure and good nature. The commonly used Archimedes Copula functions have Gumbel Copula functions, Clayton Copula functions, and Frank Copula functions. Gumbel Copula is sensitive to the tail-tailed correlation, Clayton Copula is sensitive to the tail-tail correlation, and Frank Copula is sensitive to the symmetry of the variable. Let u, v be a random variable and θ be a related parameter, then the analytical formulas of the commonly used three types of binary Archimedean Copula are shown in Table 1.

Table 1. Archimedean copula functions.

Copula function	$C(u, v \theta)$
Clayton	$(u^{-\theta} + v^{-\theta})^{-1/\theta}$
Frank	$-\frac{1}{\theta} \ln \left\{ 1 + \frac{[e^{-\theta u} - 1](e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\}$
Gumbel	$\exp \left\{ - \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{1/\theta} \right\}$

2.2 Autoregressive neural network

Some signals have strong autocorrelation, and their future time values can be predicted by learning signal autocorrelation features. Commonly used time series modelling methods include auto-regressive moving average (ARMA) models, artificial neural networks (RNN) and other methods. In this paper, a non-linear auto-regression (NAR) dynamic neural network is introduced to model the monitoring signal time series. Generally, the static neural network processes input signals from the input layer and then goes through each

layer in order, and finally outputs the output results through the neurons in the output layer. The NAR and other dynamic networks can feed back the output signal to the input end, so that the output signal can participate in the next iterative training with memory function, so it can better describe the characteristics of time-varying systems with non-stationary, nonlinear and other complex mapping relationships [9], and overcome the shortcomings of the ARMA model that can only be modelled for stationary linear signals.

The NAR neural network can be defined as shown in Eq.4

$$\hat{y}(t+1) = f[y(t), y(t-1), L, y(t-n_y)] \quad (4)$$

where $f(\mathcal{g})$ is a nonlinear mapping, $y(t), y(t-1), L, y(t-n_y)$ is the historical time series of the monitoring signal, and $\hat{y}(t+1)$ is a prediction value.

2.3 Establishing fault-free monitoring model

2.3.1 Determination and selection of Copula functions

Determining the edge distribution function of each variable is the first step in determining the Copula function, so firstly an edge distribution function for wind speed and power is built. In most cases, the monitoring signal does not completely obey the Gaussian distribution, so the Gaussian distribution cannot be used to fit the data. This paper uses the kernel density estimation (KDE) nonparametric estimation method to fit the edge distribution. Subsequently, the maximum likelihood estimation method was used to solve the parameters of the Copula function to determine the Archimedean Copula function expression. Then select the appropriate Copula function type based on the binary frequency histogram.

2.3.2 Establishing condition monitoring model based on NAR network

First determine the input data for the NAR neural network. In this paper, based on the joint probability density function of wind speed and power, it is used as a data fusion mapping model. The model is used for data fusion of multivariate parameters to obtain a new fusion data vector, which are denoted as $y_t = [y(1), y(2), L, y(t)]$.

Using the new data vector y_t as the input of the NAR neural network, the operational status of the system is described in terms of wind speed and power.

After using the Copula function to fuse the wind speed and power data in the fault-free state, the fault-free fusion data can be obtained, which is denoted as $y_{t, norm}$, then it is used as the input vector of the NAR network. After training, a single step prediction model under normal conditions can be established. This model is called a faultless condition monitoring model.

Furthermore, the single-step prediction residuals of training sample were obtained using the no-fault monitoring model, and the distribution of residual samples was analysed using statistical principles. After training, after the NAR network fully extracts the information of the training sample, the residual sample generally approximately obeys the normal distribution. After training, the NAR network can generally fully extract the effective information of the training sample, so the residual sample generally approximately obeys the normal distribution. Consider the residual as a random variable A , and then in the no-fault state, there should be $X_e : N(\mu, \sigma)$, where μ and σ are the mean and standard deviation of the residual sample.

Finally, using the residual distribution law to monitor the running status of the system in real time, it is judged whether the single-step prediction residuals of the non-fault state monitoring model obey the residual distribution law. If the single-step prediction residual is deviating from the distribution of residual, it is considered that the system has an early fault signal, the monitoring model will carry out fault early warning.

The above process is the so-called state monitoring process based on Copula function and autoregressive neural network. The process can be described as

$$\begin{cases} \mathbf{u}_t = F(\mathbf{x}_1), \mathbf{v}_t = F(\mathbf{x}_2) \\ \mathbf{y}_t = C[\mathbf{u}_t, \mathbf{v}_t] \\ \hat{y}(t+1) = f(\mathbf{y}_t) \\ e(t+1) = y(t+1) - \hat{y}(t+1) \\ X_e = \{e_1, e_2, \dots, e_N\} \\ X_e : N(\mu, \sigma) \end{cases} \quad (5)$$

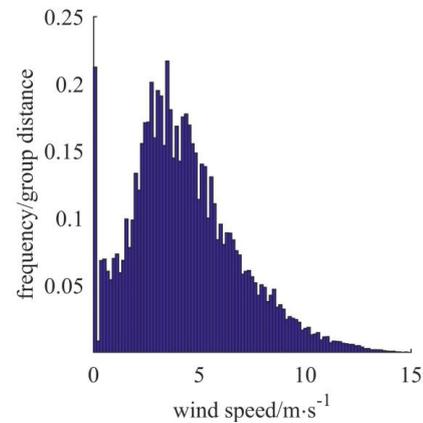
where $F(\mathbf{x}_1)$ and $F(\mathbf{x}_2)$ are the edge cumulative distribution of wind speed and power, C is the binary Copula joint probability density function, \mathbf{y}_t is the output vector of Copula function, and f is the NAR neural network mapping.

3 Application verification and analysis

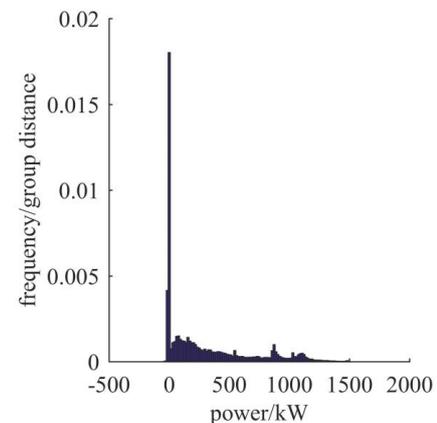
This paper uses the wind speed and power data collected by the wind turbine Supervisory Control and Data Acquisition (SCADA) system as an application verification to verify the effectiveness of the proposed method. The sampling interval for wind speed and power is 10 s. The acquisition time is from 1:00:10 on November 15, 2015 to 11:59:00 on December 2, 2015. Among them, at the time of 11:45:00 on December 2, 2015, the SCADA system showed that a right yaw reaching limit fault occurred in the system. This article will use 8640 data sets from 11:59:00 on December 1st, 2015, to 11:59:00 on December 2nd, 2015 as a test sample. This data contains failure data; use the remaining data as a training sample training model. This article will use a total of 8640×2 sets of data from 11:59:00 on December 1st, 2015 to 11:59:00 on December 2nd, 2015 as a test sample, which contains fault data of a right yaw

reaching limit fault. The remaining data are used as a training sample to train the NAR model.

The frequency histograms of the two datasets are shown in Figure 1(a) and Figure 1(b), obviously, neither of them fully obeys the normal distribution. The wind speed and power are converted into uniform distributions u and v on $[0, 1]$. The binary frequency histogram is shown in Figure 2. It can be seen that the graph has obvious symmetry, so the Frank Copula function was chosen to fit the model.



(a) Wind speed



(b) Power

Figure 1. Wind speed and power frequency histogram.

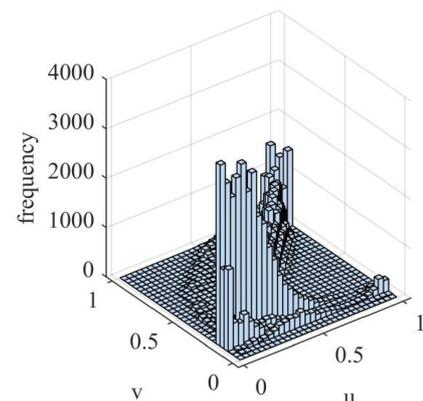


Figure 2. Binary frequency histogram.

Using KDE non-parametric estimation method to fit the edge distribution of wind speed and power data, and then using the maximum likelihood estimation method to obtain the parameter of the Frank Copula function as $\theta=11.706$, then the joint probability density function of the binary Frank Copula is

$$C(u, v; \theta) = -\frac{\ln \left[1 + \frac{(e^{-11.706u} - 1)(e^{-11.706v} - 1)}{e^{-11.706} - 1} \right]}{11.706} \quad (6)$$

Enter u, v into the joint probability density function of Eq. 6 to find the time series y_i of fusion sample points based on the Copula function. The time series of wind speed, power, and fusion data are shown in Figure 3.

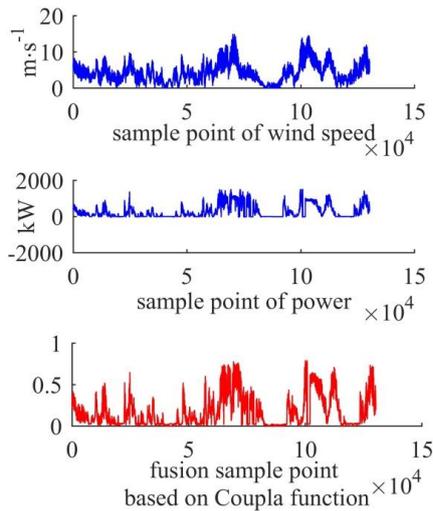


Figure 3. Wind speed, power and fusion data.

The fusion data sample is used as a training set, and further divided into sub-training sets, sub-validation sets, and sub-test sets, respectively accounting for 70%, 15%, and 15% of the total fusion data samples, and the NAR model is trained using the data set. The training error goal is set as 5×10^{-4} . After several parameter optimizations, the hidden layer of the NAR neural network selected in this paper is 30, and the delay order is 10. The training process shows that the convergence step is up to 4 steps, which satisfies the error requirements. The model establishment speed is fast and all parameters of the network meet the requirements.

The statistical analysis of the single-step prediction residual random variables of the training samples in the no-fault state can be considered as approximately obeying the normal distribution, and there is $X_e : N(-1.89 \times 10^{-4}, 0.0211)$.

Through the above training, a fault-free monitoring model based on the NAR neural network in a fault-free state can be obtained. The model is used to perform online monitoring of the test set containing the fault signal. By judging whether the residual of the single-step prediction value obeys the normal distribution, it can be judged whether or not the failure has occurred. The data of wind speed and power in the test set are input into the fault-free condition monitoring model to obtain the

residual sequence of the test set. The time series of the monitoring parameters and the residuals are displayed synchronously as shown in Figure 4.

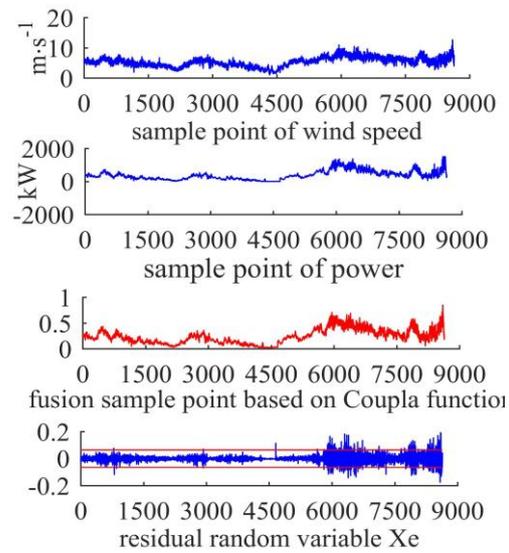


Figure 4. Parameters simultaneously displayed.

In Figure 4, the residual values of the single-step prediction clearly exceeds the confidence interval of $[\mu - 3\sigma, \mu + 3\sigma]$. In order to overcome the effects of extreme errors, an abnormal signal is considered to be present only if there is a continuous period of time exceeding the threshold and the sample point satisfying this condition appear at 3:40 on December 2, 2015. By comparison, it can be found that the fault early warning time is 8 hours earlier than the fault recording time of the SCADA system. The results show that the fault-free monitoring model proposed in this paper can effectively monitor the early abnormal signal before the failure occurs compared with SCADA system.

In order to fully illustrate the effectiveness of the method, it is applied to the actual monitoring to observe whether more faults can be monitored. After a period of testing, it is found that the method can effectively monitor the following faults: the left yaw reaching limit fault, the gearbox oil temperature overrun fault, and the generator bearing temperature overrun fault. Compared with the fault alarm time of the SCADA system, the prediction time of the three kinds of faults in advance is shown in Table 2. The residual error sequences of the above three kinds of faults are shown in Figure 5.

Table 2. Various types of fault early warning time

Type of fault detected	Early warning time / h (compared to SCADA system)
Left yaw reaching limit	4.3
Gearbox oil temperature overrun	16.2
Generator bearing temperature overrun	1.3

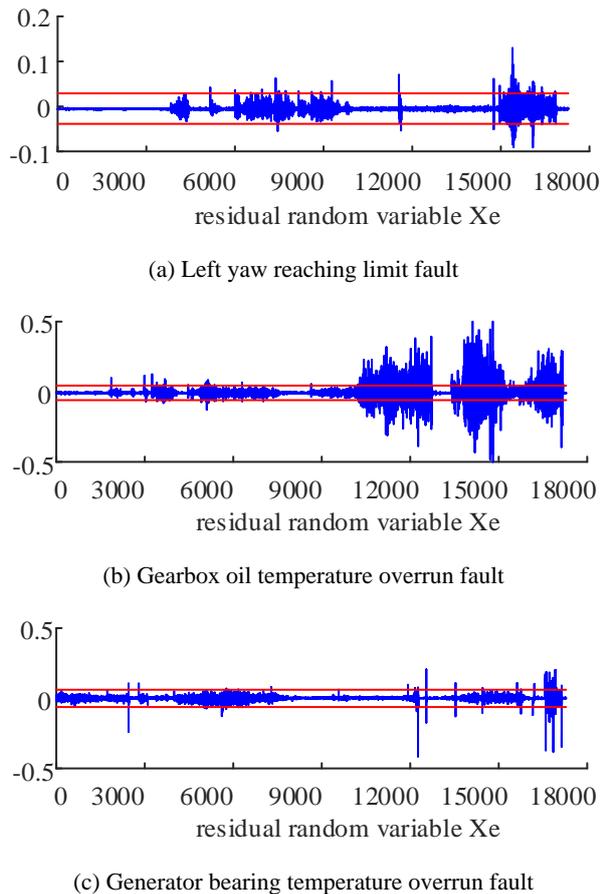


Figure 5. Various types of fault residual sequences.

As can be seen from the Table 2, this method can give an alarm earlier than the SCADA system, which is of great significance for the condition monitoring activities of wind turbines. The analysis proves that the method proposed in this paper has a high state monitoring and early warning capability.

4 Conclusions

This paper makes use of the wind speed and power parameters of wind turbines to monitor the status of the

system. A condition monitoring method based on Copula function and autoregressive neural network is proposed. The final application verification shows that the warning time of this method is 8 hours earlier than the SCADA monitoring system, which proves that this method can effectively perform state monitoring and abnormal alarm for wind turbines.

References

1. Global Wind Report 2015: Annual Market Update [R]. Global Wind Energy Council, 2016.
2. CJ Crabtree, Y Feng, PJ Tavner. Detecting Incipient Wind Turbine Gearbox Failure: A Signal Analysis Method for On-line Condition Monitoring [J]. *Journal of Organic Chemistry*, 2010, **75**(18): 6122-6140.
3. A Zaher, SDJ McArthur, DG Infield, et al. Online wind turbine fault detection through automated SCADA data analysis [J]. *Wind Energy*, 2009, **12**(6): 574-593.
4. W Yang, PJ Tavner, CJ Crabtree, M Wilkinson. Cost-Effective Condition Monitoring for Wind Turbines [J]. *IEEE Transactions on Industrial Electronics*, 2010, **57**(1): 263-271.
5. P Guo, D Infield, X Y Yang. Wind turbine gearbox condition monitoring using temperature trend analysis [J]. *Proceedings of the CSEE*, 2011, **31**(32): 129-136.
6. A Zaher, SDJ McArthur, DG Infield, et al. Online wind turbine fault detection through automated SCADA data analysis [J]. *Wind Energy*, 2009, **12**(6): 574-593.
7. RB Nelsen. *An Introduction to Copulas* (Springer Series in Statistics) [M]. 2006: xx,315.
8. M Sklar. *Fonctions de repartition a n dimensions et leurs marges* [J]. *Publinststatistunivparis*, 1959, 8.
9. M Mordjaoui, S Haddad, A Medoued, A Laouafi. Electric load forecasting by using dynamic neural network [J]. *International Journal of Hydrogen Energy*, 2017.