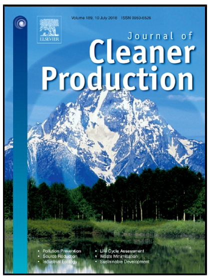


An Anomaly Identification Model for Wind Turbine State Parameters

Yiyi Zhang, Hanbo Zheng, Jiefeng Liu, Junhui Zhao, Peng Sun

PII:	S0959-6526(18)31463-X
DOI:	10.1016/j.jclepro.2018.05.126
Reference:	JCLP 12983
To appear in:	*Journal of Cleaner Production*
Received Date:	07 January 2018
Accepted Date:	16 May 2018

An Anomaly Identification Model for Wind Turbine State Parameters

Yiyi Zhang [1,†], Hanbo Zheng[1,3,*,†] ,Jiefeng Liu[1,4,*,†], Junhui Zhao[2], Peng Sun[3]

1 Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning 530004, P.R. China

2 Department of Electrical and Computer Engineering & Computer Science, University of New Haven, West Haven, CT 06516, USA

3 State Grid Henan Electric Power Research Institute, Zhengzhou, Henan, 450052, P.R. China

4Shijiazhuang Power Supply Branch of State Grid Electric Power Company, Shijiazhuang 050093, P.R. China

†*These authors contributed equally to this work.*

Co-corresponding author: Hanbo Zheng, Jiefeng Liu

Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning 530004, P.R. China

E–mail: hanbozheng@163.com (Hanbo Zheng), liujiefeng9999@163.com (Jiefeng Liu)

An Anomaly Identification Model for Wind Turbine State Parameters

Abstract

Identifying the anomalies of wind turbine (WT) and maintaining in time will improve the reliability of wind turbine and the efficiency of energy use, however it is difficult toidentify the wind turbine's abnormal operation by the traditional threshold settings because the anomalies can be induced by multiple factors.Therefore, this paper presents an anomaly identification model for wind turbine state parameters,and the model can identify abnormal state which the fluctuation range of the condition parametersis within the SCADA alarm threshold. The main work is as follows: 1) in order to increase the accuracy of the prediction model, a novel BPNN model integrated genetic algorithm (GA) was employed to optimize the training method (called GABP method), data samples, and input parameter selection, respectively; 2) on this basis, the distribution characteristics of state parameter prediction errors were depicted by a T-location scale (TLS) distribution with the shift factor and elastic coefficient; 3)error abnormal index (EAI) is defined to quantify the abnormal level of the prediction error, which is used as an indicator of the wind turbine anomaly. The proposed method has been applied on areal 1.5 MW wind turbine, and the analysis shows that the proposed method is effective in wind turbine anomaly identification.

Keywords: Wind turbine; Anomaly identification; State parameters; Back propagation neural networks (BPNN); T-location scale distribution; Error abnormal index

1. Introduction

Wind power has attracted global attention in recent years as a clean and renewable energy generation (Li et al., 2017). Wind turbines (WTs) are an emerging renewable energy technology that have the potential to provide low carbon intensity power in the future (Cruz and Martín, 2016; Demir and Taşkın, 2013; Li et al., 2016; Ortegon et al., 2013). Rapid developments of wind energy in recent years have drawn attention to issues of operation and maintenance (O&M) of wind farms (Li and Chen, 2013). Andthe detection of wind turbine faults is considered to be a cost-effective approach to improve the reliability of WTs and reduce the O&M costs of the wind farms (Li et al., 2012).

Resently, the detection of wind turbine faults becomes a hot problem. In the study (Kusiak and Verma, 2012a; Kusiak and Li, 2010), three wind turbine condition parameters, including a main bearing temperature, a lubrication oil temperature of the gearbox, and the winding temperature of the generator, were modeled in a back propagation neural network (BPNN) for the fault detection of WTs based on SCADA data (Zaher et al., 2010). How to utilize the BPNNs to model the wind turbine parameters with SCADA data was also investigated in publications (Garcia et al., 2006a; Lapira et al., 2012; Schlechtingen and Santos, 2011a; Stickland, 2012; Xiang et al., 2009). A comparative analysis of two BPNN-based models and a regression-based model was presented (Schlechtingen and Santos, 2011a) for modeling parameters of gearbox bearing temperature and generator stator temperature. Besides, certain thresholds of prediction errors are usually set to identify the anomalies in the WTs (Kusiak and Verma, 2012b; Schlechtingen et al., 2013; Schlechtingen and Santos, 2011b). Intelligent anomaly identification systems, such as the multi-agent system (Zaher and Mcarthur, 2007) and SIMAP (Garcia et al., 2006b) were developed using the prediction models of the wind turbine condition parameters such as gearbox bearing temperature, gearbox oil temperature, and generator winding temperature. However, most of the previous studies identify the wind turbine's abnormal operation by the traditional threshold settings, and it is difficult to identify abnormal state which the fluctuation range of the condition parameters is within the SCADA alarm threshold(Chandola et al., 2009; Sun et al., 2016; Sun et al., 2016).

To solve the problem, the paper developed an anomaly identification model for the wind turbine state parameters. The main work is as follows. First, in order to increase the accuracy of the prediction model, a novel BPNN model integrated genetic algorithm (GA) was employed to optimize the training method (called GABP method), data samples, and input parameter selection, respectively. On this basis, the distribution characteristics of state parameter prediction errors were depicted by a T-location scale (TLS) distribution with the shift factor and elastic coefficient. After that, the estimated residual anomaly intensity of the turbine state parameter was quantized and indicated. Finally, the proposed method has been applied to a real 1.5 MW wind turbinefor verification.

2. General anomaly identification model of WT state parameters

A prediction model is set up to identify the turbine's abnormal operation by analyzing the prediction residuals according to the following steps:

(1) Based on the data of the SCADA system, optimize the initial weights and thresholds of the BPNN by using a GABP-based training method for the prediction model;

(2) Establish data samples of state parameters in various distribution intervals and then train the prediction model by a 10-fold cross-validation;

(3) Analyze the precision of the prediction model by a suitable analysis index;

(4) Estimate the state parameters in the period concerned by the established state parameter prediction model to gain the prediction residual sequence;

(5) Study the statistics of the residual sequence's distribution characteristics; gain the TLS fitting parameters by fitting the residual errors with TLS distribution, and then divide the ranges of residual anomaly intensity;

(6) Quantitatively analyze the intensity of residual anomaly in the period concerned and calculate a residual error abnormal index to identify any anomaly of the wind turbine's state parameters.

The general anomaly identification model of WT state parameters is shown in Figure 1.

[Figure1 could be here]

3. PREDICTION MODEL OF WIND TURBINE STATE PARAMETERS

3.1GABP-based prediction model

Known for its strength to map complex nonlinear and unknown relationships, the BPNN is usually used by researchers to model unstructured problems (Goh, 1995; Haykin, 2009; Yao, 1999). The multilayer feed-forward architecture of the BPNN is shown in Figure 2. This neural network is trained by using a back-propagation algorithm, which utilizes gradient descent to iteratively update the weights and biases of the neural network, minimizing the performance function commonly measured in terms of an error goal between the actual and predicted output. Due to easy implementation, the BPNN is well adopted as a universal function approximator. However, its drawbacks of getting trapped in slow convergence and local minima also needs to be solved. These drawbacks are mainly attributed to random initialization of weights and biases before training a neural network.

[Figure2 could be here]

In order to screen out the best network model for the BPNN, an effective methodology for improving its prediction performance and convergence to global optima is developed. There are many optimization methods that can be used to optimize BPNN parameters, such as GA, backtracking search algorithm (BSA), rain-fall optimization algorithm (RFO), artificial cooperative search and multi-objective PSO algorithm (Mostafa et al., 2016; Kaboli et al., 2017; Kaboli et al., 2016; Rafieerad et al., 2016), which provide the possibility of optimizing BPNN parameters. The GA is a gradient-free global optimization and search method which imitates natural biological evolution (Srinivas and Patnaik, 2002; Whitley, 1994). Compared to traditional search and optimization procedures such as enumerative and calculus-based strategies, the GA is a promising optimization technique inspired by evolutionary processes, namely, natural selection and genetic variation (Grefenstette, 1986; Konak et al., 2006; Shrouf et al., 2014). The GA allows the simultaneous search for optimal solutions in different directions to minimize the chance of getting trapped in a local minimum and speed up its convergence. In this study, the GA is utilized to optimize the performance of BPNN. The flowchart of this hybrid GABP method is presented in Figure 3.

[Figure3 could be here]

The proposed hybrid GABP method has the followingsteps:

(1) Population initialization

Real-number encoding is applied to the individuals, each of which is a real number string composed of an input layer, hidden layer connection weight, hidden layer threshold, hidden layer, output layer weight and output layer threshold. The individuals, which incorporate all weights and thresholds of a neural network, can form a neural network with determined structure, weights, and threshold if the network structure is known.

(2) Fitness function

The BPNN is trained with training data; then the anticipated output is worked out.

Utilizing the GA to optimize the BPNN, the optimality is measured by the fitness functions that are defined in relation to the considered optimization problem. In the process of training and testing, the BPNN aims to improve the generalized performance of the regression model; in other words, to minimize the deviation of the testing samples between the expected values and forecasting values. Therefore, the fitness function can be defined as follows:

$$F = k(\sum_{i=1}^{n} abs(y_i - o_i)) \tag{1}$$

where n is the total number of neural network output nodes, $y_i$is the expected output at node i, o_i is the forecasting output at node i, and k is the coefficient.

(3) Selection

For the selection operation of GA, several methods are available, (e.g., tournament and roulette) which are more commonly used in applications. The latter (roulette) is a selection tactic based on the fitness proportion, with p_i,the probability of selecting each i,as follows:

$$f_i = k / F_i \tag{2}$$

$$p_i = f_i / \sum_{i=1}^{N} f_i \tag{3}$$

where, k is the coefficient, N isthe population size, and F_i is the fitness of individual i, the smaller, the better. The fitness value is reciprocated before the individual is selected.

(4) Crossover

Given real number encoding for individuals, real number crossover is applied for this procedure.

The operation method of the k-th chromosome a_k and the l-th chromosome a_l at j position can be expressed as

$$\begin{cases} a_{kj} = a_{kj}(1-b) + a_{lj}b \\ a_{lj} = a_{lj}(1-b) + a_{kj}b \end{cases} \tag{4}$$

where b is a random number at [0, 1].

(5) Mutation

The gene a_{ij} of the i-th individual is selected for mutation, and the mutation operation method is shown as

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) \times f(g) & r > 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) \times f(g) & r \le 0.5 \\ f(g) = r_2 \times (1 - g / G_{\max})^2 \end{cases} \tag{5}$$

where $a_{\max}$and $a_{min}$are the upper and lower limits of gene a_{ij}respectively, r_2 is a random number, g is the current iterations, $G_{max}$is the max iteration number, and r is a direct random number at [0, 1].

3.2Selection of sample data and the input parameters

The status of WTs can be reflected by many parameters, among which the component temperature is of great importance. In addition, the condition parameters of WTs can be affected by wind speed, ambient temperature and the other components. Therefore, in traditional prediction techniques, the training model which uses recent samples does not consider the overall operation conditions of the WTs, and a significant prediction error may be induced once the other operating points of the parametric sequence appear. Since wind energy is random, the real-time sample model to a certain extent is for a single working condition. The forecast error can be increased when the volatility of unit operating conditions is high. Given the above challenges, in this study, we attempt to determine the section of monitored condition parameters of WTs to obtain the unabridged training samples for the condition parameters. Obviously, the condition parameters are predominantly wind speed.

In this section, the distribution of wind speed is first analyzed. The cut-in wind speed of the WTs is 3m/s, and only the wind speed greater than the cut-in speed is considered. Figure 4 shows the distribution of wind speed values from ten WTs on a wind farm. It is observed that the scope of wind speeds on all WTs are similar, from 5m/s to 10m/s, and only a few values are higher than 15m/s. There are differences in the maximum, 75% and 25% of the WTs, because there is a wake effect in the wind speed.

[Figure4 could be here]

As an external factor of wind turbine’s operation, wind speed is related to the state parameters of the wind turbine. However, it is difficult to categorize the operating conditions of the wind turbine only by the wind speed, because the values of each state parameter may include the entire range of values in different wind speed intervals. Moreover, there are some correlations between the state parameters, and the other state parameters witha higher correlation degree can be selected to divide the state parameter interval.

Taking the temperature of gearbox input shaft as an example, when the wind speeds are 3-4m/s or 4-5m/s, the temperature range of the gearbox in each wind speed range is 20-80°C, so the range cannot be divided. Therefore, according to the wind speed and the correlation levels of other parameters, the temperatures can be divided by the sample intervals. As shown in Figure 5, the wind speeds of 3-25m/s are divided into a Wx range, while the relevant monitoring parameters are divided into a Wy range. Each section is a sub-sample of the state parameters, and all the sub-samples together constitute the sample data of the state parameter.

[Figure5 could be here]

Taking the state parameter data of #10 wind turbine in the SCADA system of a wind farm as an example, the data samples are divided into groups by taking into account the distinction between the samples and the number of samples. Table 1 shows the temperature intervals of gearbox input shaft in #10 wind turbine, which is composed of a total of 10 sub-samples. Among them, the number of samples larger than the rated wind speed range is limited, and these samples are individually divided into an interval.

In the state parameter prediction model, the selection of input parameters is a premise in simplifying models and ensuring prediction accuracy. In this paper, the input parameters are selected according to the correlation between the state parameters. Meanwhile, to avoid information redundancy, the condition parameters of the same type for the same components need to be removed. For instance, when the temperature of wind turbine winding (U_1) is being predicted, although the temperature of wind turbine winding (U_2) is closely related to U_1 and a higher prediction precision can be achieved using U_2 as the input parameter, the redundancy information would prevent the identification of abnormal information, hence we can select the temperature of wind turbine bearing as an input parameter. Besides, the wind speed is selected as one of the input parameters for all condition parameters.

[Table 1 could be here]

3.3 Precision analysis for prediction model

In this paper, the goal of wind turbine state parameter prediction is evaluated by the prediction accuracy. In order to comprehensively assess the effectiveness of the forecasting method, the following three indicators are used to measure the prediction accuracy.

The mean square error (MSE) can be presented by

$$MSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t^{'} - y_t)^2} \tag{6}$$

The mean absolute error (MAE) can be calculated by

$$MSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t^{'} - y_t)^2} \tag{7}$$

The mean absolute percent error (MAPE) can be calculated by

$$MSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t^{'} - y_t)^2} \tag{8}$$

wherein (6)-(8),y_t' is the predicted value at time t, $y_t$is the measured value at time t, and n is the length of the sample sequence.

(1) Effect of training methods on prediction precision

In this part, we take the SCADA data of a wind farm in China as an example to predict the temperature of wind turbine bearing (front) using BPNN and GABP respectively. For comparison, the input samples of the two prediction models are the normal sample data of the wind turbine in the most recent three months. The input parameters are the ambient temperature, wind speeds, and the temperatures of generator bearing at the last moment, and the output is the temperature of generator front bearing at the next time unit. Three prediction time intervals, namely 1 minute, 10 minutes and 15 minutes are set for the analysis and prediction of the temperature.

[Figure6 could be here]

[Figure7 could be here]

[Figure8 could be here]

Figures 6-8 present the temperature prediction results of generator front bearing under different prediction time intervals. Table 2 shows the results of prediction precision based on GABP and BP. It can be seen that the prediction errors of the two prediction methods become more significant as the prediction time scale increases. However, under the same prediction time scale, the prediction error of GABP is less than that of the BP neural network. Therefore, the GABP improves the prediction accuracy of the wind turbine state parameters and is used in this study.

[Table 2 could be here]

(2) Effect of training samples on prediction precision

Based on the actual data of wind farm SCADA, the prediction accuracy is compared, based on the two kinds of training samples (the recent samples and our training samples) using the GABP method. For comparison, the input parameters are the ambient temperature, wind speeds, and the previous temperatures of generator bearing (front), and the output parameter is the temperature of generator bearing at the next moment. The sample analysis period is from May 10 to May 26 in #10 wind turbine. The current samples are selected based on the normal data of #10 wind turbine from February to April.

Figures 9-10 show the prediction error sequences and residual distribution results of the two prediction methods in the analysis period, respectively. From Figure 9, the prediction error range of the training model with the current condition samples is large, and there are obvious errors between the anticipated value and real value at the moments 417, 1240 and 1974, especially. According to the survey, there is no fault record for #10 wind turbine in this period, and this may be caused by the insufficiency of training samples. As seen in Figure 10, the models' trained based on the presented data sampling method (seen in Section 3.2) can give a smaller fluctuation range of prediction error. Moreover, over 99% of errors appear at the temperature interval of [-5oC , 5oC]. Table 3 shows the results of prediction precision based on two kinds of samples. It can be observed that our training samples, to some extent, can improve the prediction precision of condition parameters.

[Figure9 could be here]

[Figure10 could be here]

[Table 3 could be here]

(3) Effect of input parameters on prediction precision

For comparison, the wind turbine gearbox main shaft bearing temperature is taken as an example. Four input parameter~~s~~ models of the main bearing gearbox side temperature are established under the same data sample and training method. The models of the four input parameters are: (a) wind speeds and the previous monitored values; (b) ambient temperatures, and the previous monitored values; (c) wind speeds, ambient temperatures and the previous monitored values; and (d) wind speeds, the temperature of gearbox input bearing, and the previous monitored values.

Figure 11 presents the prediction error residuals based on the four models. From 11(a) and (b), it can be observed that the prediction error is large only when the wind speed or the ambient temperature is the input parameter. This is because the main bearing temperature is affected by the wind speed and the ambient temperature at the same time. Figure 11(c) shows that the prediction error is better improved when the input parameters are both wind speed and ambient temperature. From Fig. 11 (d), it can be seen that the accuracy of the prediction is further improved when the ambient temperature is replaced by the temperature of gearbox input bearing. The comparison in Table 4 shows that the input parameters have a great influence on the accuracy of the prediction model. In particular, selecting the state parameters with wind speed and the parameters with large correlation as input parameters can improve the prediction accuracy.

[Figure11 could be here]

[Table 4 could be here]

4. ANOMALY IDENTIFICATION MODEL FOR TURBINE STATE PARAMETERS

4.1Distribution characteristics of state parameter prediction errors

In the literature, a normal distribution is always used to describe the prediction model of wind turbine state parameters (Schlechtingen et al., 2014). However, many tests have shown that the wind turbine state parameters show a "pointed apex and thick tail" distribution characteristic. Thus the normal distribution may not be suitable to describe the actual data.

The t location-scale (TLS) distribution is usually used to model data distributions with heavy tails (more prone to outliers). The TLS approaches the normal distribution as its shape parameter approaches infinity, and small shape parameters yield heavy tails. In this study, the TLS is used to depict the state parameter errors, and the fitting effect of the TLS is also verified.

The probability density function of the TLS distribution can be presented as

$$f(x,\mu,\sigma,v)=\frac{\Gamma(\frac{v+1}{2})}{\sigma\sqrt{v\pi}\Gamma(\frac{v}{2})}[\frac{v+(\frac{x-\mu}{\sigma})}{v}]^{-(\frac{v+1}{2})} \tag{9}$$

where Γ is the gamma function, μ is the location parameter, σ is the scale parameter, and v is the shape parameter.The fitting effect of the TLS distribution is compared with the normal and logistic distributions. For more intuitive comparison, the fitting index is utilized to describe the difference between the probability density curve and the square probability density distribution; that is, the probability density distribution curve of the actual data. The fitting index *I* is defined as

$$I = \sum_{i=1}^{M} (y_i - \overline{N_i})^2 ,\ y_i = f(\overline{C_i}) \quad i = 1, 2, \ldots, M \tag{10}$$

where *M* is the number of groups of the frequency distribution histogram, N_i and C_i are the height and the central positions of the *i*-th square column respectively, *f* is the fitted probability density function, and y_i is the value corresponding to the probability density function on the C_i. The smaller the fitting index *I*, the more accurate the fit.

The GABP method is used to predict the temperature of generator front bearing and the wind speed, whenthe temperature of generator front bearing at the previous moment and main bearing gearbox-side temperature (data of the farm's #17 turbine in May 2012) are the input parameters. The prediction errors at time intervals of 1 minute, 10 minutes and 15 minutes are obtained.

Table 5 lists the fitting precisions of the prediction errors under three time-intervals based on the normal, logistic, and the TLS distributions, respectively. Under the three prediction time-intervals, the TLS distribution is more accurate than the other two distributions. Figures 12(a)~(c) show the fitting results of the prediction error of generator front bearing temperatures. It illustrates that the normal distribution has a large amplitude error at the tail of the fitting error distribution. In addition, the logistic distribution model has a large amplitude error in the central region of the fitting error distribution. Therefore, the overall probability density curve of TLS distribution is more accurate to describe the prediction error in the interval within the probability.

[Table 5 could be here]

[Figure12 could be here]

4.2Distribution characteristics of state parameter prediction errors

In this section, the TLS distribution is used to analyze the difference of the error distribution under different forecasting time-intervals and different operating conditions. The prediction errors of the generator bearing temperature (data of #17 turbine in May 2012) are analyzed. The change in wind speeds mainly leads to the variation of the operating condition of wind turbine. In order to ensure the number of sample data in each wind speed range, the values of wind speeds are divided into four wind speed ranges which are 3-6m/s, 7-9m/s, 9-11m/s, and 12-25m/s). In Table 6, the residual fitting parameters of TLS distributions in different wind speed ranges under three prediction time scales are compared. It can be seen that the standard deviation of TLS distribution increases with respect to the increase of the wind speed at the same prediction time, indicating that the error dispersion increases with the increase of the wind speed. In the same wind speed range, the standard deviation of different prediction time interval is different. The larger the prediction time interval, the larger the standard deviation, indicating that the error disperses more widespread as the prediction time interval increases.

Figures 13 (a)-(c) show the prediction error probability density curves for the generator bearing temperatures under the predicted winds at three time intervals. It can be seen that the error probability density curves vary greatly at different wind speeds under the same interval and the error probabilities deviate to different degrees from the zero. In addition, the high wind speed leads toa flat probability curve and wide error distribution. Therefore, it is concluded that the prediction error of the wind turbine state parameters is highly related to the wind speed. Moreover, the probability density functions are different with different wind speed intervals for the prediction error of the same state parameter.

[Table 6 could be here]

[Figure13 could be here]

4.3Quantification of the abnormal level of prediction residuals

When the wind turbine is in normal operation, the prediction residual of the state parameter is the same as the distribution characteristic of the training error. Therefore, abnormal state parameters can be identified based on the characteristics of their residual distribution. At present, the standard normal distribution is mainly chosen to identify the anomaly of the parameters in the large data samples, but the accuracy of the state parameter fitting is low when applied in wind turbine state parameters.

Figures 14(a)-(d) show the probability density curves of the temperature residual of generator front bearing in different operating conditions. It can be seen that the TLS distribution has a good fitting effect with different wind speed ranges in this study.

[Figure14 could be here]

The traditional method of identifying the residual error is to determine the threshold by setting the confidence interval of the prediction residual. However, the data of the farm’s

SCADA system can be affected by multiple non-turbine-fault factors(e.g., sensor faults and interfered data transmission resultedinexcessive residuals). Moreover, the distribution characteristics of prediction residuals are based on the statistical rule of massive data, and the judgment of a single data's threshold is not able to present the general anomaly information of the turbine operation state. Therefore, an abnormal index is presented in this paper based on the residual distribution of turbine state parameters to identify the anomaly of these parameters.

For normal state parameters, the prediction residual errors stay within the wide range of the probability density. However, narrow ranges might indicate anomaly of state parameters. Therefore, we set quantiles at 0.025, 0.25, 0.75 and 0.975, based on the probability distribution of the parameter residuals. Combined with the parameters of the TLS distribution, the residual value ranges can be divided as shown in Figure 15.

[Figure15 could be here]

The error abnormal index (EAI) refers to the residual error's intensity of anomaly and can be calculated by

$$EAI = 1 - \frac{N_1 C_1}{\sum_{i=1}^{3} N_i C_i} \tag{11}$$

where N_i refers to the number of residual sequences in interval i, C_i is the penalty factor in interval i (set to [1, 3, 5]), N_1 is the number of residual sequences in the first interval, and C_1 is set to 1. A greater EAI means a higher intensity of residual anomaly. According to the data tests and statistics in our studies, the wind turbine state parameters can be directly judged abnormal when EAI exceeds 0.8.

5. CASE ANALYSIS

This case study is based on the SCADA data of a real wind farm. The farm's repair record showed that the #17 turbine was shut down on May 30, 2012, due to the overheating of the generator rear bearing. However, the SCADA system did not have any status parameter limit alarm records before the fault occurred. In order to analyze the pre-fault changes of generator rear bearing temperature, the analysis period is two months before the fault occurred, from April 1 to May 30, 2012. The analysis of the identification process is as follows:

(1) Temperature prediction of generator rear bearing

First, the temperatures of generator rear bearing during the above-mentioned 60 days are predicted every 10 minutes by the GABP method. In the training phase, the first correlation quantity in the samples is the wind speed, and the second correlation quantity is the generator winding temperature, both of which are used to establish the training data samples of the temperatures. In the prediction phase, the data is tested by cross-validation based on different wind speed ranges. The input parameters of the prediction model are wind speed, generator winding temperature, and generator rear bearing temperature at the last time unit, and the output parameter is the generator rear bearing temperature at the next time unit.

Figures 16(a) and (b) show the prediction results of generator rear bearing temperatures by GABP on April 1, 2012, with the following prediction precisions of MSE=0.91°C, MAE=0.49°C and MAPE=1.23%. It can be seen that the GABP has an excellent prediction performance for the generator rear bearing temperatures.

According to the above prediction process, a rolling prediction method is employed to get the residual sequences from April 1 to May 30. Figure 17(a) shows the prediction results of generator rear bearing temperatures in this study period. From Figure 17(a), the maximum temperature of generator rear bearing is 95.20oC in. However, the upper limit warning temperature of generator rear bearing is usually set at 110oC. Hence the monitoring of SCADA fails to alarm the fault because the maximum temperatures in the whole study period are less than 110oC. From the residual sequences shown in Figure 17(b), it can be seen that the residual error increases abruptly at the moment 6500, to a max of 7.4°C and -4.23°C. The residual error changes in a wide range which necessitates a further quantitative analysis of the residual sequence in thefollowing:

[Figure16 could be here]

[Figure17 could be here]

(2) Quantitative analysis of TLS fitting residuals

Table 7 lists the TLS distribution parameters of sample data under the four wind speed ranges. Figures 18 (a)-(d) show the TLS distribution characteristics. It can be seen that the fitting curves are symmetrical about the maximum value, but the mean values have different degrees of deviation from zero. In addition, Table 8 shows the results of fitting accuracy. From the table, we can conclude that the TLS has an excellent fitting performance in the application of fitting residuals for generator rear bearing temperatures.

[Table 7 could be here]

[Figure18 could be here]

[Table8 could be here]

(3) EAI analysis

According to Section 4.4 (Quantification of the residual anomaly), the division results of the sample residuals are shown in Table 9. In this case, the sliding window period is set to be one day for such an extended analysis period. The EAIs are calculated according to (11), and the result is shown in Figure 19. The figure shows that the EAIs increase continuously from Day 40 to Day 46. The EAI reaches 0.8 on Day 46, and the abnormal index of more than 0.8 lasts four days. Afterwards, the EAIs drop down, and then they quickly rise again before the failure with the abnormal index of more than 0.8, lasting four days,occurs again. These observations show that the state parameter is abnormal.

In this case, although the bearing temperature is abnormal due to the fault, the fluctuation range of the state parameter is still within the SCADA alarm threshold. Thus, the abnormal state cannot be detected in time based on the monitoring of the SCADA alarm system. In this paper, a state parameter prediction model is established by normal sample data based on GABP. The EAI is used to quantify the anomaly and reflects the anomaly of the state parameters. It is found that there is a large error between the predicted value and the actual value of the generator bearing temperatures at the early stage of the fault, and the residual distribution is different from normal. Therefore, the case analysis verifies the validity of the proposed identification model. Moreover, this method shows advantages of detecting the abnormal information of the state parameters in advance and then avoiding the occurrence of severe faults of the WTs.

[Table 9 could be here]

[Figure19 could be here]

6. Conclusions

An anomaly identification model for wind turbine state parameters was presented in this paper. The conclusion can be summarized as follows:

Firstly, the wind turbine state parameter prediction model has been developed and trained by the GABP algorithm. The influence of the training algorithm, data samples and input parameters on the prediction performance of the developed models has been analyzed. 1) The GABP optimized model can provide much higher accuracy than the BPNN based model: the MSE(°C), MAE(°C), and MAPE(%) of GABP and BP with 1 minute interval time are (0.0530, 0.0367, 0.0964) and (0.0705, 0.0556, 0.1476), respectively; the MSE(°C), MAE(°C), and MAPE(%) of GABP and BP with 10 minute interval time are (0.1998, 0.1625, 0.4380) and (0.2884, 0.2255, 0.6247), respectively; the MSE(°C), MAE(°C), and MAPE(%) of GABP and BP with 15 minute interval time are (0.2701, 0.2096, 0.5435) and (0.3095, 0.2385, 0.5977), respectively. 2) The accuracy of prediction models developed by using the proposed data sampling method is higher than that trained by the current data: the MSE(°C), MAE(°C), and MAPE(%) of recent samples and samples of this paper with 1 minute interval time are (1.3822, 1.0563, 2.0231) and (1.0977, 0.6448, 1.4491), respectively; the MSE(°C), MAE(°C), and MAPE(%) of recent samples and samples of this paper with 10 minute interval time are (1.5479, 0.9888, 2.1844) and (1.2514, 0.9661, 1.8789), respectively; the MSE(°C), MAE(°C), and MAPE(%) of recent samples and samples of this paper with 15 minute interval time are (1.7643, 1.2633, 2.3431) and (1.4327, 1.1608, 2.0742), respectively. 3) Selecting the state parameters with wind speed and the parameters with large correlation as input parameters can further improve the prediction accuracy: the MSE(°C), MAE(°C), and MAPE(%) of Input parameters including a) Wind speeds and the previous monitored values, b) Ambient temperatures and the previous monitored values, c) Wind speeds, ambient temperatures and the previous monitored values, d) Wind speeds, temperature of gearbox input bearing and the previous monitored values are (1.4473, 1.0901, 2.1886), (1.2327, 0.9505, 2.0491), (1.1356, 0.8087, 1.5926), and (1.1180, 0.7692, 1.3826), respectively.

Secondly, the TLS distribution is employed to characterize the distribution of condition parameter prediction error under different wind speed intervals which is better than Normal distribution and Logistic distribution. And the Fitting precisions of the prediction errors of the three distributions (Normal distribution, Logistic distribution, and TLS distribution) in 1-minute interval time, 10-minutes interval time, and 15-minutes interval time are (0.1054, 0.0815, 0.0493), (0.0779, 0.0943, 0.0772), and (0.1760, 0.1301, 0.1245), respectively.

In addition, A case study for an onshore wind farm has been carried out and analyzed. In this case, although the bearing temperature is abnormal due to the fault, the fluctuation range of the state parameter is still within the SCADA alarm threshold. Thus, the abnormal state cannot be detected in time based on the monitoring of SCADA alarm system by regular method. However, by the proposed method, the EAI reaches 0.8 on the Day 46, and the abnormal index of more than 0.8 lasts four days, which shows that the state parameter is abnormal. The results show that the proposed method is effective in anomaly identification of WTs and can provide an early warning before the wind turbine faults occur.

Although the proposed model can provide a good performance for anomaly identification of wind turbine, some improvements of the model are needed. For example, GABP can provide good performance in the forecasting process in the paper, however, there are many effective optimization methods that can be used to optimize BPNN parameters, such as RFO, artificial cooperative search algorithm and optimized gene expression programming (Kaboli et al., 2017; Kaboli et al., 2016; Kaboli et al., 2017), which may have a similar performance with GABP. Thus, a subsequent work needs to be addressed in the near future, although not demonstrated in this paper.

Acknowledgements

The authors acknowledge the National Basic Research Program of China (973 Program, 2013CB228205), and the National High-tech R&D Program of China (863 Program, 2015AA050204), the Natural Science Foundation of Guangxi (2015GXNSFBA139235), the Foundation of Guangxi Science and Technology Department (AE020069), and the Foundation of Guangxi Education Department (T3020097903) in support of this work. The authors also thanks the anonymous reviewers and the editor for their valuable comments.

References

Chandola, V., Banerjee, A., Kumar, V., 2009, Anomaly detection: A survey. ACM COMPUT SURV 41, 1-58.

Cruz, V.D.L., Martín, M., 2016, Characterization and optimal site matching of wind turbines: Effects on the economics of synthetic methane production. J CLEAN PROD 133, 1302-1311.

Demir, N., Taşkın, A., 2013, Life cycle assessment of wind turbines in Pınarbaşı-Kayseri. J CLEAN PROD 54, 253-263.

Garcia, M.C., Sanz-Bobi, M.A., Pico, J.D., 2006a, SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. COMPUT IND 57, 552-568.

Garcia, M.C., Sanz-Bobi, M.A., Pico, J.D., 2006b, SIMAP: Intelligent System for Predictive Maintenance : Application to the health condition monitoring of a windturbine gearbox. COMPUT IND 57, 552-568.

Goh, A.T.C., 1995, Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering 9, 143-151.

Grefenstette, J.J., 1986, Optimization of Control Parameters for Genetic Algorithms. Systems Man & Cybernetics IEEE Transactions on 16, 122-128.

Haykin, S.S., 2009, Neural networks and learning machines. Pearson Schweiz Ag.

Kaboli, S.H.A., Selvaraj, J., Rahim,N.A., 2016, Long-term electric energy consumption forecasting via artificial cooperative search algorithm. ENERGY 115, 857-871.

Kaboli, S.H.A., Selvaraj, J., Rahim, N.A., 2017, Rain-fall optimization algorithm: a population based algorithm for solving constrained optimization problems. J COMPUT PHYS 19, 31-42.

Aghay, S. H., 2017, Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. ENERGY, 126, 144-164.

Konak, A., Coit, D.W., Smith, A.E., 2006, Multi-objective optimization using genetic algorithms: A tutorial. RELIAB ENG SYST SAFE 91, 992-1007.

Kusiak, A., Li, W., 2010, Virtual Models for Prediction of Wind Turbine Parameters. IEEE T ENERGY CONVER 25, 245-252.

Kusiak, A., Verma, A., 2012a, Analyzing bearing faults in wind turbines: A data-mining approach. RENEW ENERG 48, 110-116.

Kusiak, A., Verma, A., 2012b, Analyzing bearing faults in wind turbines: A data-mining approach. RENEW ENERG 48, 110-116.

Lapira, E., Brisset, D., Ardakani, H.D., Siegel, D., Lee, J., 2012, Wind turbine performance assessment using multi-regime modeling approach. RENEW ENERG 45, 86-95.

Li, J.S., Chen, B., Chen, G.Q., Wei, W.D., Wang, X.B., Ge, J.P., Dong, K.Q., Xia, H.H., Xia, X.H., 2017, Tracking mercury emission flows in the global supply chains: A multi-regional input-output analysis. J CLEAN PROD 140, 1470-1492.

Li, J.S., Chen, G.Q., 2013, Energy and greenhouse gas emissions review for Macao. RENEW SUST ENERG REV 22, 23-32.

Li, J.S., Duan, N., Guo, S., Ling, S., Lin, C., Wang, J.H., Hou, J., Hou, Y., Meng, J., Han, M.Y., 2012, Renewable resource for agricultural ecosystem in China: ecological benefit for biogas by-product for planting. ECOL INFORM 12, 101-110.

Li, J.S., Xia, X.H., Chen, G.Q., Alsaedi, A., Hayat, T., 2016, Optimal embodied energy abatement strategy for Beijing economy: Based on a three-scale input-output analysis. RENEW SUST ENERG REV 53, 1602-1610.

Modiri-Delshad, M., Kaboli, S.H.A., Taslimi-Renani, E., Rahim, N.A., 2016, Backtracking search algorithm for solving economic dispatch problems with valve-point effects and multiple fuel options. ENERGY 116, 637-649.

Ortegon, K., Nies, L.F., Sutherland, J.W., 2013, Preparing for end of service life of wind turbines. J CLEAN PROD 39, 191-199.

Rafieerad, A. R., Bushroa, A. R., Nasiritabrizi, B., Kaboli, S. H., Khanahmadi, S., & Amiri, A., 2016, Toward improved mechanical, tribological, corrosion and in-vitro bioactivity properties of mixed oxide nanotubes on Ti-6Al-7Nb implant using multi-objective PSO. Journal of the Mechanical Behavior of Biomedical Materials, 69: 1.

Schlechtingen, M., Santos, I.F., 2011a, Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mechanical Systems & Signal Processing 25, 1849-1875.

Schlechtingen, M., Santos, I.F., 2011b, Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mechanical Systems & Signal Processing 25, 1849-1875.

Schlechtingen, M., Santos, I.F., Achiche, S., 2013, Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description. Applied Soft Computing Journal 13, 259-270.

Schlechtingen, M., Santos, I.F., Achiche, S., 2014, Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description, Elsevier Science Publishers B. V.

Shrouf, F., Ordieres-Meré, J., García-Sánchez, A., Ortega-Mier, M., 2014, Optimizing the production scheduling of a single machine to minimize total energy consumption costs. J CLEAN PROD 67, 197-207.

Srinivas, M., Patnaik, L.M., 2002, Genetic algorithms: a survey. COMPUTER 27, 17-26.

Stickland, M., 2012, International Standard IEC61400-12-1 : Wind Turbines-Part 12-1: Power performance measurements of electricity producing wind turbines: Annex G.

Sun, P., Li, J., Wang, C., Lei, X., 2016, A generalized model for wind turbine anomaly identification based on SCADA data. APPL ENERG 168, 550-567.

Whitley, D., 1994. A Genetic Algorithm Tutorial., Statistics and Computing, pp. 65-85.

Xiang, J., Watson, S., Liu, Y., 2009, Smart Monitoring of Wind Turbines Using Neural Networks, Springer Berlin Heidelberg.

Yao, X., 1999, IEEE Xplore - Evolving artificial neural networks.

Zaher, A., Mcarthur, S.D.J., Infield, D.G., Patel, Y., 2010, Online wind turbine fault detection through automated SCADA data analysis. WIND ENERGY 12, 574-593.

Zaher, A.S., Mcarthur, S.D.J., 2007. A Multi-Agent Fault Detection System for Wind Turbine Defect Recognition and Diagnosis., Power Tech, 2007 IEEE Lausanne, pp. 22-27.

Figure captions

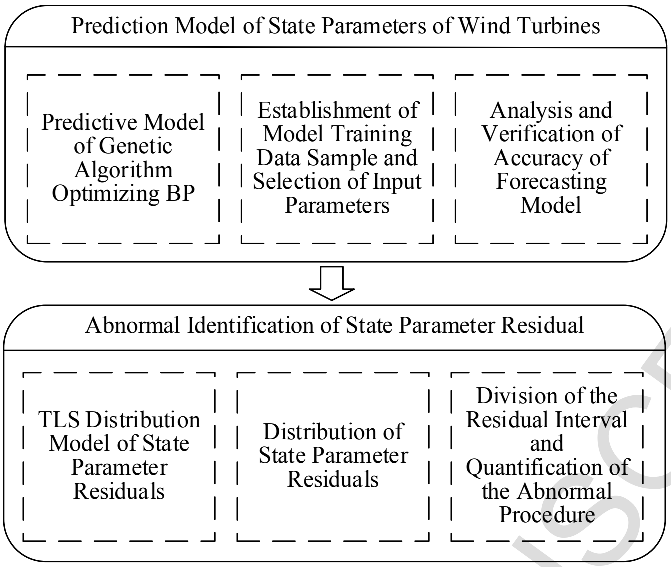

Prediction Model of State Parameters of Wind Turbines
Predictive Model of Genetic Algorithm Optimizing BP
Establishment of Model Training Data Sample and Selection of Input Parameters
Analysis and Verification of Accuracy of Forecasting Model
Abnormal Identification of State Parameter Residual
TLS Distribution Model of State Parameter Residuals
Distribution of State Parameter Residuals
Division of the Residual Interval and Quantification of the Abnormal Procedure

Figure 1 General anomaly identification model of WT state parameters

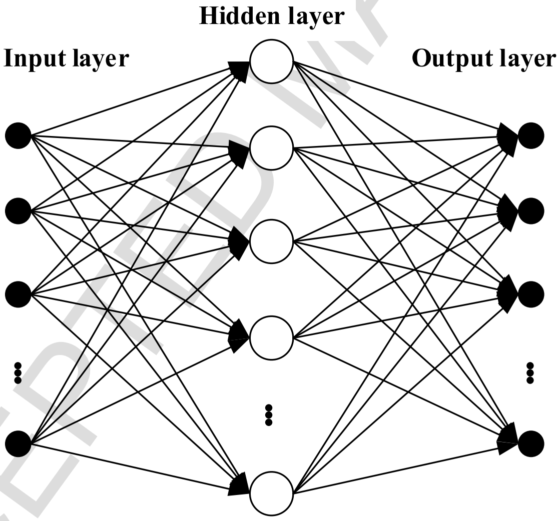


Figure 2 The architecture of BPNN

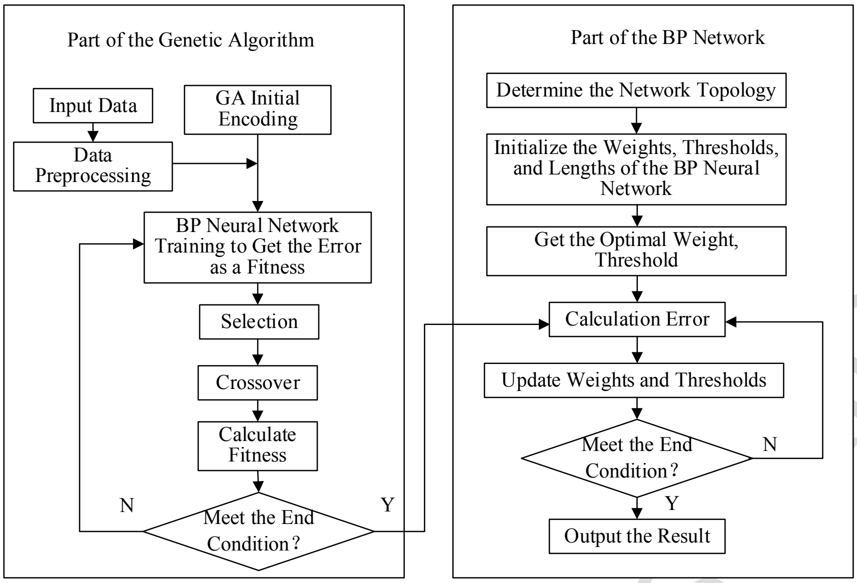

Part of the Genetic Algorithm
Input Data
GA Initial Encoding
Data Preprocessing
BP Neural Network Training to Get the Error as a Fitness
Selection
Crossover
Calculate Fitness
N
Y
Meet the End Condition?
Part of the BP Network
Determine the Network Topology
Initialize the Weights, Thresholds, and Lengths of the BP Neural Network
Get the Optimal Weight, Threshold
Calculation Error
Update Weights and Thresholds
Meet the End Condition?
N
Y
Output the Result

Figure 3 Flowchart of the proposed hybrid BPNN and GA algorithm

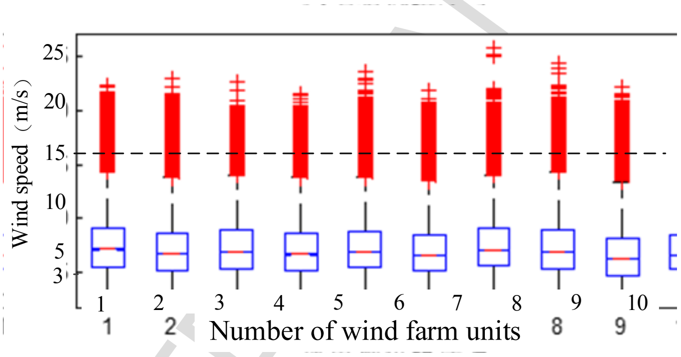


Figure 4 Boxplot of WT wind speeds

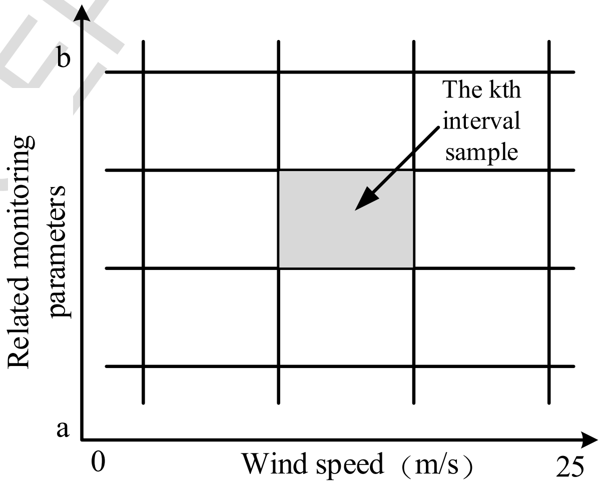


Figure 5 Classifications of WT State Parameter Samples

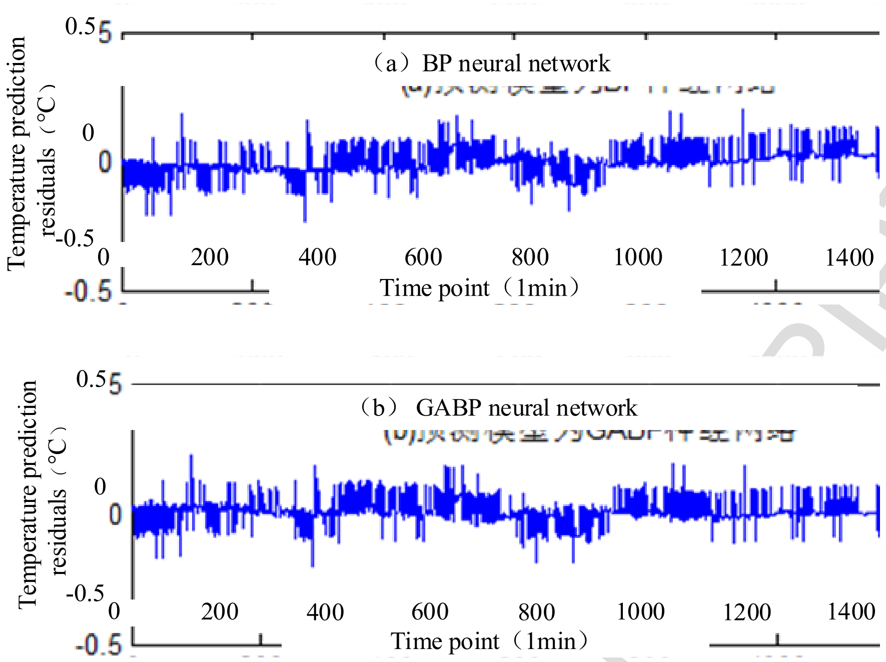


Figure 6 Prediction results of generator front bearing under time interval of 1 minute

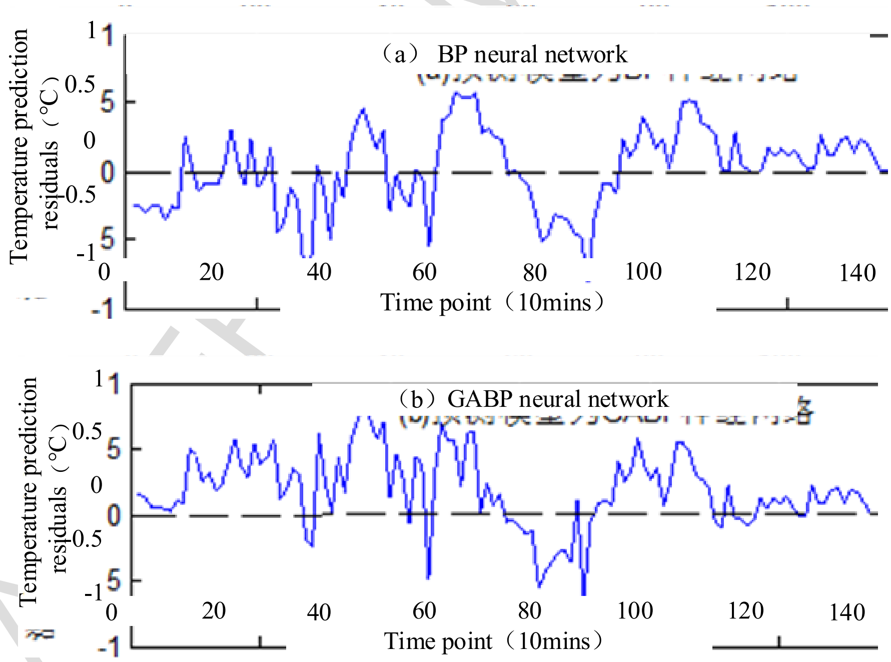


Figure 7 Prediction results of generator front bearing under time interval of 10 minutes

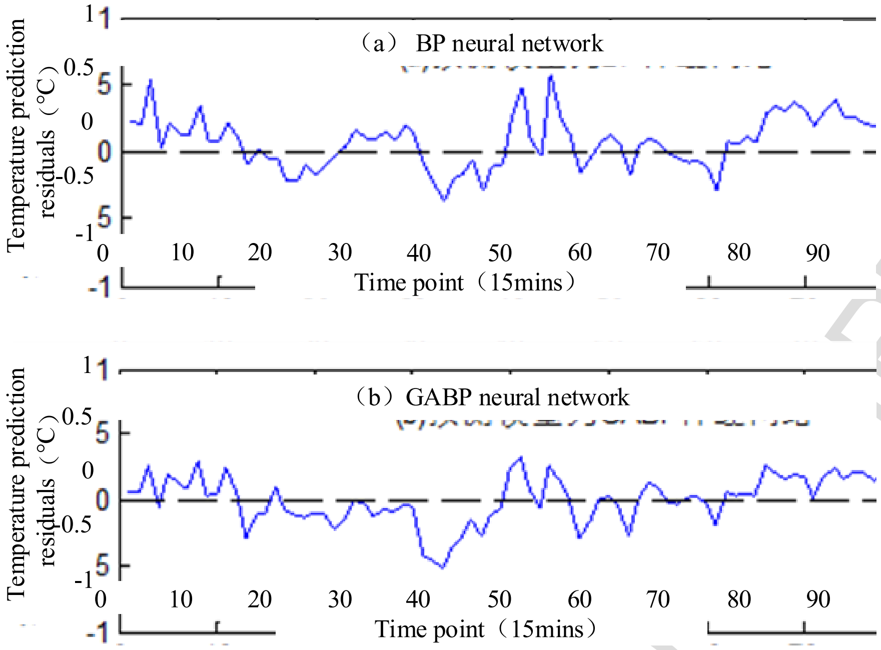


Figure 8 Prediction results of generator front bearing under time interval of 15 minutes

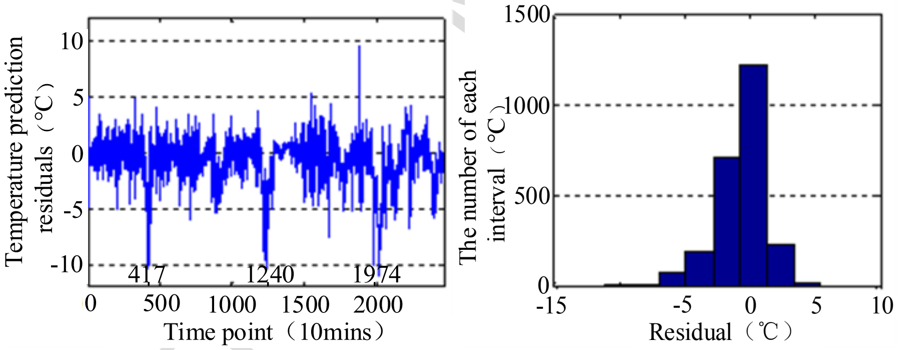


Figure 9 Prediction error sequences and residual distribution results based on the current samples

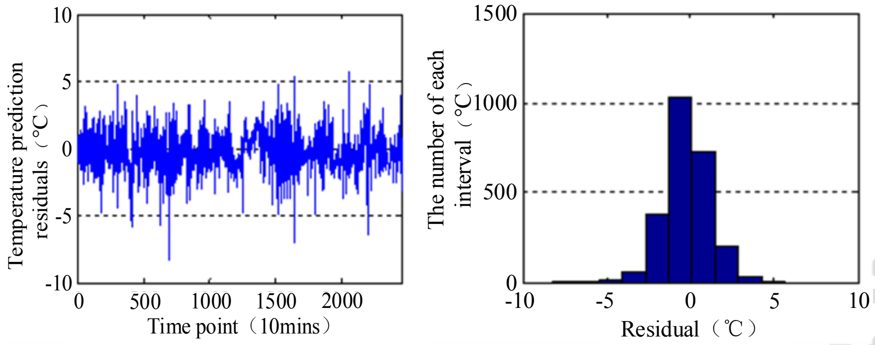
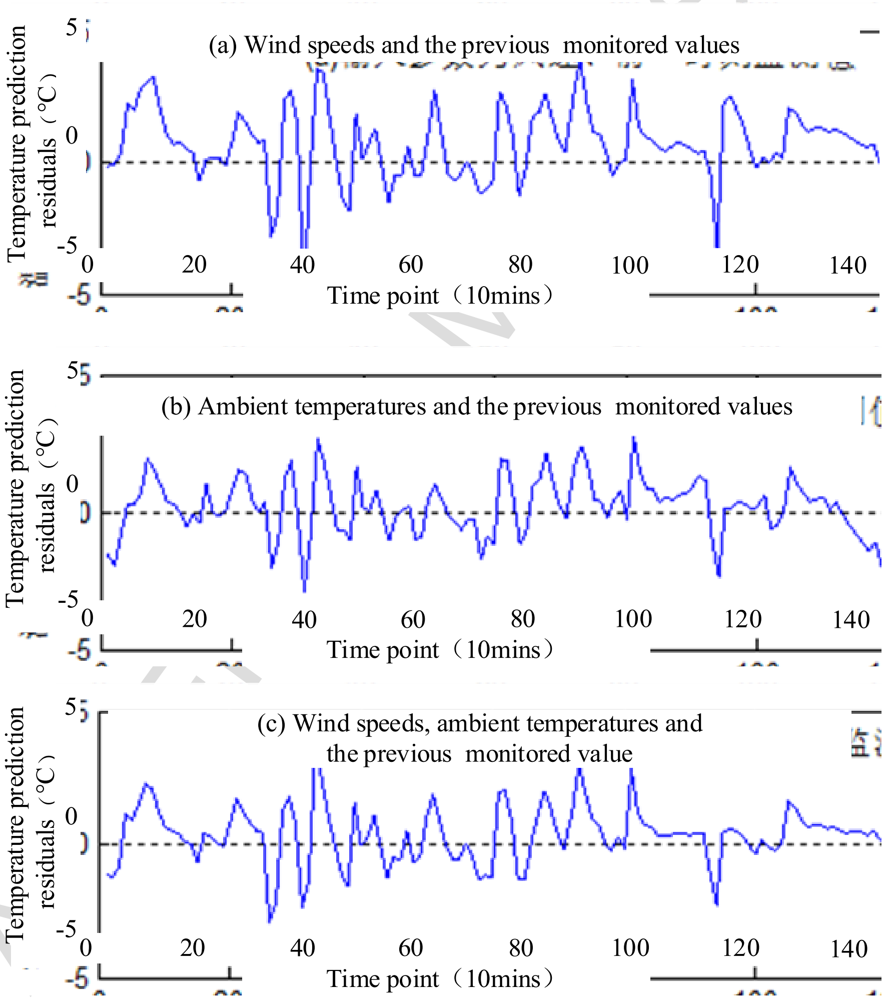


Figure 10 Prediction error sequences and residual distribution results based on our training samples of this paper



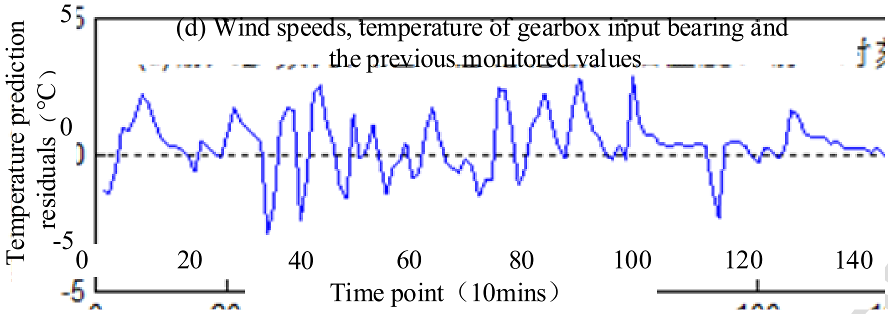
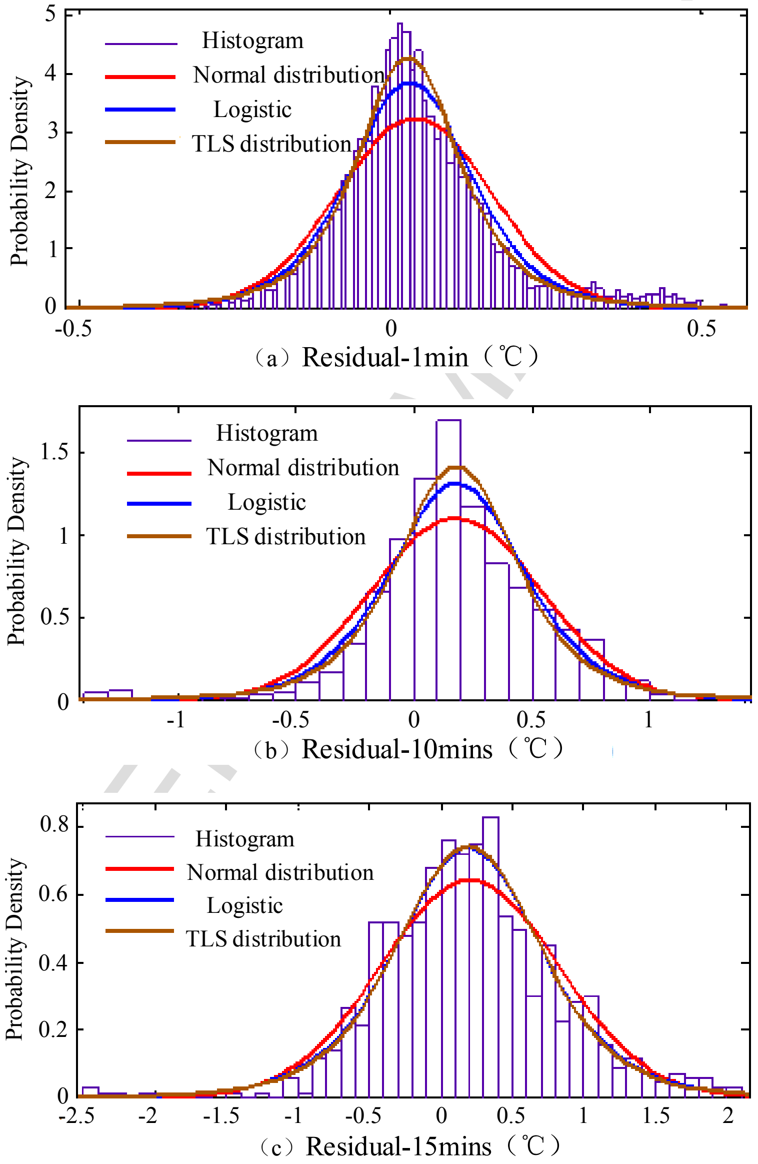


Figure 11 Prediction error residuals based on the four models



（a）Residual-1min（℃）

（b）Residual-10mins（℃）

（c）Residual-15mins（℃）

Figure 12 Fitting results of prediction error of generator front bearing temperatures

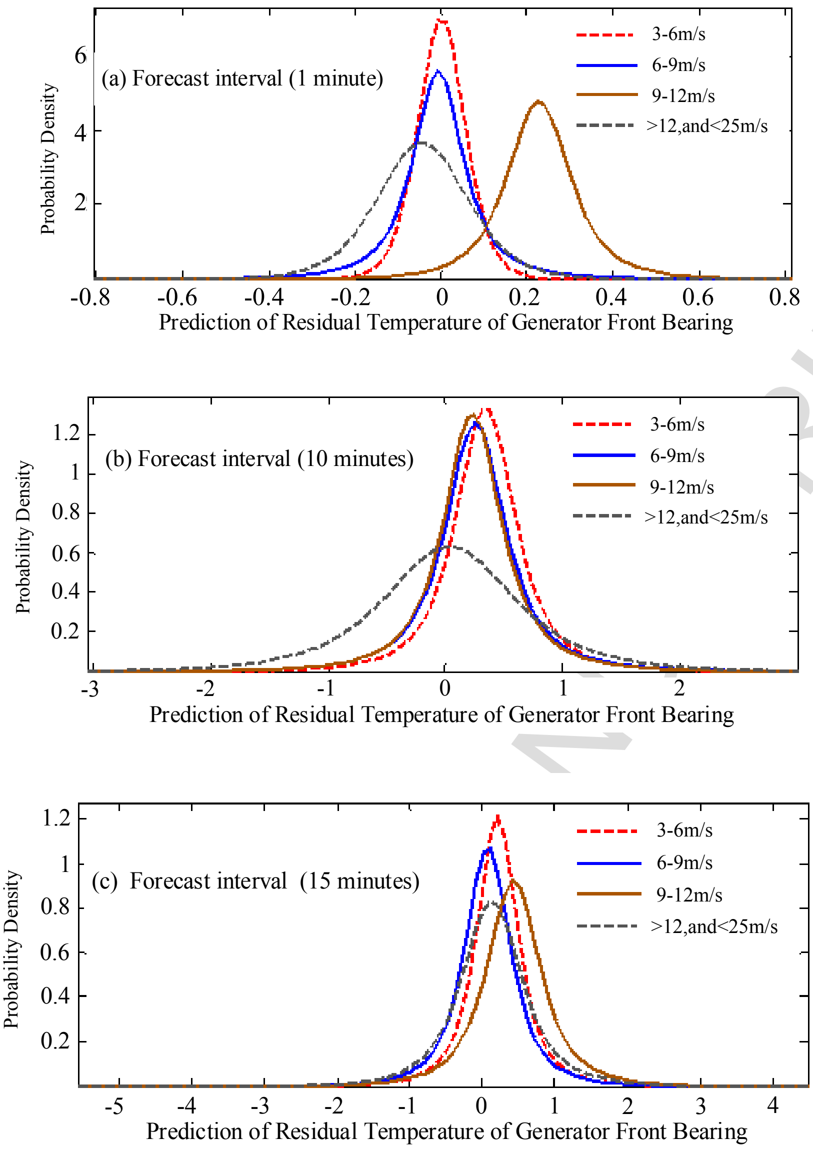
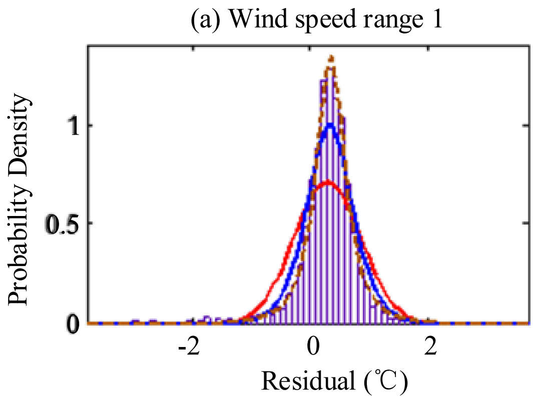


Figure 13 Prediction error probability density curves of the generator bearing temperatures



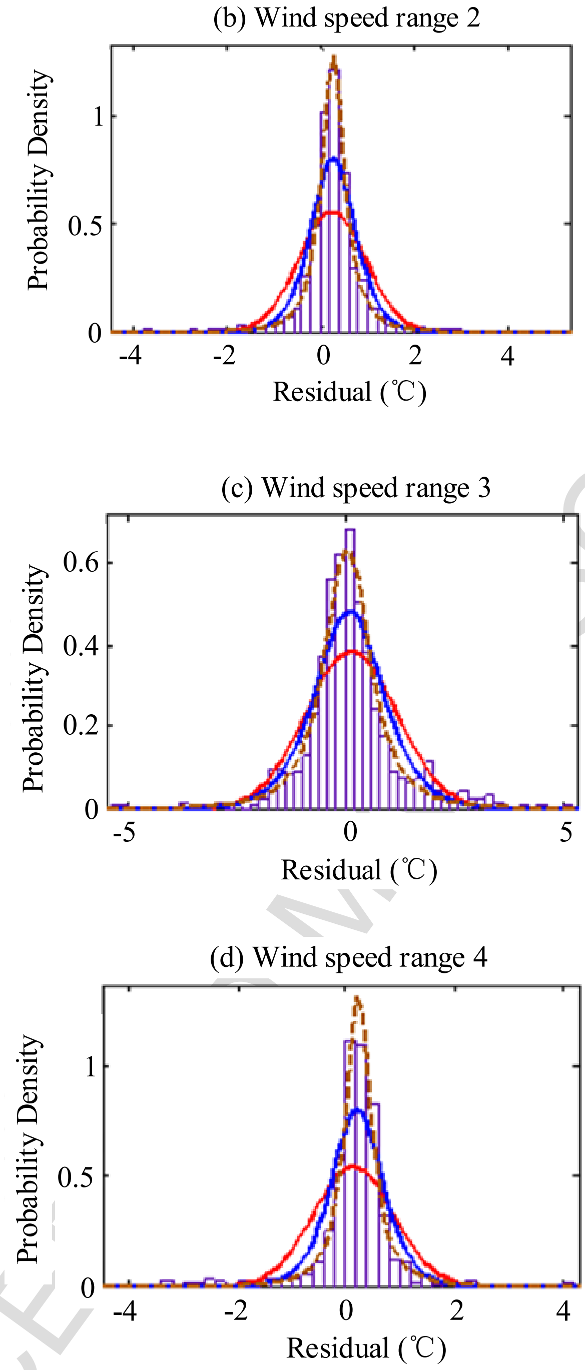


Figure 14 Probability density curves of temperature residual

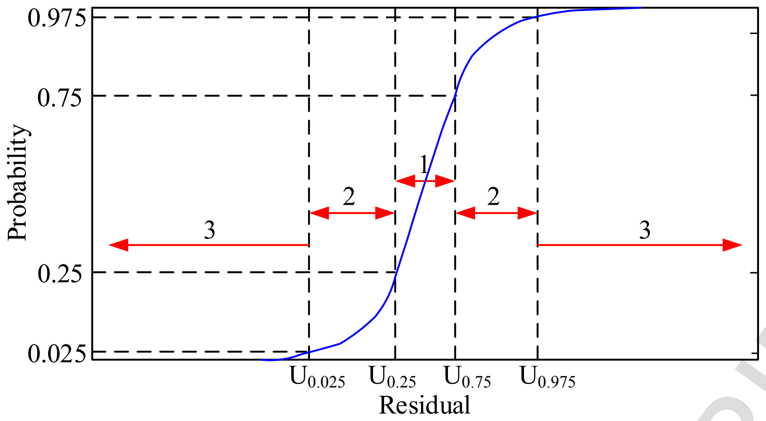


Figure 15 The interval of the prediction residuals

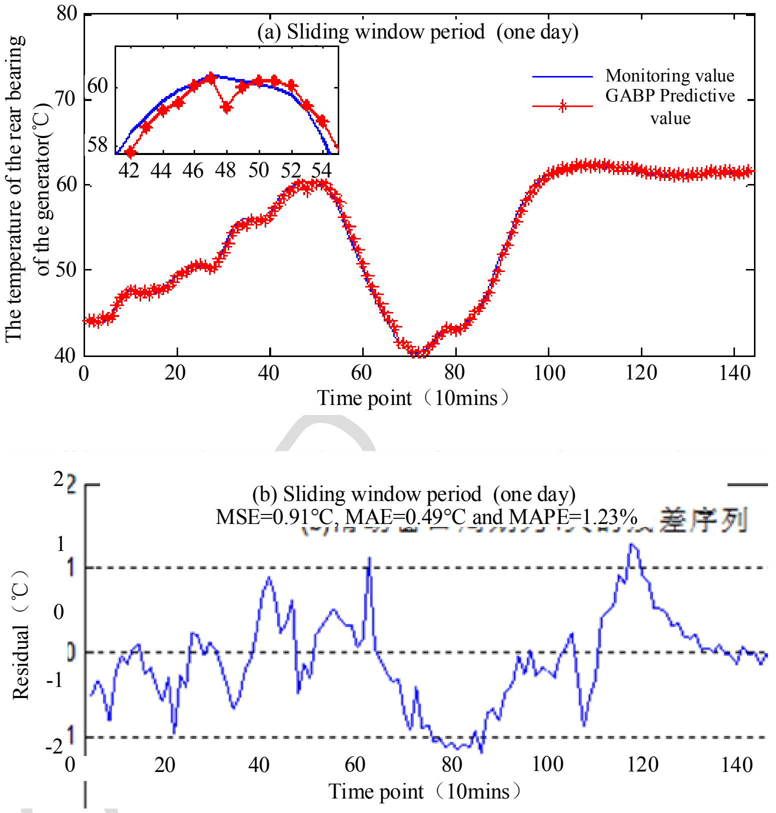


Figure 16 Prediction results of generator rear bearing temperatures

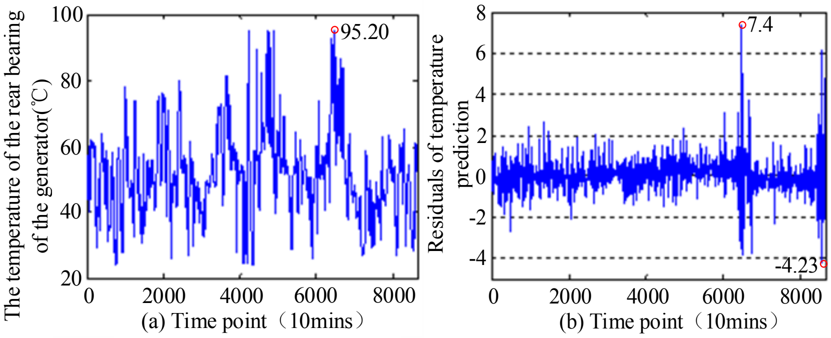
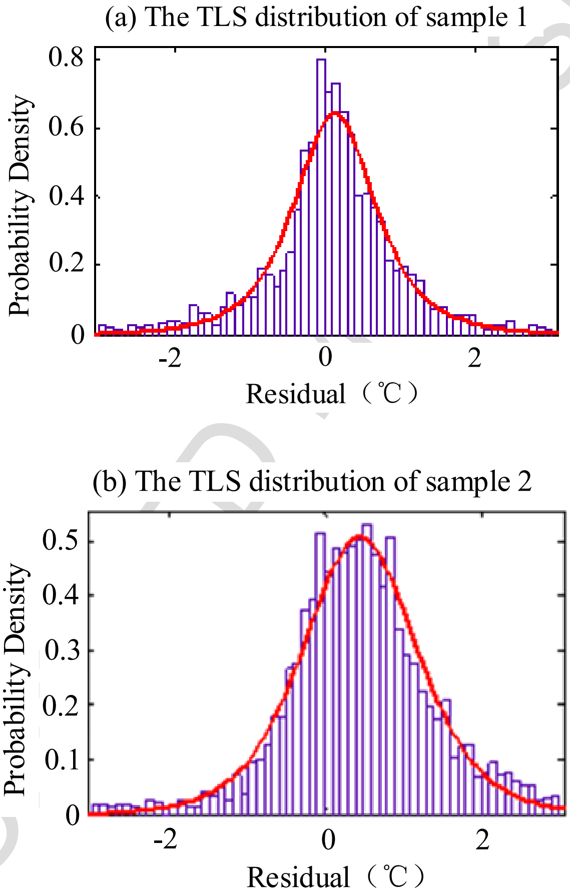
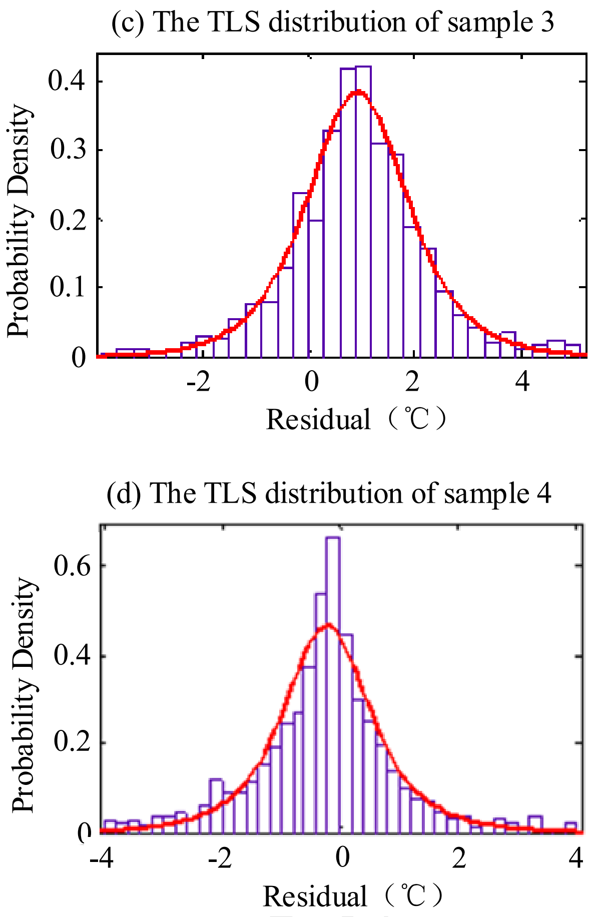


Figure 17 Prediction residual results of generator rear bearing temperatures





(c) The TLS distribution of sample 3

(d) The TLS distribution of sample 4

Figure 18 Sectoral nexus impacts on energy and water systems

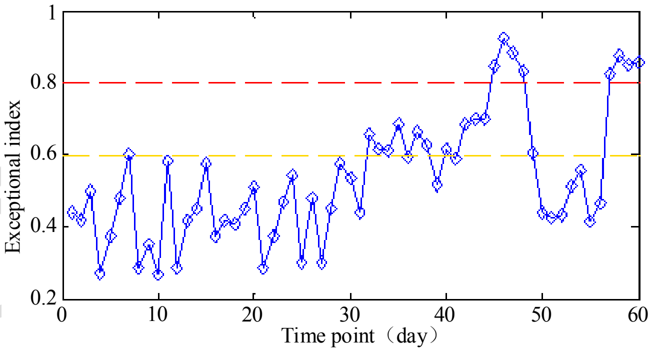


Figure 19 Results of EAI.

Table list

Table 1 Temperature intervals of gearbox input shaft in #10 WT

Range of wind speed range	Range of main bearing temperature	Sample of input shaft temperature	
		number	sample size
3-6m/s	5-20°C	Y_{11}	4237
	20-35°C	Y_{12}	79174
	35-50°C	Y_{13}	89592
6-9m/s	5-20°C	Y_{21}	1048
	20-35°C	Y_{22}	25094
	35-50°C	Y_{23}	4596
9-12m/s	5-20°C	Y_{31}	0
	20-35°C	Y_{32}	10305
	35-50°C	Y_{33}	11804
>12 and <25m/s	5-50°C	Y_{41}	2580

Table 2 Prediction precision results based on GABP and BP

Interval of forecast time	Training methods	Indicators of evaluation		
		MSE(°C)	MAE(°C)	MAPE(%)
1 min	BP	0.0705	0.0556	0.1476
	GABP	0.0530	0.0367	0.0964
10 mins	BP	0.2884	0.2255	0.6247
	GABP	0.1998	0.1625	0.4380
15 mins	BP	0.3095	0.2385	0.5977
	GABP	0.2701	0.2096	0.5435

Table 3 Results of prediction precision based on two kinds of samples

Samples	Sample selection	Indicators of evaluation		
		MSE(°C)	MAE(°C)	MAPE(%)
1 min	recent samples	1.3822	1.0563	2.0231
	samples of this paper	1.0977	0.6448	1.4491
10 mins	recent samples	1.5479	0.9888	2.1844
	samples of this paper	1.2514	0.9661	1.8789
15 mins	recent samples	1.7643	1.2633	2.3431
	samples of this paper	1.4327	1.1608	2.0742

Table 4 Prediction error results based on the four models

Input parameters	Indicators of evaluation		
	MSE(°C)	MAE(°C)	MAPE(%)
Wind speeds and the previous monitored values	1.4473	1.0901	2.1886
Ambient temperatures and the previous monitored values	1.2327	0.9505	2.0491
Wind speeds, ambient temperatures and the previous monitored values	1.1356	0.8087	1.5926
Wind speeds, temperature of gearbox input bearing and the previous monitored values	1.1180	0.7692	1.3826

Table 5 Fitting precisions of the prediction errors

Interval of forecast time	Distribution model		
	Normal distribution	Logistic distribution	TLS distribution
1 min	0.1054	0.0815	0.0493
10 mins	0.0779	0.0943	0.0772
15 mins	0.1760	0.1301	0.1245

Table 6 Residual fitting parameters of TLS distributions

Interval of forecast time	wind speed *m*/*s*	*μ*/°C	*σ*/°C	*v*
1 min	3-6	-0.0061	0.0632	2.0053
	6-9	0.0016	0.0552	8.88081
	9-12	0.2272	0.0764	2.8462
	>12 and<25m/s	-0.0429	0.1051	6.9641
10 mins	3-6	0.3413	0.2668	2.21639
	6-9	0.2541	0.2726	1.68373
	9-12	0.0327	0.2830	2.1786
	>12 and<25m/s	0.2293	0.3046	1.6602
15 mins	3-6	0.2033	0.2971	2.2708
	6-9	0.0715	0.3322	2.2233
	9-12	0.1257	0.4355	2.5481
	>12 and<25m/s	0.4271	0.4703	2.3381

Table 7 TLS distribution parameters of different sample data

Sample number	*μ*/°C	*σ*/°C	*v*
Sample 1	0.1256	0.5722	2.9020
Sample 2	0.4419	0.7543	5.7670
Sample 3	0.9168	0.8847	4.2980
Sample 4	0.2301	0.98441	2.7151

Table 8 TLS fitting accuracy results of different samples

Sample number	Indicator value
1	0.0824
2	0.0752
3	0.1046
4	0.1244

Table 9 Results of the residual quantiles under four kinds of wind speeds

Sample number	Quantile			
	0.025	0.25	0.75	0.975
1	-1.7307	-0.3140	0.5652	1.9819
2	-1.4220	-0.1007	0.9845	2.3058
3	-1.7440	0.1923	1.6413	3.5776
4	-2.8814	-0.8383	0.3781	2.4212