applied sciences

Article

Wind Turbine Anomaly Detection Using Mahalanobis Distance and SCADA Alarm Data

Jui-Hung Liu [1,*], Nelson T. Corbita, Jr. [2], Rong-Mao Lee [3] and Chun-Chieh Wang [4]

1 Department of Mechanical Engineering, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan
2 Department of Mechanical Engineering, University of Science and Technology of Southern Philippines, Cagayan de Oro 9000, Philippines
3 Department of Intelligent Automation Engineering, National Chin-Yi University of Technology, Taichung 41170, Taiwan
4 Mechanical and Mechatronics Systems Research Laboratory, Industrial Technology Research Institute, Hsinchu 310401, Taiwan
* Correspondence: dofliu@stust.edu.tw

Abstract: Wind energy is becoming a common source of renewable energy in the world. Wind turbines are increasing in number, both for onshore and offshore applications. One challenge with wind turbines is in detecting anomalies that cause their breakdown. Due to the complex nature of the wind turbine assembly, it is quite an extensive process to detect causes of malfunctions in the system. This study uses the Mahalanobis distance (MD) to detect anomalies in wind turbine operation, using SCADA alarm data as a comparison. Different predictive models were generated as the bases for analyses in MD computations. Using the SCADA alarm data as a reference, trend patterns that deviated from the threshold value were compared. Results showed that the MD could be used to detect anomalies within a group of data sets, with behaviors learned based on the model used. A large portion of those data sets deviated from the threshold level, corresponding to serious alarms in the SCADA data. We concluded that the MD can detect anomalies in different wind turbine components, based on this study. MD analysis of models can be used in conditions monitoring systems of wind turbines.

Keywords: wind turbine; Mahalanobis distance; anomaly detection; SCADA alarm; predictive models

Citation: Liu, J.-H.; Corbita, N.T., Jr.; Lee, R.-M.; Wang, C.-C. Wind Turbine Anomaly Detection Using Mahalanobis Distance and SCADA Alarm Data. *Appl. Sci.* **2022**, *12*, 8661. https://doi.org/10.3390/app12178661

Academic Editors: Richard (Chunhui) Yang and Chengjung Yang

Received: 26 July 2022
Accepted: 26 August 2022
Published: 29 August 2022

1. Introduction

1.1. Background of the Study

One crucial aspect of wind turbine maintenance is the reliability of the anomaly detection system used for failures/faults. Such a system can improve the performance of the wind turbine, as the component where failure is detected would be given proper maintenance [1]. This method also reduces costs associated with maintenance, as affected components will become the focus of maintenance work. With the current trends in improving monitoring systems through research, anomaly detection is a significant part of this.

Different methods in anomaly detection can be developed and incorporated into the wind turbine system. Alarms in the SCADA system can be used as references for determining the types of anomalies. Still, due to the complex and solitary function of the SCADA system, it is seldom used in conditions monitoring systems [2]. In-depth failure analysis of wind turbine components takes much time and effort, and comes with corresponding costs. Currently, an emerging method in anomaly detection is data analysis in anomaly detection [3–5]. Contemporary research uses data analysis in detecting anomalies, conveniently with the aid of computing software. In this paper, our aim is to use SCADA data and the Mahalanobis distance approach as the primary method of detecting anomalies.

The main objective of this study is to incorporate the use of operational SCADA data into the anomaly detection feature of a condition monitoring system. Different models will be created and used for the data analysis using the Mahalanobis distance approach. A comparison with SCADA alarm data will verify the model's performance using the MD. The operational data used in this study were extracted from a wind farm in Taiwan with fourteen wind turbines, during the year 2019.

The focus of this study is the development of an anomaly detection method for wind turbines using SCADA operational data and the MD approach. Since this study is still in the developmental stage, our approach was not yet integrated into a wind turbine's actual CMS. Furthermore, since different parameters are involved in different wind turbine SCADA designs, this means that minor changes in this approach should be applied when testing other wind farms [6]. Furthermore, the wind farm tested in this study is an onshore, direct-drive type system.

This study will use the Mahalanobis distance to measure distance variations in the data set generated from predictive models. This measure can provide insights into the duration of the generated data where abnormal behavior can be observed. In this way, abnormalities in the wind turbine can be checked using the SCADA alarm data generated throughout the study.

In this study, anomalies detected by the MD method were analyzed based on the SCADA alarm data. No count or measure of the frequency of various anomalies was conducted. This means that there was no further analysis conducted in terms of the frequency of alarms that were generated during the duration of the anomaly detected from the MD analysis.

The following sections describe the Mahalanobis distance, which is the main approach used in this study, followed by a summary of related research focused on the use of the MD in fault detection and SCADA systems for conditions monitoring of wind turbines. A description of the wind turbine data used and a description of the proposed method are then detailed in the Materials and Methods section. This is then followed by a discussion of the results and the conclusions, which address the future scope of this research.

1.2. Mahalanobis Distance

The Mahalanobis distance is the distance between two points in multivariate space. This is similar to the Euclidian distance, except there is also a correlation between variables that is involved for the MD approach [7]. One of the most common purposes of the MD is in locating outliers in the multivariate space. This is commonly used in applications such as fault detection in some measuring tools. The Mahalanobis distance between two objects can be defined as follows by Equation (1):

$$d_{MD} = \sqrt{(x_b - x_a)^T * C^{-1} * (x_b - x_a)} \quad (1)$$

where x_a and x_b refer to a pair of objects, while C is the sample covariance matrix.

The advantage of using the Mahalanobis distance lies in its capacity to solve for the limitation of the Euclidian distance. Since using the Euclidian distance in a multivariate space would be inappropriate as it only deals with distances between points, the Mahalanobis distance solves this by measuring the distance between the point and the distribution itself [8]. This is achieved by using the covariance of the matrix distribution in the calculation, as shown in Equation (2):

$$cov_{x,y} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{N-1} \quad (2)$$

where x_i and y_i represent data values of the distributions, and $\overline{x}$ and $\overline{y}$ refer to the means of the distributions. Figure 1 is an example plot of the Mahalanobis distance in which a contour plot overlays the scatterplot of 100 random draws from a bivariate normal distribution with a mean of zero, unit variance, and 50% correlation. The centroid defined by the marginal means is noted by a blue square. It should be noted that the Mahalanobis distance is capable of detecting outliers in the bivariate space.

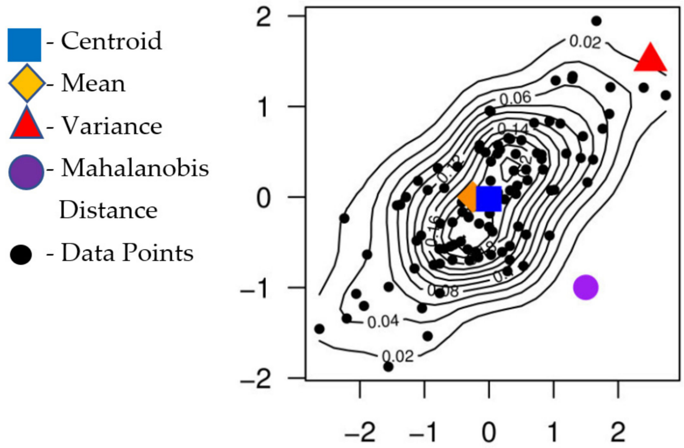

Figure 1. Illustration of the Mahalanobis distance in a bivariate space adapted from [9]; the contours represent the behavior of the MD in the bivariate space. The Mahalanobis distance can detect outliers in the bivariate space.

1.3. Related Studies

One study on the statistical evaluation of SCADA data for condition monitoring indicates that there are advantages and limitations of using SCADA in condition monitoring [10]. One such limitation is in its capacity to detect anomalies by itself. SCADA data are limited to the operational values of the wind turbines which can not function for anomaly detection alone. Statistical analyses can be conducted in order to address this limitation, and one tool that can be used is the "Mahalanobis distance".

Based on the characteristics of the Mahalanobis distance, its applicability in fault detection for other applications has already been studied. One study made use of the MD approach in detecting incipient sensor faults. In this study, the model used the MD approach to compare with a conventional approach. The results showed that the MD approach's performance was comparable with the conventional one [11]. It was emphasized in this paper that the MD approach is a data-driven monitoring method. Another study that used the MD approach in detecting anomalies involved detecting multivariate outliers in a data set. Multivariate outliers can cause disruptions in the data's behavior, resulting in erroneous results. This led to the application of the MD approach to detect these outliers in the data set. The study results showed that the MD approach was able to detect multivariate outliers, even though the accuracy was relatively low compared to other methods [12].

A study in [13] conducted an incipient fault diagnosis using the MD where the detection was based on empirical probability estimation. The results from the simulation showed that the proposed methodology was effective for non-Gaussian data, and at the same time sensitive for incipient fault detection. The case study presented also highlights the benefits of the proposed methods compared to state-of-the-art-based solutions. The MD approach was also used in methane detection using spectroscopy. The study aimed to incorporate the MD approach in gas detection and calculate the corresponding effects of errors in the detection method. Their results showed that the error was minimized for the methods that incorporated the MD approach [14]. A better error detection performance was achieved when the MD approach was incorporated into the anomaly detection method.

In terms of anomaly detection in wind turbines, several methods using data analysis were developed. The paper in [15] developed an anomaly detection method based on SCADA data mining. This method extensively used the data in the SCADA to create correlations between abnormalities and performances of wind turbines. The results showed a promising use of SCADA data in detecting abnormalities in wind turbines.

Two studies in [16,17] proposed new methods that can be used for wind turbine diagnosis using SCADA data. Parameter features from the SCADA data were used in [16], while vibration measurements in the tower were used in [17]. Both of these methods were seen to be promising in developing future wind conditions monitoring systems. Another paper [18] showed how SCADA data for a gearbox was used in creating a new anomaly detection technique. The paper used machine learning algorithms in the creation of a model that was used for anomaly detection. Compared to other techniques, this technique was able to reach a comparable efficiency.

Using SCADA data in anomaly identification was emphasized in another article [19]. In this study, an improvement in the deep belief network was used as the basis for anomaly detection. Results showed considerable efficiency in terms of accurately detecting abnormalities in wind turbines. These papers used SCADA data and analyzed them in a variety of ways to create an abnormal detection method.

Mahalanobis distance control charts were used in determining anomalous behaviors in [20], which were compared with other methods, particularly the multivariate exponentially weighted moving average. It was found that using MD control charts were more effective in identifying outliers within the data set, and could moreover be used for detecting anomalies in data-extensive applications. Another study in [21] showed how the MD can be used in identifying diagnoses in non-linear systems. Since non-linear systems are more prevalent in real-world scenarios, it was advised that the MD be used to obtain more reliable diagnostics of systems. A different study conducted also assessed the structural health of offshore wind turbines, using autoencoders and the MD distance [22]. The combination of the MD and autoencoders resulted in a more robust unsupervised novelty detection pipeline for structural health monitoring.

The proposed method in this study will focus on the use of operational data of the SCADA system paired with the MD approach in order to investigate abnormalities in the operation of wind turbines. This is an improved version of the related literature cited above, as this method will be able to relate the MD approach to the actual alarm event. This could be beneficial to the future of conditions monitoring for wind turbines.

2. Materials and Methods

The general data of the wind turbine from the wind farm used in the study are presented in Table 1. It should be noted that in using the SCADA system of the wind turbine, knowledge of the different characteristics of the wind turbine is essential. Knowing the different characteristics of the wind turbines used in the study can help ensure accuracy in the conclusions generated. For this study, a direct-drive onshore wind turbine located in Taiwan was used.

Table 1. General data of the wind turbine and SCADA system from the wind farm used in this study.

Parameter	Data
Location	Taiwan
Number of wind turbine heads	Fourteen (14)
Nominal power capacity of wind turbine	2.0 MW
Transmission connection	Direct-drive
Number of parameters in SCADA system	Thirty-two (32)
Number of component models generated	Seven (7)

A flowchart presented in Figure 2 is used to accomplish the study's objectives. It begins with the extraction of operational and alarm data from the SCADA system. These data are then pre-processed to filter abnormal data underlying different periods. The pre-processed data are then used to generate predictive models using the algorithm presented in [23].

Figure 2. Flowchart used in the study.

Table 2 shows the specifications of the models that use neural networks. Based on the SCADA parameters used, seven output component models were generated. Using the different models generated from the predictive model algorithm, the MD was calculated between predicted and actual values. In the generation of the predictive models and the calculation of the Mahalanobis distance, the computing programming software MATLAB was used.

Table 2. Specification of the models that use neural networks.

Artificial Neural Network Type	**Feed Forward Neural Network**	
Layer	Hidden	Output
Neurons	20	1
Activation function	Sigmoid	Threshold
List of possible inputs	Air gap temperature, generator voltage, generator current, inverter current, inverter voltage, cooling water low pressure, cooling water low pressure, cooling water temperature, control cabinet temperature, impeller angle, motor current, blade angle, rotor temperature, wind speed, wind direction, vibration	
List of possible outputs	Generator speed, generator stator temperature, alternator bearing temperature, cooling water high pressure, rotor speed, brake hydraulic pressure, active power	

The MD for the ten-minute duration points for anomaly detection was averaged for the first twelve hours of the day and for the following twelve-hour points. The method employed provides two averaged data points for each day. These averaged data points are then plotted for the entire year or duration of the study to show abnormalities based on the threshold value. The threshold value for the study was calculated based on the typical characteristics of the data points [24]. The healthy data of the SCADA system was used in calculating the threshold value using the same MD approach. The averaged MD value was used as the threshold value.

In order to backtrack the estimated date of occurrence of the anomaly, reverse engineering was used. The point or time of the year where the anomaly occurred was divided by two, and this value represents the date in the year of the occurrence. The time of the year was divided by two, since the averaging of the MD distance was done twice a day. Calculating this would result in the estimated date of occurrence of the anomalies.

Anomalies detected with the MD approach were then compared with the SCADA alarm data for the duration of the study. The component of concern was also part of the comparison with the SCADA alarms. The different input parameters were also considered in comparison with the SCADA alarms. In other words, the input parameters were checked to see if they coincided with the SCADA alarms. Table 3 shows the list of these input parameters for each output model used in the study. The different parameters were combined with their correlations to the output parameters [25]. Using MATLAB software, the different input and output parameters were used for the generation of the specified model. This ensured that these input parameters were related to the output parameters in terms of data behavior.

Table 3. List of input parameters for each output model.

Output Model	Input Parameters
Generator speed	average alternator bearing temperature generator voltage generator current active power
Stator temperature	average generator air gap temperature average alternator bearing temperature generator current active power
Bearing temperature	average generator stator temperature generator speed average generator air gap temperature generator voltage
Cooling water high pressure	average inverter current average inverter voltage cooling water low pressure active power
Rotor speed	blade 2 motor current blade 3 motor current average blade angle active power
Brake hydraulic pressure	XY direction resultant force vibration X direction vibration Y direction vibration active power
Active power	average generator stator temperature generator current average inverter current rotor speed

The procedure for checking the anomaly was to determine the days covered by certain deviations from the threshold. The days covered by this anomaly were then backtracked to trace the month and day of the occurrence. The alarms in the SCADA system were then checked using the dates of occurrence of the deviations.

The threshold for the data set was calculated by using the data set for normal behaviors. In this way, the MD calculated represented the threshold of such a data distribution. Paper [26] illustrated this type of method in determining the threshold of such data distribution; the healthy data values are used in both model generation and calculating the MD in this instance. The averaged MD value represents the threshold of the data distribution.

3. Results and Discussion

The Mahalanobis distance for each model and wind turbine is plotted for the entire duration. Figure 3 shows the results of these plots for the brake hydraulic pressure model for all the wind turbines in the wind farm. The plot shows some abnormal behavior for some wind turbines. An example would be wind turbine H11, where abnormal behavior can be seen at the end of the year. The different output models show different durations or periods of occurrence of abnormal behavior, and these are subject to comparison with the SCADA alarm data.

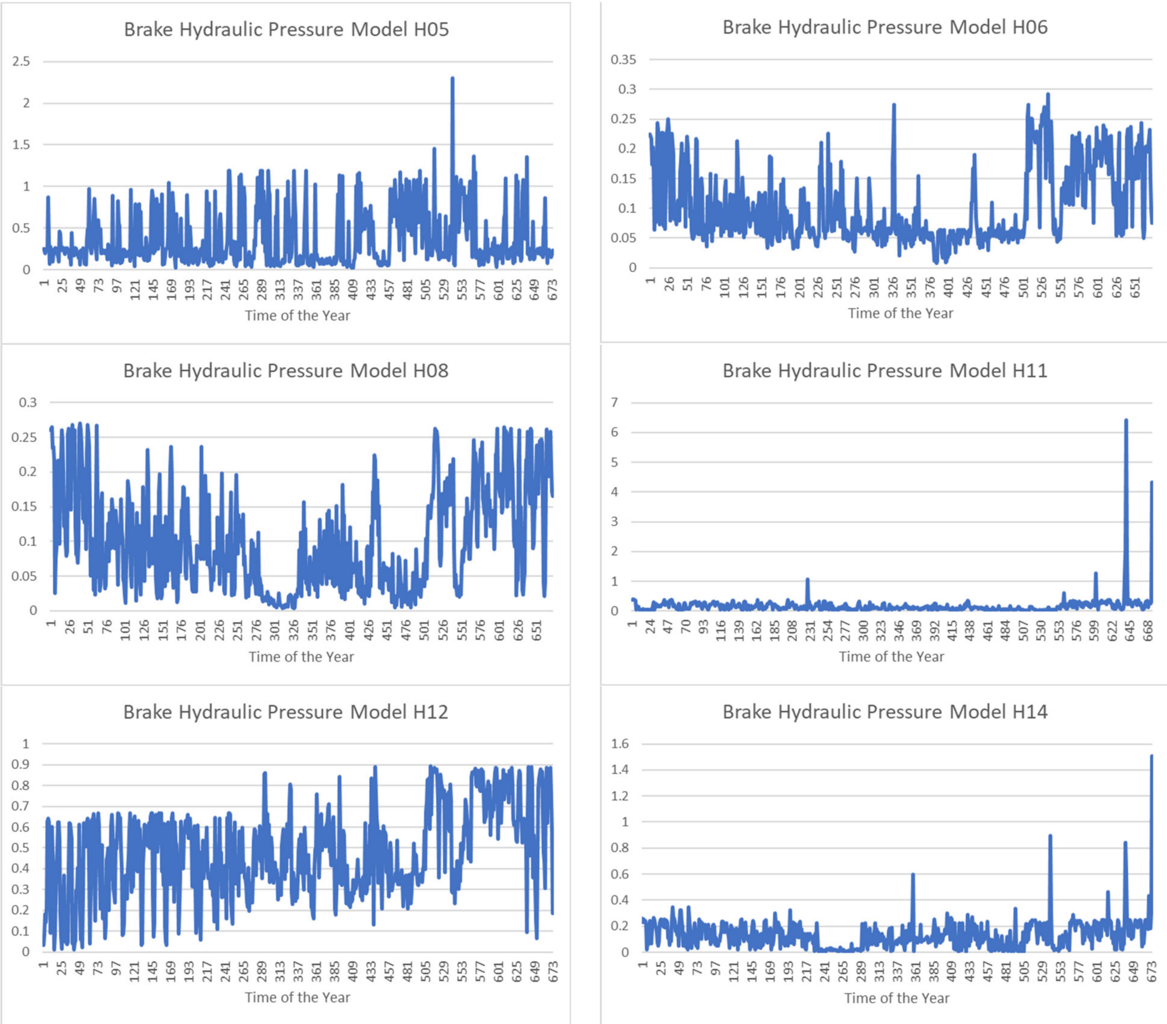

Figure 3. Plotted Mahalanobis distances based on the brake hydraulic pressure model of some wind turbines in the wind farm.

The results show that abnormal behavior can be detected in the calculated MD for the data set. This shows the ability of the MD to detect faults in a component where data is extensively used, and where data characterizes the system's performance [27]. The expected behavior of the data set shows slight variations, while portions with faulty behavior show widespread deviations with the rest of the data set.

Using the plotted results in Figure 3 can help assess the wind turbine behavior for different parameter data sets. Observing the plot can help one infer the different wind turbines where clear abnormal behavior in the data set was generated, based on the MD measure. Thus, this plot can be useful for generating observations that can lead to the detection of anomalies based on the MD measure.

An example would be a plot based on the brake hydraulic pressure model where the MD measures for the different wind turbines are shown. It can be noted that different behaviors can be observed for the wind turbines in the wind farm. This observation can also be used in implying the efficiency of the generated models in going deep into the predictive models. However, this is not covered in this study.

As shown in the plot for the MD measure of different wind turbines for the brake hydraulic pressure model, certain observations can be generated for durations where there is abnormal behavior. A good observation example is the end part of the year for wind turbine H11, where there is a clear abnormal behavior in the data set. This will then be tested using the threshold value generated from the normal behaviors of the data set. This shows that having a summary of plotted MD measures for different wind turbines of different predictive models can be useful in making observations that may lead further to anomaly detection.

The brake hydraulic pressure model for wind turbine H11 shows a period where abnormal behavior is detected. Figure 4 shows this period emphasized inside a dashed red box within the plot. The thick threshold line shows within the plot which data are in typical or faulty behavior.

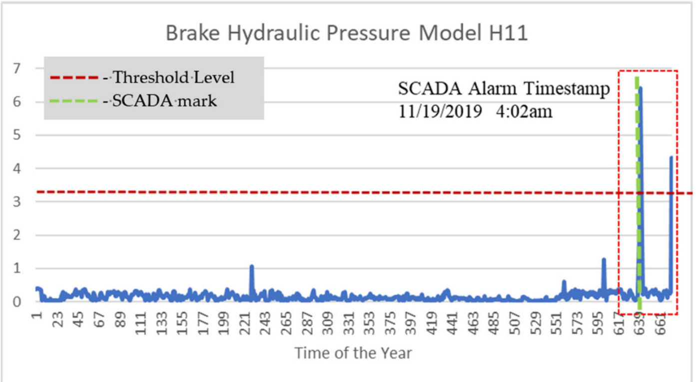

Figure 4. Mahalanobis distance based on the brake hydraulic pressure model for H11.

The period of occurrence of the faulty behavior is calculated by tracing the period corresponding to the day of the year of the occurrence. It should be noted that the day of the year presented in the plot should be divided into two, since the averaged MD is calculated twice a day. Table 4 further looks into the MD value and time-stamp around the SCADA alarm region. It can be observed that the highest MD value of 6.42 occurred at time-point 645, which means a time period between 19 November 2019 00:00 and 19 November 2019 12:00.

Table 4. Detailed MD value and timestamp list around the SCADA alarm region for H11.

Time-Points of the Year	Real Time-Stamp	MD Value
641	2019/11/17 00:00	0.077828383
642	2019/11/17 12:00	0.200583207
643	2019/11/18 00:00	1.202948234
644	2019/11/18 12:00	2.032181356
645	**2019/11/19 00:00**	**6.421187127**
646	2019/11/19 12:00	1.909682528
647	2019/11/20 00:00	0.221972715
648	2019/11/20 12:00	0.273416611

The different SCADA alarm data are then checked for correspondence with the component where the MD faulty behavior is detected. Relevant data in Table 5 show that alarm data for November for wind turbine H11 indicate a very low pressure on the yaw brake system. This corresponds to the detected anomaly in the calculated MD of the brake hydraulic pressure model. In terms of the component analysis of the different parameters, the MD approach detected the faulty behavior of the wind turbine.

Table 5. Part of the SCADA alarm system corresponding to anomalies for H11.

Wind Turbine	Alarm Code	Alarm Description	Date/Time
H11	228	Yaw Brake Pressure Very Low	19 November 2019 4:02
H11	228	Yaw Brake Pressure Very Low	19 November 2019 17:15

Another example is based on the stator temperature model plots for the calculated MD. Wind turbine H07 shows faulty behavior among the plots in the wind farm for stator temperature. Figure 5 shows the plot for the calculated MD for the stator temperature model for H07. Tracing the period of occurrence for this abnormal duration results in October.

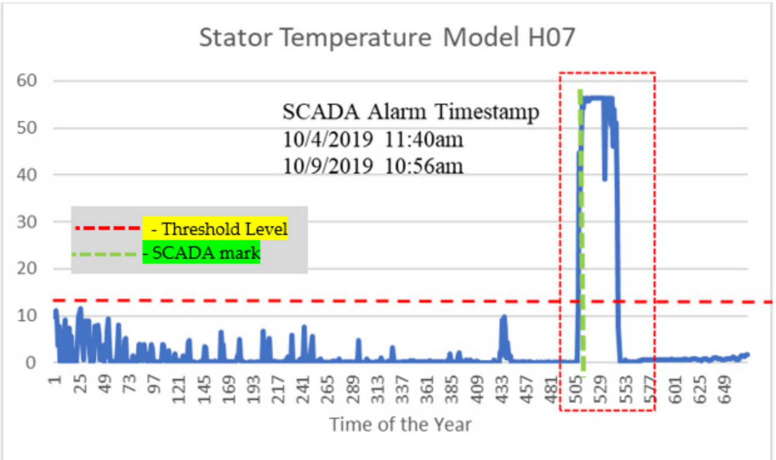

Figure 5. Mahalanobis distance based on the stator temperature model for H07.

The SCADA alarm data for wind turbine H07 shown in Table 6 shows a very high temperature in the generator stator and bearing temperature fault for October. It should be noted that bearing temperature is part of the input parameters for the stator temperature models. In terms of parameters in different components, the MD approach was able to detect the faulty behavior of the wind turbine.

Table 6. Part of the SCADA alarm system corresponding to anomalies for H07.

Wind Turbine	Alarm Code	Alarm Description	Date/Time
H07	249	generator stator temperature very high	9 October 2019 10:56
H07	242	bearing temperature fault	4 October 2019 11:40

In summary, calculating the percentage of the detected faulty behaviors based on the SCADA alarm data shows that over 97 percent of the faulty behaviors were detected using the MD approach based on different models of the wind turbines in the wind farm. This accuracy is comparable to the studies presented in [28,29], as shown in Table 7. These papers used anomaly detection methods involving data analysis and machine learning methods. They were able to generate efficiencies of more than 95%, which is comparable to the efficiency generated in this study.

Table 7. Efficiency comparison of proposed method to other methods.

Method	Percentage Accuracy
Deep Small-World Neural Network [28]	95%
Statistical Process Control and Machine Learning [29]	92.16%
SCADA Alarm Data and MD approach (proposed method)	97%

Results of the plotted MD for each wind turbine model show that abnormal behaviors can be detected based on the threshold of the data set. The characteristics of the Mahalanobis distance can be used to detect faulty behaviors in the wind turbine and can be integrated into the condition monitoring system. This could imply using the MD approach in a condition monitoring system that uses data extensively.

Comparison with the SCADA alarm data shows that parameters used in the generation of the predictive models affect how the MD approach detects the faulty behaviors of the wind turbine. Different components of the predictive models could detect abnormalities corresponding to the wind turbine component concerned. This is important in the construction of the predictive models to be used in the condition monitoring system. The results also show that the MD approach was able to detect faulty behaviors at a high percentage. This is important in considering the MD, a data-intensive approach for fault detection in wind turbines. A higher percentage of fault detection using a data-driven approach could be sustainable in the long run [30].

Although the MD measure can be effective in detecting abnormalities in wind turbines, certain defects and challenges can still occur while using this method. One sophisticated challenge and defect is integrating the MD measure into the available CMS of wind turbines. Having a data-intensive method may create problems in maintaining the performance of the system incorporated into the wind turbine. Another possible defect of the method is its reliance on powerful computing software, which may affect the performance of the CMS in the wind turbines. Although this defect can be addressed with improvements in technology, further studies need to be developed to cope with this situation.

There are also uncertainties in fault detection using the proposed method. Aside from SCADA network problem where data can be erratic or not at all available, missing SCADA data will affect calculations of times of fault occurrences in the MD approach. This can result in uncertainty in the method if not addressed properly. An uncertainty model presented in [31] is applied in the proposed approach. This model is presented in Equation (3), as follows:

$$B = \frac{X_{true}}{X_{pred}} \tag{3}$$

where B is the model uncertainty, X_{true} is the true value (SCADA alarm value) and X_{pred} is the predicted value (MD distance). The calculated model uncertainty was averaged for instances of failures detected, and the results show a final uncertainty of 0.965.

4. Conclusions

Based on the results of this study, the following conclusions can be generalized:

1. The Mahalanobis distance approach can detect faulty behaviors of wind turbines at a high accuracy rate of 97%.
2. The input and output parameters of the predictive models where the Mahalanobis distance approach is calculated are essential factors in determining the component affected by faulty behavior.
3. In a data-driven monitoring system, the Mahalanobis distance approach can be used to further sustain and enhance the features of such a system.

In order to further improve this study, future researchers can address the following aspects of the study:

1. Integrate the proposed approach into the CMS of the wind turbine, and investigate its efficiency in the system.
2. Improve the reliability of the Mahalanobis distance approach by creating variations in the duration averaging of the approach, and choose the optimum in the process.
3. Integrate the Mahalanobis distance approach and the condition monitoring system into an automatic system environment to facilitate the recording of failure reports and further enhance productivity.
4. Develop a predictive model for the MD method that can detect anomalies at early stages.

Author Contributions: Conceptualization, J.-H.L. and N.T.C.J.; data curation, J.-H.L.; formal analysis, J.-H.L., R.-M.L. and N.T.C.J.; investigation, J.-H.L. and C.-C.W.; methodology, J.-H.L. and N.T.C.J.; project administration, J.-H.L.; validation, N.T.C.J. and C.-C.W.; writing—original draft, N.T.C.J.; writing—review and editing, R.-M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Economic Affairs grant number M353CH1330.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: The support from the Industrial Technology Research Institute and Taipower Company were highly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

d_{MD}	Mahalanobis distance
x_a, x_b	a pair of objects for the MD calculation
C	the sample covariance matrix
$cov_{x,y}$	covariance of distribution
x_i, y_i	data values of the distributions
$\overline{x}$	mean value of x distributions
B	model uncertainty
CMS	condition monitoring system
MD	Mahalanobis distance
SCADA	Supervisory Control and Data Acquisition system
Hxx (H01~H14)	represents the turbine number in the harbor area

References

1. Zhang, Y.; Zheng, H.; Liu, J.; Zhao, J.; Sun, P. An anomaly identification model for wind turbine state parameters. *J. Clean. Prod.* **2018**, *195*, 1214–1227. [CrossRef]
2. Chen, L.; Xu, G.; Zhang, Q.; Zhang, X. Learning deep representation of imbalanced SCADA data for fault detection of wind turbines. *Measurement* **2019**, *139*, 370–379. [CrossRef]
3. Pang, Y.; Jia, L.; Zhang, X.; Liu, Z.; Li, D. Design and implementation of automatic fault diagnosis system for wind turbine. *Comput. Electr. Eng.* **2020**, *87*, 106754. [CrossRef]
4. Fang, R.; Wu, M.; Guo, X.; Shang, R.; Shao, P. Identifying early defects of wind turbine based on SCADA data and dynamical network marker. *Renew. Energy* **2020**, *154*, 625–635.
5. Xiang, L.; Wang, P.; Yang, X.; Hu, A.; Su, H. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement* **2021**, *175*, 109094. [CrossRef]
6. Du, Y.; Zhou, S.; Jing, X.; Peng, Y.; Wu, H.; Kwok, N. Damage detection techniques for wind turbine blades: A review. *Mech. Syst. Signal Process.* **2019**, *141*, 106445. [CrossRef]
7. Christensen, R. *Advanced Linear Modeling: Statistical Learning and Dependent Data*; Springer: Cham, Switzerland, 2019.
8. Brereton, R.G.; Lloyd, G.R. Re-evaluating the role of the Mahalanobis distance measure. *J. Chemom.* **2016**, *30*, 134–143. [CrossRef]
9. Hill, T.; Lewicki, P.; Lewicki, P. *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining*; StatSoft Inc.: Street Tulsa, OK, USA, 2006.
10. Gonzalez, E.; Tautz-Weinert, J.; Melero, J.J.; Watson, S.J. Statistical Evaluation of SCADA data for Wind Turbine Condition Monitoring and Farm Assessment. *J. Phys. Conf. Ser.* **2018**, *1037*, 032038. [CrossRef]
11. Ji, H. Statistics Mahalanobis distance for incipient sensor fault detection and diagnosis. *Chem. Eng. Sci.* **2020**, *230*, 116233. [CrossRef]
12. Leys, C.; Klein, O.; Dominicy, Y. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *J. Exp. Soc. Psychol.* **2017**, *74*, 150–156. [CrossRef]
13. Yang, J.; Delpha, C. An incipient fault diagnosis methodology using local Mahalanobis distance: Detection process based on empirical probability density estimation. *Signal Process.* **2021**, *190*, 108308. [CrossRef]
14. Chen, P.; Kan, L.; Song, X.; Wang, X.; Jiang, C. Application of VMD and Mahalanobis distance combination algorithm in TDLAS methane gas detection. *Optik* **2020**, *228*, 166114. [CrossRef]
15. Liu, X.; Lu, S.; Ren, Y.; Wu, Z. Wind Turbine Anomaly Detection Based on SCADA Data Mining. *Electronics* **2020**, *9*, 751. [CrossRef]

16. Zhang, C.; Hu, D.; Yang, T. Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108445. [CrossRef]
17. Castellani, F.; Garibaldi, L.; Daga, A.P.; Astolfi, D.; Natili, F. Diagnosis of Faulty Wind Turbine Bearings Using Tower Vibration Measurements. *Energies* **2020**, *13*, 1474. [CrossRef]
18. McKinnon, C.; Carroll, J.; McDonald, A.; Koukoura, S.; Infield, D.; Soraghan, C. Comparison of new anomaly detection technique for wind turbine condition monitoring using gearbox SCADA data. *Energies* **2020**, *13*, 5152. [CrossRef]
19. Long, X.; Li, S.; Wu, X.; Jin, Z. Wind Turbine Anomaly Identification Based on Improved Deep Belief Network with SCADA Data. *Math. Probl. Eng.* **2021**, *2021*, 8810045. [CrossRef]
20. Rashid, K. Comparative Between the MEWMA and Mahalanobis Distance Control Chart. *Kurd J. Appl. Res.* **2021**, *6*, 94–104.
21. Daroogheh, N.; Meskin, N.; Khorasani, K. An improved particle filtering-based approach for health prediction and prognosis of nonlinear systems. *J. Frankl. Inst.* **2018**, *355*, 3753–3794. [CrossRef]
22. Weil, M.; Weijtjens, W.; Devriendt, C. Autoencoder and Mahalanobis distance for novelty detection in structural health monitoring data of an offshore wind turbine. *J. Phys. Conf. Ser.* **2022**, *2265*, 032076. [CrossRef]
23. Liu, J.H.; Corbita, J.N.T. Performance analysis of different predictive models for condition monitoring of direct drive wind turbine generator. *Meas. Control* **2021**, *54*, 374–384. [CrossRef]
24. Zaher, A.S.A.E.; McArthur, S.D.J.; Infield, D.G.; Patel, Y. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy Int. J. Prog. Appl. Wind Power Convers. Technol.* **2009**, *12*, 574–593. [CrossRef]
25. Bette, H.M.; Jungblut, E.; Guhr, T. Non-stationarity in correlation matrices for wind turbine SCADA-data and implications for failure detection. *Wind Energy Sci. Discuss.* **2021**, in press. [CrossRef]
26. Wang, B.; Gu, C. Comparison of the projection onto convex sets and iterative hard thresholding methods for seismic data interpolation and denoising. *Explor. Geophys.* **2018**, *49*, 825–832. [CrossRef]
27. Lahdhiri, H.; Taouali, O.; Elaissi, I.; Jaffel, I.; Harakat, M.F.; Messaoud, H. A new fault detection index based on Mahalanobis distance and kernel method. *Int. J. Adv. Manuf. Technol.* **2017**, *91*, 2799–2809. [CrossRef]
28. Li, M.; Wang, S.; Fang, S.; Zhao, J. Anomaly Detection of Wind Turbines Based on Deep Small-World Neural Network. *Appl. Sci.* **2020**, *10*, 1243. [CrossRef]
29. Hsu, J.-Y.; Wang, Y.-F.; Lin, K.-C.; Chen, M.-Y.; Hsu, J.H.-Y. Wind Turbine Fault Diagnosis and Predictive Maintenance Through Statistical Process Control and Machine Learning. *IEEE Access* **2020**, *8*, 23427–23439. [CrossRef]
30. Ouyang, T.; He, Y.; Huang, H. Monitoring wind turbines' unhealthy status: A data-driven approach. *IEEE Trans. Emerg. Top. Comput. Intell.* **2018**, *3*, 163–172. [CrossRef]
31. Bai, Y.; Jin, W.-L. Random Variables and Uncertainty Analysis. *Mar. Struct. Des.* **2016**, *12*, 615–625.