An Anomaly Detection Approach Using Wavelet Transform and Artificial Neural Networks for Condition Monitoring of Wind Turbines' Gearboxes

Yue Cui, *Student Member, IEEE*
Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden
ycui@kth.se

Pramod Bangalore
Greenbyte AB
Gothenburg, Sweden
pramod@greenbyte.com

Lina Bertling Tjernberg, *Senior Member*, *IEEE*
Electrical Engineering
KTH Royal Institute of Technology
Stockholm, Sweden
linab@kth.se

***Abstract*—This paper presents an anomaly detection approach using artificial neural networks and the wavelet transform for the condition monitoring of wind turbines. The method aims to attain early anomaly detection and to prevent possible false alarms under healthy operations. In the approach, nonlinear autoregressive neural networks are used to estimate the temperature signals of the gearbox. The Mahalanobis distances are then calculated to measure the deviations between the current states and healthy operations. Next, the wavelet transform is applied to remove noisy signals in the distance values. Finally, the operation information is considered together with the refined distance values to detect potential anomalies. The proposed approach has been tested with the real data of three 2 MW wind turbines in Sweden. The results show that the approach can detect possible anomalies before failure events occur and avoid reporting alarms under healthy operations.**

***Index Terms*—condition monitoring system, neural networks, the wavelet transform, and wind power.**

I. INTRODUCTION

Renewable energy has been developing rapidly in the past years. It is projected that the European electricity demand will reach nearly 400GW by 2030 and 28.5% of the total demand is expected to come from wind power [1]. However, wind power is often criticized for its high operation and maintenance (O&M) cost [2]. It is reported that this expenditure can account for 23~30% of the total electricity production cost of a wind turbine [3]. One possible solution is to develop condition-monitoring systems for wind turbines to find those hidden malfunctions. By conducting condition monitoring, incipient anomalies can be detected and corrected before they turn to severe faults. This makes it possible to maintain the wind turbines' reliability by using preventive maintenance, which could be more cost effective than corrective maintenance [4]. The corresponding maintenance activities can be also optimized with lower cost.

Traditional condition monitoring for wind power mainly focuses on analyzing vibration signals. These signals are of high frequencies and are sampled during the operations of rotational components. However, the measurement requires auxiliary sensors, hence results in additional expenditure [5], [6]. Another method is to directly exploit the data stored in supervisory control and data acquisition (SCADA) systems. The rich information makes it feasible to conduct systematic condition monitoring for wind turbines. The alarm data produced by SCADA systems can also help to identify the operation states of wind turbines [7]. In the recent literature, machine learning techniques are introduced into this application. Artificial neural networks (ANN) have been successfully applied to assess the performance of wind turbines [8] and detect failures of subassemblies, including the gearbox [9], generator [10], blade [11] and bearing [12]. Other techniques, like the fuzzy technology [13] and multi-agent systems [14], are applied to the filed as well. However, possible false alarms under healthy operations have not been fully discussed yet, although these papers are tested to be capable of detecting anomalies. The false alarms can not only decrease the sensitivity of condition monitoring systems, but also cause unnecessary inspections. Hence the essential problem is how to inform risks before failures happen and to avoid false alarms under healthy operations.

This paper aims to provide an anomaly detection approach, which is based on the self-evolving maintenance scheduler (SEMS) framework [9]. The main contribution of this paper is the introduction of wavelet denoising as posterior processing applied to the SCADA based condition monitoring method presented within the SEMS framework. After removing noisy signals, the denoised results help to detect incipient malfunctions in advance and to reduce the possibilities to report false alarms under healthy operations.

The paper is organized as followed. Section II introduces the theoretical tools of the anomaly detection approach. The details of the approach are presented in Section III. Section IV introduces the results and corresponding analysis of case studies. The conclusions are summarized in Section V.

II. PRELIMINARIES

A. Nonlinear Autoregressive Neural Networks

Nonlinear autoregressive neural networks with exogenous inputs (NARX) networks are a kind of dynamic artificial neural networks. They are widely used to predict one time series with its given past values and the other independent time series. The NARX model can be expressed as followed:

$$y(t) = f\begin{pmatrix} y(t-1), y(t-2), \cdots, y(t-n_y), \\ u(t), u(t-1), \cdots, u(t-n_u) \end{pmatrix} \quad (1)$$

where $y(t)$ is the output series and $u(t)$ is the independent input series. n_y and n_u are respectively the delay number of the two series. Fig. 1 presents the structure plot of a NARX network.

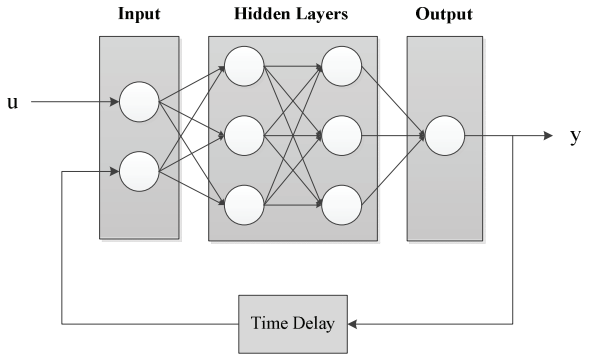

Figure 1. The structure plot of NARX networks.

B. The Wavelet Transform

A wavelet is defined as a function of $\psi \in L^2(R)$ which satisfies the admissibility condition:

$$C_\psi \equiv \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (2)$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. Given a single mother wavelet function, a family of its daughter wavelet functions can be obtained by using dilations and translations:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \ a, b \in R, \ a \neq 0 \quad (3)$$

where a is a dilation parameter which measures the scale effect and translation b is a translation parameter which determines the time location of the wavelet.

For $\psi \in L^2(R)$, the continuous wavelet transform W_ψ is an integral transformation which is defined on $L^2(R)$:

$$W_\psi[f](a, b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(t) \bar{\psi}_{a,b}(t) dt \quad (4)$$

where $\psi_{a,b}(t)$ is a daughter function of the mother function ψ.

The continuous wavelet transform is capable of extracting features in both time and frequency domains, however it involves much computation work during the implementation process [15].

The discrete wavelet transform is developed as a discretized version of the continuous wavelet transform, which can be implemented in high-dimension calculations [16], [17]. The dyadic wavelet transform is a popular form to discretize the continuous wavelet transform with the dilation parameter $a = 2^j$ and the translation parameter $b = 2^j n$. The discrete wavelet transform can be defined as the following doubly indexed sequence:

$$DTW_{j,n}(f) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t) \bar{\psi}_{a,b}\left(\frac{t}{2^j} - n\right) dt, \ j, n \in Z \quad (5)$$

In this paper, the discrete wavelet transform is implemented by the pyramid algorithm [18], which is approximated at different decomposition levels.

III. THE ANOMALY DETECTION APPROACH

In this paper, an anomaly detection approach for wind turbines is proposed to monitor temperature signals, to estimate current operation states and to report warnings and alarms if necessary. The approach mainly consists of four steps:

Step 1. Nonlinear autoregressive neural networks with exogenous inputs are trained and applied to estimate the temperature signals;

Step 2. The Mahalanobis distances are calculated to assess the deviations between the estimated values and the actual values measured by SCADA systems;

Step 3. The wavelet transform is applied to remove noisy signals in the distance values;

Step 4. Anomaly analysis based on the operation information and the refined distance values.

where the step 1&2 are presented in [9] and step 3&4 are developed and extended in this paper. The flow chart of the detailed approach is shown in Fig. 2.

The data from SCADA systems are applied as data input, which sampled and averaged every ten minutes. The ANN models are trained to emulate the equilibrium condition during normal operation of the wind turbines; i.e. when the turbines are producing power. Hence, the application of the proposed approach is also done only when the wind turbine is producing power.

In the preprocessing part, the raw data are processed by three steps to remove inconsistent data. First, the records with missing attribute values are deleted in the filter. Second, the data are classified into clusters which denote the different states under normal operations. The clustering is conducted based on the approach that is presented in [19]. In each cluster, the Mahalanobis distances of the data points are calculated. The values are assumed to be represented by probability functions. In this paper, normal functions are considered. If the distance value of a record locates outside three standard

deviations from the mean value, then it is assumed to be an outlier and is deleted in the filter. Third, the remained data are normalized to the standard normal distribution in case of the scale effect.

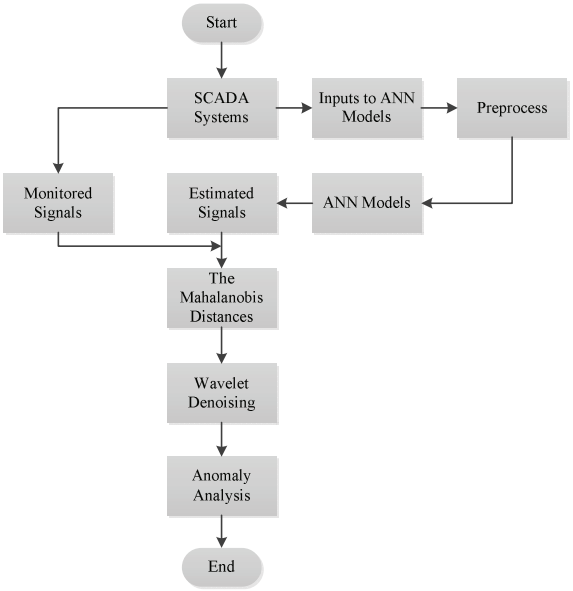

Figure 2. Flow chart of the proposed anomaly detection approach.

At Step 1, NARX networks are trained by Levernberg-Marquardt back propagation algorithm. The training performance is evaluated by mean squared errors.

At Step 2, the approach applies the Mahalanobis distance to assess the deviations between the measured values of the temperature signals and their estimated values calculated from the NARX networks. The Mahalanobis distance measures the similarity between two sample sets. Unlike the Euclidean distance, it considers the correlation of different variables. For two independent sample set x and y, if they are assumed to be produced by the same probability distribution, then their Mahalanobis distance is calculated as:

$$D_M(x,y) = \sqrt{(x-y)^T\, C^{-1}(x-y)} \tag{6}$$

where C is the covariance matrix of x and y. When C is a unit variance matrix, then the Mahalanobis distance corresponds to the Euclidean distance. The approach first calculates the absolute errors of estimated values and real measurements of the temperature signals. Second, the Mahalanobis distances of the errors against those under the healthy states are calculated as the evaluated indicators to measure the deviations between current states and healthy operations.

At Step 3, the Mahalanobis distances obtained from the previous step are filtered by the wavelet denoising. The distances are decomposed at different levels based on the wavelet base functions. And, the obtained detail coefficients are limited to the threshold at each level. Subsequently, the denoised distances are reconstructed based on the modified coefficients and the wavelet base functions.

At Step 4, the operation information is considered together with the refined distances to determine the operation states. If an anomaly is detected, then corresponding warnings and alarms will be reported to inform the potential operation risks. The anomaly analysis mainly consists of three stages and its flow chart is presented in Fig. 3.

Stage I. Define threshold.

The threshold determines the upper limit of the Mahalanobis distances of healthy states. The approach calculates the denosied Mahalanobis distances under healthy operations. The distances are assumed to be represented by a certain probability distribution. Since the samples are large enough, normal distributions are considered in this paper. Next, the assumption is tested by Kolmogorov-Smirnov statistics test. If it cannot be rejected, then a normal distribution is fitted and the threshold is defined at a low probability (0.001). Otherwise, new probability distributions are tested until the corresponding assumptions cannot be rejected.

Stage II. Compare distance values with the threshold.

The denosied Mahalanobis distances are compared with the threshold that is calculated at previous stage. If the distance is below the threshold, then the wind turbine is considered healthy and no anomalies exist in the component. However, if the calculated distances exceed the threshold, and the wind turbine is producing power, then the process steps in to the next stage.

Stage III. Report warnings and alarms.

At this step, the approach defines warnings and alarms to measure the different risk degrees of the possible anomalies. A warning is triggered when the distance value crosses the threshold, which indicates that the possible operation risks exist but not dangerous. If warnings are reported continuously, say more than two hours, then an alarm is triggered. This warns operators of the potential incipient anomalies and further inspection activities should be arranged in time in case of severe damages of the component.

IV. CASE STUDIES

The proposed approach is examined with the data from the experience of three different wind turbines. All the wind turbines are rated 2 MW and onshore located in Sweden. This paper mainly focuses on the gearbox component. Since the failures in the gearbox tend to cause abnormal increase in temperature signals [20], two signals, including gearbox bearing temperature and lubrication oil temperature, are respectively tested as the output of two NARX networks. Based on [13], four signals are tested as the input signals, including the nacelle temperature, rotor speed, ambient temperature, and grid production power. Both the networks have same signals as input. In the following three examples, the networks are respectively trained and tested in different periods, where the detailed information is shown in Table I.

TABLE I. THE TRAINING AND TEST PERIODS OF THE THREE TEST CASES.

Cases	Training Data	Test Data

A	Jan. 01, 2010 - Dec. 31, 2010	Jan. 01, 2011 - Nov. 19, 2011
B	Jan. 01, 2011 - Dec. 31, 2011	Jan. 01, 2012 - Feb. 15, 2013
C	Jan. 01, 2010 - Dec. 31, 2010	Jan. 01, 2011 - Feb. 29, 2012

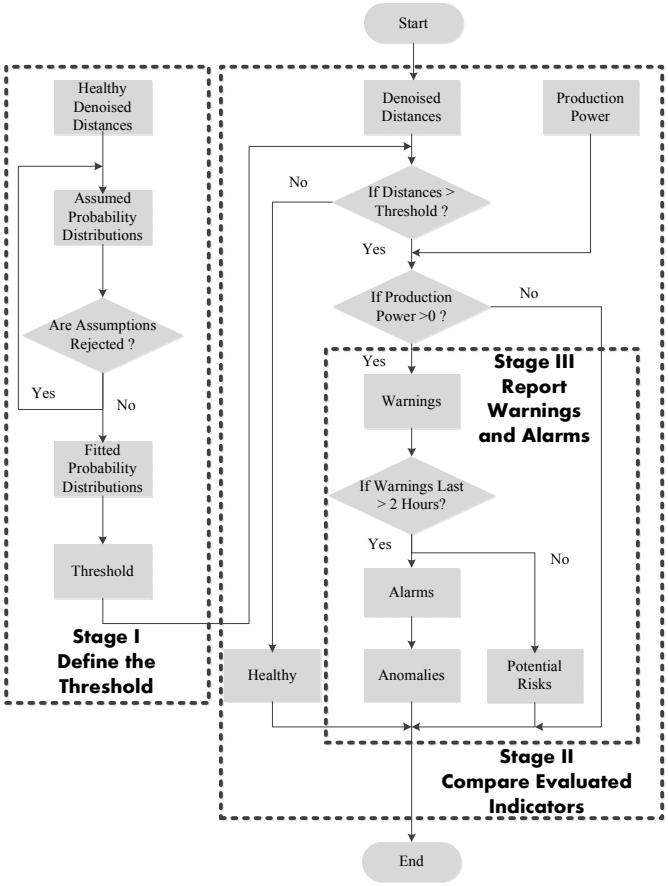

Figure 3. Flow chart of the anomaly detection analysis.

A. Case A: a wind turbine with a fault in the intermediate shaft bearing

In this example, the data of the wind turbine were recorded from the January, 2010. A fault was detected in the intermediate shaft bearing of the gearbox on November 19, 2011. The replacement was implemented later on after that.

The data records from the year of 2010 are applied for the training of NARX networks. In the clustering filter, the input data are clustered based on four signals, which are the rotor speed, the wind speed, the production power and the blade pitch angle [19]. These signals reflect the different operation states of main subsystems, like the drive train and the control system etc. To visualize the data that are removed in the filter, Fig. 4 and 5 respectively presents three performance curves before and after the clustering filter. The dots in different colors denote the data records that come from in different clusters. It can be seen that the dots which locate on the edge of the clusters are removed and the trends of the performance curves become clean after the filter.

In this example, the bearing network is applied with one input delay, one feedback delay and the lubrication network is applied with one input delay, two feedback delays through comparing different delays (from 0 to 2) in the trial test. Each network is trained three times and the network with least performance function is chosen in the approach. Fig. 6 compares the estimated values of the bearing temperature produced by the NARX networks and the corresponding measured values in SCADA systems. It can be seen that the estimated values exactly follow the fluctuations of the measured values and their deviations are not large from the figure. Furthermore, the relative errors of the two signals are presented as well and most errors locate within the range of ±5 percent.

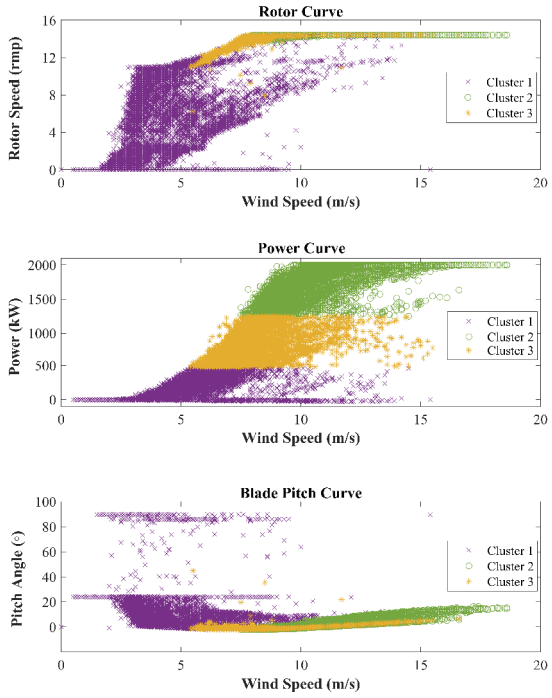

Figure 4. The rotor curve, power curve and blade pitch curve in the training period before the clustering filter in Case A.

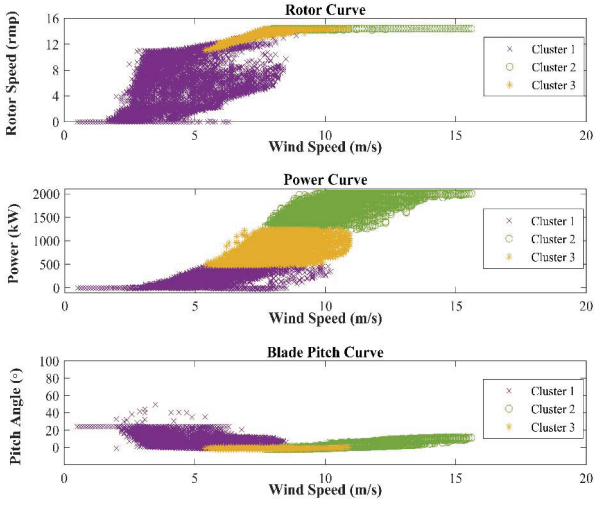

Figure 5. The rotor curve, power curve and blade pitch curve in the training period after the clustering filter in Case A.

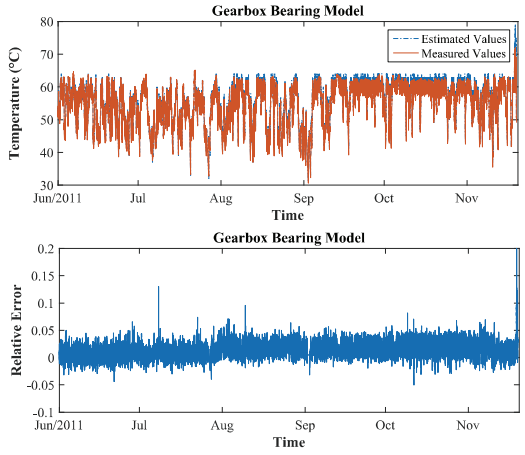

Figure 6. Comparison of the estimated values of the temperature signals from the bearing model and the relative errors in Case A.

During performing the wavelet denoising, different wavelet base functions are tested and compared in the trial test, including the Daubechies, Symlets and Coiflets wavelet. By comparing the results at same decomposition levels, the Daubechies wavelet is applied as the base function in the paper, since it can maintain information at low frequencies while deleting excessive noise signals. The denoising effect of the Mahalanobis distances with five-level denoising.is presented in Fig. 7. It can be seen that the distance curves become thinner after the denoising and the details are easy to identify. More levels are tested as well, like eight-level denoising, which is presented in Fig. 8. However, compared with Fig. 7, more levels do not show prominent improvement in the denoised signals. This indicates that the main noises have been removed through five-level denoising and further denosing may lead to the information losses. Hence, the wavelet denoising with five levels is applied in the following anomaly analysis.

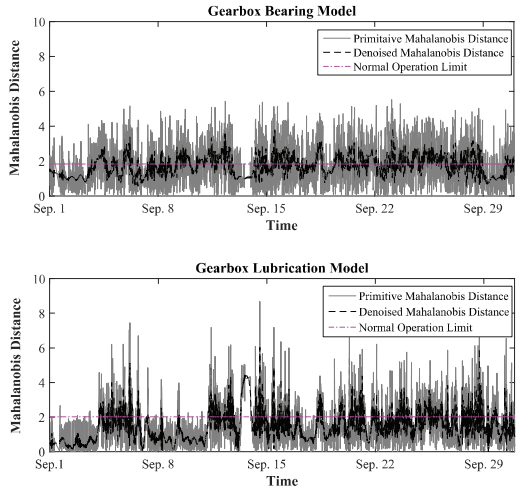

Figure 7. Comparison of primitive and denosied Mahalanobis distances at five-level denoising in Case A.

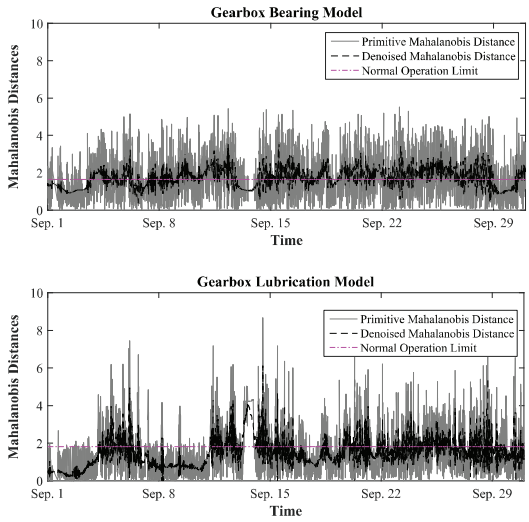

Figure 8. Comparison of primitive and denosied Mahalanobis distances at eight-level denoising in Case A.

Fig. 9 shows the anomaly analysis results in Case A. In the bearing model, the Mahalanobis distance reaches its peak values at July 27, 9:10, 2011; while in the lubrication model, the distance reaches its the second highest value at the same time. However, since the corresponding production power is -3.2 kW, which means that the wind turbine is outside normal operations, neither any warnings nor alarms are triggered at that moment. During the test period, 9381 warnings and 4301 warning are respectively reported in the bearing model and the lubrication model. 1280 alarms are totally reported in the bearing model and 244 alarms in the lubrication model. Table II summarizes the detailed information of the alarms that are reported in the two models. The first alarm of the bearing model happens on June 30, 17:20, and that of the lubrication model happens on June 6, 18:40. These alarms are five months earlier than the failure time.

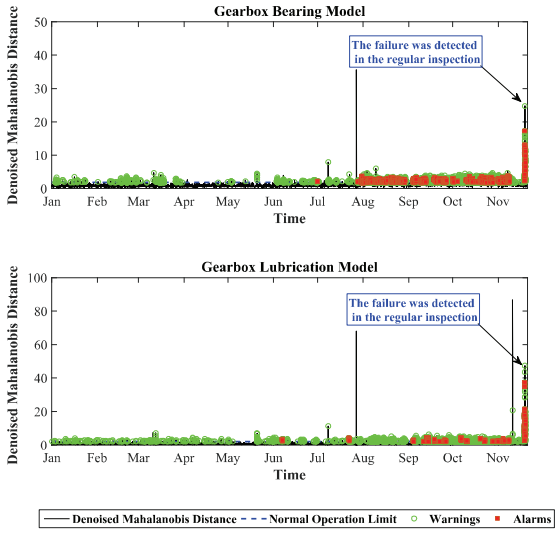

Figure 9. The anomaly analysis results in Case A.

TABLE II. THE DETAILS OF THE ALARMS THAT REPORTED IN CASE A.

Time / Models	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.
The bearing model	1	91	459	361	188	180
The lubrication model	18	9	13	60	50	94

In addition, the results without wavelet denoising are tested and shown in Fig. 10. In the bearing model, the first alarm is triggered only 13 hours ahead of the failure event; while in the lubrication model, alarms are reported on July 22nd, September 13th, October 6th, November 18th and 19th. The anomaly detection results of averaged Mahalanobis distances over three days that are used in [9] are presented in Fig. 11. The average distances do not cross the predefiend threshold so neither warnings nor alarms are triggered during the test period.

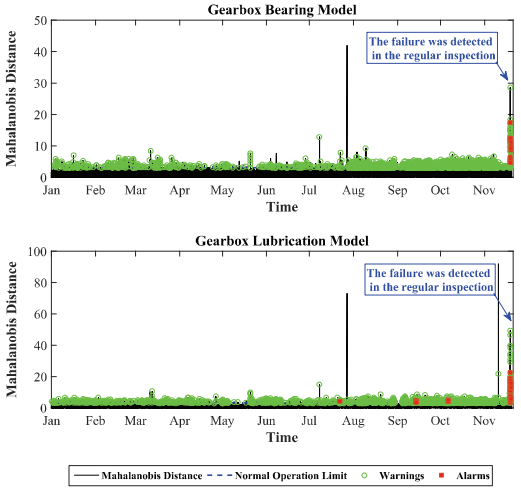

Figure 10. The anomaly analysis results without the wavelet denoising in Case A.

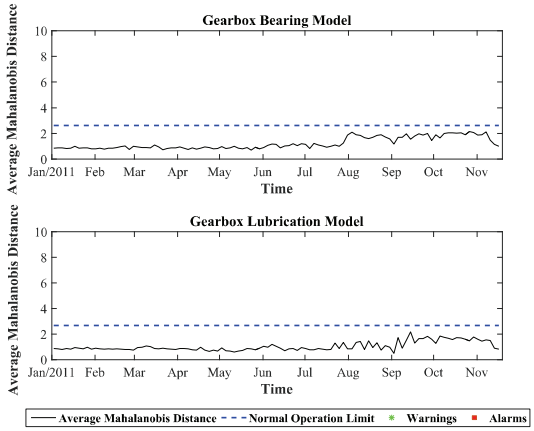

Figure 11. The anomaly analysis results with the average Mahalanobis distances in Case A.

It should be noted that the time from the first alarm to the failure time can be calculated with some uncertainty, but the effect of this unceratinty can be reduced by repeating the implementations of the approach. In the following period, more alarms are continuously reported, which demonstrates the operaion condition of the gearbox component is deteriorating. The discussion about the number of alarms is not included in the paper, which will be implemented in the further work. But what can be perceived is that more alarms indicate the potential changes in the component's health with higher possibilities compared with one single alarm.

B. Case B: A wind turbine with a fault in the rotor end bearing

In this example, the data of the wind turbine were recorded starting from November 20, 2010. A fault located in the rotor end bearing, was detected in the gearbox component on November 23rd, 2012. The replacement was implemented on February 15th, 2013. The models that are used in this example are same to Case A with the same parameters. The anomaly analysis results are shown in Fig. 12. 5012 warnings and 296 alarms are reported in the bearing model and 3666 warnings and 2 alarms in the lubrication model. In the former model, the first alarm is triggered on September 9, 7:50, 2012, which is two months before the failure event. In the latter model, the alarms are reported only on December 24, 2012. The results imply that the anomaly in the bearing is not severe to cause obvious health deterioration in the lubrication system.

C. Case C: A wind turbine without reported failures

In this case, the wind turbine was monitored from December 15th, 2009 to February 29th, 2012. During the monitoring period, no fault was detected in the regular inspections. Fig. 13 presents the anomaly detection results in the test period. Considering that it is difficult to determine an alarm is true or false at the time before a failure happens, this paper only discusses false alarms under healthy operations. Unlike Case A and B, all the denosied Mahalanobis distances are far below the threshold. No warnings nor alarms are triggered in both the models. This example certifies that the proposed approach can avoid reporting false alarms under healthy operations.

From the above results in the three cases, it is concluded that the proposed anomaly detection approach, on one side, can achieve the detection of incipient malfunctions in wind turbines before failure events; on the other side, it can also help to prevent reporting false alarms under healthy operations. It should be stated that the time from the first alarm to a failure can vary from cases to cases, depending on the failure types, the severe degree and many factors. But no matter when an anomaly is detected, the proposed approach does make sense if it can be corrected in time before it turns to a failure. What should be also noted is that an approach should be generalized after examined by more test cases. However since SCADA data are not public, the test examples are constrained by the access to the data. In future work, more

data will be used to examine and improve the proposed approach.

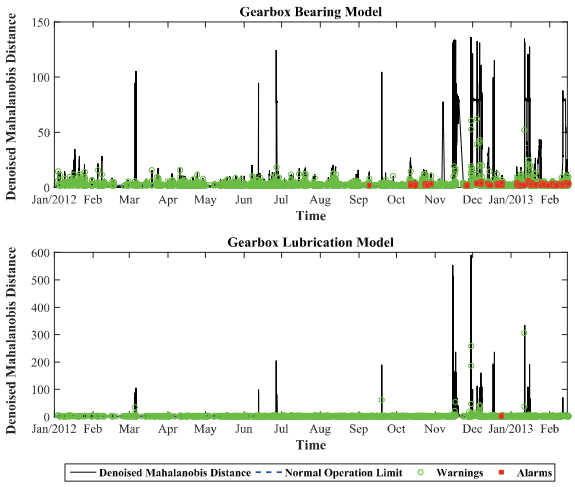

Figure 12. The anomaly analysis results in Case B.

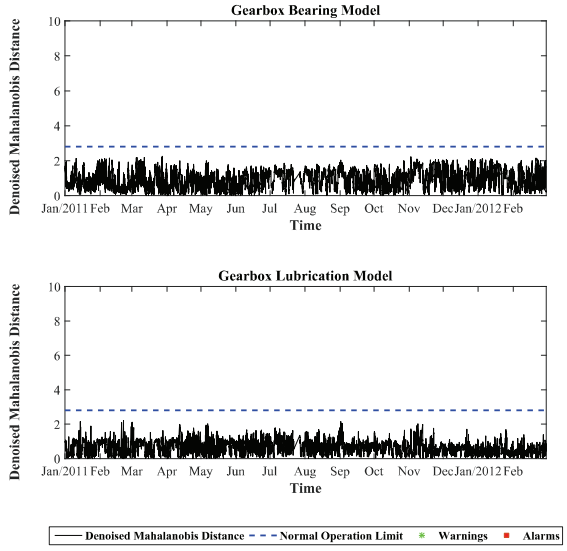

Figure 13. The anomaly analysis results in Case C.

V. CONCLUSIONS

In this paper, the authors propose an anomaly detection approach using artificial neural networks and the wavelet transform for the condition monitoring of wind turbines. The approach integrates machine learning technologies with the operation information to implement anomaly detection. The main contribution of this paper, compared with [9], is to apply the wavelet denoising as post processing to delete the effect of noise signals. The propose approach is tested with the data from three wind turbines. The results demonstrate that on one side, the approach can detect possible amonalies and inform potential operation risks before failure events; on the other side, the approach is capable of preventing false alarms towards healthy operation states. The further work in the next step is to test and improve the approach with more test data and apply the approach to other components.

ACKNOWLEDGMENT

This work was supported by Chinese Scholarship Council. Also, the authors would like to acknowledge the data support from Stena Renewables to this paper.

REFERENCES

[1] J. Moccia et al., "Pure power-wind energy targets for 2020 and 2030," European Wind Energy Association, 2011.
[2] T. W. Verbruggen, "Wind turbine operation and maintenance based on condition monitoring WT-Ω Final Report," Energy Research Center of the Netherlands, ECN-C-03-047, 2003.
[3] E. J. Wiggelinkhuizen et al., "Conmow Final Report," Energy Research Center of the Netherlands, ECN-E-07-044, 2007.
[4] L. Bertling Tjernberg, *Infrastructure Asset Management with Power System Examples*, CRC Press Taylor and Francis, pp. 580, 2018.
[5] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: a review," Renewable and Sustainable Energy Reviews, vol. 13, pp. 1-39, 2009.
[6] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez and M. Papaelias, "Condition monitoring of wind turbines: techniques and methods," *Renewable Energy*, vol. 46, pp. 169-178, 2012.
[7] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," *Renewable Energy*, vol. 36, pp. 16-23, 2011.
[8] E. Lapira, D. Brisset, H. D. Ardakani, D. Siegel and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach," *Renewable Energy*, vol. 45, pp. 86-95, 2012.
[9] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 980-987, 2015.
[10] M. Schlechtingen and I. F. Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical systems and Signal Processing*, vol. 25, pp. 1849-1875, 2011.
[11] L. Wang, Z. Zhang, J. Xu and R. Liu, "Wind turbines blade breakage monitoring with deep autoencoders," *IEEE Transactions on Smart Grid*, in press.
[12] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: a data-mining approach," *Renewable Energy*, vol.48, pp. 110-116, 2012.
[13] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on SCADA data using normal behavior models. part 1: system description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259-270, 2013.
[14] A. Zaher, S. D. J. McArthur, D. G. Infield and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574-593, 2009.
[15] L. Debnath and F. A. Shah, *Wavelet transforms and their applications*, Second Edition, New York: Springer Science, 2015.
[16] S. J. Waston, B. J. Xiang, W. Yang, P. J. Tavner and C. J. Crabtree, "Condition monitoring of the power output of wind turbine generators using wavelets," *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 715-721, 2010.
[17] A. Singh and A. Parey, "Gearbox fault diagnosis under fluctuating load conditions with independent angular resampling technique, continuous wavelet transform and multilayer perceptron neural network," *IET Science, Measurement & Technology*, vol. 11, pp. 220-225, 2017.
[18] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, 1989.
[19] A. Kusiak and A. Verma, "Monitoring wind farms with performance curves," *IEEE Transactions on Sustainable Energy*, vol.4, pp. 192-199, 2013.
[20] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis-part I: components and subsystems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6536-6545, 2015.