Anomaly detection for wind turbine gearbox oil pressure difference based on SCADA data

Y F Bo¹, X J Zeng¹, M Yang¹,3 and Y Zhu²

¹Key Laboratory of Power Intelligent Dispatch and Control (Shandong University), Ministry of Education, Jinan 250061, China
²Economic & Technology Research Institute of State Grid Shandong Electric Power Company, Jinan, China

E-mail: myang@sdu.edu.cn

Abstract. In this paper, a SCADA data-based fault detection method for gearbox oil pressure difference is proposed, and the core of the method is Spare Bayesian Learning (SBL) algorithm. The gearbox oil pressure difference probability estimation model can be constructed by training the historical normal operating based on SBL. Then, the probability distribution interval of the gearbox oil pressure difference can be estimated. The abnormal state of gearbox oil pressure difference can be judged by observing whether the actual value within the probability distribution interval. In addition, statistical hypothesis testing method is used to verify the reliability of the anomaly detection results. Through the method, the gearbox oil pressure difference anomaly detection problem can be transformed into a parameter estimation problem with low computational complexity. Case studies are conducted on two actual WTs with known gearbox oil pressure difference faults, and the results demonstrate the feasibility and effectiveness of the method.

1. Introduction

Wind power generation has developed rapidly in the past decade, and still maintains a high installed capacity every year. However, due to the complex structure and harsh operating environment of wind turbines (WTs), the failure rates of WTs are usually higher than that of conventional steam/hydro/gas turbines [1]. Therefore, more operation and maintenance (O & M) costs are required to ensure the reliable operation of WTs. It is reported that the O & M costs can account for 15%~20% of the total electricity production cost of a WT, and this ratio will be higher for off-shore WT because of their higher initial capital cost and limited accessibility [2]. Wind turbine anomaly detection based on Supervisory Control and Data Acquisition (SCADA) is a possible solution, which can detect the incipient anomalies of WTs before they turn to serious faults. Then, the failure rates of WTs can be effectively reduced by developing appropriate maintenance plans, and the O & M costs can be significantly reduced.

Anomaly detection based on SCADA data is often defined as the process of detecting the measurements change of WTs (e.g., temperature, pressure and vibration), and the significant change of those measurements under normal operation condition may indicate the early failures [3]. Since the gearbox directly affects the energy conversion efficiency of WTs, the downtime and replacement costs caused by gearbox fault are usually higher the other components of WT [4,5]. Therefore, it is of great significance to perform anomaly detection on the gearbox of WTs, which can extend the service life of
the gearbox and reduce downtime.

There are many efforts having been made on gearbox anomaly detection based on SCADA data. Because the gearbox oil is easy to measure and directly contact with the mechanical equipment inside the gearbox, such as gear, bearing and shaft, its abnormal state can reflect the abnormal state of the gearbox. Therefore, anomaly detection based on gearbox oil is one of the main methods for gearbox failure detection. Huang et al [6] combined principal component analysis (PCA) and dynamic neural network to prediction the gearbox oil temperature, from which the anomaly of gearbox oil temperature can be detected when the residuals between the actual values and the predicted values of oil temperature become significant. However, the reliability of the anomaly detection results dependent on the choice of the threshold, which lacks the corresponding selection criteria. Wang and Infield [7] proposed a method based on nonlinear state estimation technique (NSET) to detect the state of gearbox oil temperature, and Welch’s t-test was also used to enhance the accuracy of the detection results. However, the accuracy of the detection results will be affected if there are not enough state parameters included in the NSET model. Zhang and Qian [8] used a support vector machine (SVM) regression to construct a gearbox oil temperature prediction model, and the abnormal features are determined by the sequential feature selection algorithm. However, the SVM is prone to over-fitting and the computational efficiency will significantly reduced when the training data is large.

It can be seen that although the above oil temperature-based anomaly detection method can detect the incipient failures of the gearbox, there are also some disadvantages. Moreover, the SCADA data of oil temperature can be noisy due to the oil temperature is sensitive to the environmental conditions [9]. However, compared with the oil temperature, the gearbox oil (lubricant) pressure is less sensitive to external factors, can be considered as an alternative target for monitoring WT gearboxes. And the gearbox oil is applied to smooth the gearbox operation and its pressure can vary with the accumulation of swarf produced by the mechanical wear [10]. Therefore, the gearbox incipient faults can be reflected based on the gearbox oil pressure anomaly detection, and [11] detected the abnormal state of gearbox oil pressure based on the deep neural network. Although the method can find some gearbox oil pressure anomalies before the SCADA system fault alarm, the gearbox oil pressure is easily affected by temperature which may result mis-judgement.

In this paper a gearbox oil pressure difference anomaly detection method is proposed based on the SCADA data of WTs. The detection object is the difference between the inlet oil pressure and the outlet oil pressure of the gearbox filter. Then, the influence of ambient temperature on the gearbox oil pressure can be eliminated. Firstly, the Spare Bayesian Learning (SBL) algorithm is used to construct the gearbox oil pressure difference probability estimation model. Next, the probability distribution interval of the pressure difference can be estimated. Finally, according to observing whether the actual values fall into the estimated interval to judge the state of gearbox oil pressure difference. If the actual oil pressure difference exceeds the estimated interval, the phenomenon will be considered anomaly. And the judgement results are further tested by hypothesis testing. The feasibility and effectiveness of the proposed method are validated on one actual WT with known gearbox oil pressure faults.

The remainder of this paper is organized as follows. In section II, the gearbox oil pressure monitoring system and the principle of SBL algorithm are briefly introduced. In Section III, the anomaly detection process of gearbox oil pressure is designed. And the case study is conducted in Section IV. Finally, the conclusion of this paper is given in Section V.

2. The principle description

2.1. Gearbox oil pressure monitoring system

Gearbox oil plays an important role in the gearbox which can maintain the normal function of the gearbox and prolong its service life. According to forming an oil film on the gear surface and the bearing, the friction and wear between the mechanical components can be reduced and the corrosion of mechanical components by moisture and oxygen can be prevented. Meanwhile, the gearbox oil can take away a large amount of heat generated by the movement between the friction components. Based
on this principle, it is possible to judge whether the gearbox operating in abnormal state by detecting the temperature of gearbox oil. In addition, another important and easily overlooked function is that the gearbox oil can carry away impurities and debris from equipment wear in the gearbox. And in view of this characteristic, other gearbox oil anomaly detection methods are also used to determine whether the gearbox is malfunctioning or not. Such as viscosity analysis-based, particle counting and identification-based and oil pressure analysis-based. However, the first two approaches require special sensors and instruments for sampling and analysis, and it is difficult to perform timely detection because the analysis is usually periodically [12]. On contrast, gearbox oil pressure-based detection can be conducted using the SCADA data, which is simpler, faster and cheaper than the above two approaches.

Figure 1 shows the diagram of the gearbox oil pump system, it can be seen that there are two pressure sensors (A and B) which are located at the inlet and outlet of the gearbox oil filter, and one differential pressure detector. The sensors A and B detect the gearbox oil pressure at the inlet and outlet of the gearbox oil filter, respectively. The differential pressure controller is used to detect the oil pressure difference between the inlet oil pressure and out oil pressure. If the pressure difference is more than 0.5 bar and last for 20 s, the differential pressure controller will send an alarm signal, then, the WT will automatically stop. The reason results in the pressure difference is that there are too many impurities and debris blocking the filter element of filter, and this means that the moving components in the gearbox are seriously worn. Therefore, based on this principle, the abnormal operating state of the gearbox can be reflected by detecting the anomaly of gearbox oil pressure difference. Since the influence of ambient temperature can be avoided, using gearbox oil pressure difference to detect the abnormal state of the gearbox is more accurate.

![Figure 1. the diagram of the gearbox oil pump system.](image)

2.2. The principle of Bayesian inference

In general regression problems, when the relationship between the target value and the input vector is expressed:

\[ y_n = f(x_n, \omega) + \varepsilon_n \]  \hspace{1cm} (1)

where \( x_n \) is the input vector of the \( n \)th group of data; \( y_n \) is the target value of the \( n \)th group of data, which belongs to normal distribution \( N(f(x_n, w), \sigma^2) \); \( \omega = [w_0, w_1, \ldots, w_m]^T \) is the weight vector, \( \varepsilon_n \) as independent identical normal distribution \( N(0, \sigma^2) \), which is the forecast error. It can be seen from
The anomaly detection steps proposed in this paper are as follows:

3. **The process of anomaly detection**

The anomaly detection steps proposed in this paper are as follows:

- **Step 1:** According to the domain knowledge and the fault mechanism analysis, selecting the appropriate monitoring variables to estimate the gearbox oil pressure difference.
- **Step 2:** Pre-processing the original data to eliminate the noise and useless information, using Principal Component Analysis (PCA) to reduce the dimension of the data to improve the learning efficiency of the model.
- **Step 3:** Estimating the distribution of the parameters of SBL model by learning the training data. Then, use the reliability index to select the optimal parameters to construct the gearbox oil pressure difference model.
- **Step 4:** Based on the constructed model, estimating the probability distribution interval of the gearbox oil pressure difference in the target time period. Then, determining whether the pressure difference of gearbox oil is abnormal by observing whether the actual values of oil pressure difference fall into the estimated interval, and the anomaly detection results will be
further verified by HT.

4. Case study

In order to verify the effect of the proposed method for gearbox oil pressure difference anomaly detection, an experiment was conducted based on the SCADA data of a wind turbine with known gearbox oil pressure difference fault. The sampling interval of SCADA data is 1 minute, and one-week normal operation data were used for training the parameters of the model. 12-hour normal operation data was used for testing the estimated model. Another 12-hour data which contains the fault alarm information was used for gearbox oil pressure difference anomaly detection. In addition, the reliability of the estimation model at a certain confidence level is evaluated by the coverage index in the test model.

\[
P_c = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(y^*_{i} \in [U^\alpha_i, L^\alpha_i])
\]

where \(P_c\) is the interval coverage index, \(N\) is the total number of samples, subscript \(i\) denotes the \(i\)th detection point, \([U^\alpha_i, L^\alpha_i]\) denotes the distribution interval corresponding to the \(i\)th detection point at a given confidence level \((1- \alpha)\). The larger the value of \(P_c\), the more actual value is included in the estimation interval which indicates that the more reliable the estimated result.

Figure 2 shows the estimation results at the confidence of 90\% based on the 12-hour testing data, respectively. The red line represents the actual values of the gearbox oil pressure difference. The light blue area represents the estimated distribution range of gearbox oil pressure difference. The blue line represents the expectations of gearbox oil pressure difference, which always locate in the middle of estimation interval. It can be seen from figure 2 that the actual value of the oil pressure difference under normal operation condition is basically included in the estimated interval. The variation trend of the expected curve of gearbox oil pressure difference is similar to that of the actual pressure difference curve. Meanwhile, the coverage indicator under 90\% confidence level is 0.9431. It can be seen that the estimated interval can well cover the actual value pressure difference under normal operation condition, which indicates that the model has good reliability. Therefore, the estimation model can be used for gearbox oil pressure difference detection.

![Figure 2](image_url)

**Figure 2.** The estimation results based on testing data at different confidence.
Based on the same model, the abnormal gearbox oil pressure difference in the target period was detected. According to the failure log, there is a fault shutdown caused by gearbox oil pressure difference over-limit during the period. Figure 3 shows the estimated results at the confidence of 90%. It can be seen that although the actual values of the gearbox oil pressure difference occasionally exceed the estimated interval before the 506th point, most of the actual values are covered in the estimated interval. The estimation result is consistent with the estimation results of the testing data, which indicates that during this period the gearbox oil pressure difference is still within the normal range. However, from the 506th point, the pressure difference exceeds the upper limit of the estimated range and has a continuous upward trend until the fault shutdown that the pressure difference drops rapidly. Therefore, it is reasonable that the gearbox oil pressure difference is abnormal during this period (the ellipse region). Namely, our method can find the gearbox oil abnormal oil pressure difference one hour before the fault alarm.

![Figure 3. The gearbox oil pressure difference anomaly detection results at 90% confidence.](image)

In addition, from the perspective of physical mechanism, the phenomenon that the gearbox oil pressure difference continues to be high during the period (the ellipse region) can also be explained. As mentioned before, the pressure difference between the inlet and outlet of the gearbox oil pump filter is to ensure the lubricant passes through the filter element, then, to remove the impurities and debris from the oil. The gearbox oil can pass through the filter element smoothly under normal condition, and the pressure difference is small. However, when the impurities and debris accumulate too much on the filter element, the flow of the oil will be hindered. Then the oil pump will push the lubricant pass through the filter element by increasing the inlet oil pressure which results in the pressure difference increased. Moreover, when the pressure difference reaches the upper limit, the SCADA system will send a fault alarm which results in the WT shutdown automatically. The oil pump will also stop working and the pressure difference of the oil drops rapidly. This is the reason why the pressure difference curve drops rapidly after the fault alarm in figure 3.

In order to verify the accuracy of the anomaly detection results based on the observation method. A statistical hypothesis testing method was utilized. Since the value of the actual pressure difference has only two states relative to the estimation interval: within or outside the interval. If the value within the interval is defined as 1, and the value outside the interval is defined as 0. Then, the original problem can be converted to the Bernoulli hypothesis testing problem. We selected all points before the fault as samples and set the following assumption at the 0.05 significance level:
\( H_0: \) the actual values within the estimation area less than 90%.
\( H_1: \) more than 90% actual values fall within the prediction range.

The hypothesis testing was conducted on MATLAB software and the result accepted the null hypothesis \( H_0 \). Therefore, the method for gearbox oil pressure difference proposed in this paper is reliable.

5. Conclusion
The anomaly detection for gearbox operation state based on SCADA data can effectively reduce the O & M costs. The current SCADA data-based methods mainly use temperature as detection object, but the temperature is susceptible to the environment, resulting in more noise in the data, which will reduce the accuracy of the anomaly detection. Therefore, in this paper, anomaly detection based on gearbox oil temperature difference was proposed, which can eliminate the influence of temperature. Meanwhile, a detection method based on SBL was presented, which can reliably estimate the probability distribution interval of gearbox oil pressure difference through data training. Then the anomaly detection results can be obtained by observing whether the actual values fall into the estimated interval. In addition, the hypothesis testing method was also used to verify the reliability of the anomaly detection results. Finally, the case study was conducted on an actual WT with known gearbox oil pressure faults, to validate the feasibility and effectiveness of the proposed method, and the gearbox oil pressure difference can be found one hour before the alarm fault.

Through this research, a simpler, faster but more reliable and effective wind turbine gearbox anomaly detection method is heralded. Should be noted that the parameters used to construct the model in this paper can be further optimized. Therefore, the method will be further developed to improve the accuracy of the estimation model, and more actual WTs will be used to verify the reliability of the detection results.

Acknowledgments
This work was supported by a Science and Technology Project of State Grid Corporation of China (No. 52060018000X)

References
[1] Qiao W and Lu D 2015 A survey on wind turbine condition monitoring and fault diagnosis— Part I: Components and subsystems IEEE T. Ind. Electron. 62 6536-45
[2] Cui Y, Bangalore P and Tjernberg L B 2018 An anomaly detection approach using wavelet transform and artificial neural networks for condition monitoring of wind turbines' gearboxes 2018 Power Systems Computation Conf (PSCC) (Dublin) pp. 1-7
[3] Dao P B, Staszewski W J, Barszcz T and Uhl T 2018 Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data Renew. Energ. 116 107-22
[4] Feng Y, Qiu Y, Crabtree C J, Long H and Tavner P J 2013 Monitoring wind turbine gearboxes Wind Energy 16 728-40
[5] Wang J, Gao R X and Yan R 2014 Integration of EEMD and ICA for wind turbine gearbox diagnosis Wind Energy 17 757-73
[6] Huang Z S, Xiang D and Wei Y Z 2018 Prediction of oil temperature variations in a wind turbine gearbox based on PCA and an SPC-dynamic neural hybrid J Tsinghua Univ (Sci & Technol) 58 539-46
[7] Wang Y and Infield D 2013 Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring Renewable Power Generation Int 7 350-8
[8] Zhang D and Qian Z 2017 Probability warning for wind turbine gearbox incipient faults based on SCADA data 2017 Chinese Automation Congress (CAC) (Jinan, China) pp. 3684-3688
[9] Sequeira C, Pacheco A, Galego P and Gorbena E 2019 Analysis of the efficiency of wind turbine gearboxes using the temperature variable Renew. Energ. 135 465-72
[10] Nutakor C, Talbot D and Kahraman A 2019 An experimental characterization of the friction coefficient of a wind turbine gearbox lubricant *Wind Energy* **22** 509-22

[11] Wang L, Zhang Z, Long H, Xu J and Liu R 2017 Wind turbine gearbox failure identification with deep neural networks *IEEE T. Ind. Inform.* **13** 1360-8

[12] Qiao W and Lu D 2015 A survey on wind turbine condition monitoring and fault diagnosis—Part II: Signals and signal processing methods *IEEE T. Ind. Electron.* **62** 6546-57

[13] Tipping M E 2001 Sparse Bayesian learning and the relevance vector machine *J. Machine Learning Research* **1** 211-44