3-10-2019

Classification of Wind Turbine Blade Performance State Through Statistical Methods

Jones Shen
University of Windsor

Classification of Wind Turbine Blade Performance State Through Statistical Methods

By

Jones Shen

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Mechanical, Automotive and Materials Engineering

in Partial Fulfilment of the Requirements for

the Degree of Master of Applied Science at the

University of Windsor

Windsor, Ontario, Canada

Classification of Wind Turbine Blade Performance State Through Statistical Methods

By

Jones Shen

APPROVED BY:

T. Bolisetti
Department of Civil and Environmental Engineering

B-A. Schuelke-Leech
Department of Mechanical, Automotive & Materials Engineering

R. Carriveau, Co-Advisor
Department of Civil and Environmental Engineering

D. S-K. Ting, Co-Advisor
Department of Mechanical, Automotive & Materials Engineering

January 18, 2019

DECLARATION OF CO-AUTHORSHIP/PREVIOUS PUBLICATIONS

I hereby declare that this thesis incorporates material that is the result of joint research, as follows:

Thesis Chapter	Details
Chapter 2	This thesis also incorporates the outcome of a joint research undertaken in collaboration with Ryan Francis under the supervision of Dr. Lindsay Miller, Dr. Rupp Carriveau, Dr. David S-K. Ting, Dr. Marianne Rodgers, and Mr. J.J. Davis. In all cases, the key ideas, primary contributions, data analysis and interpretation, were performed by the author, and the contribution of the co-author and Ms. Carina Xue Luo (Leddy Library, University of Windsor) was primarily through the provision of data processing.
Chapter 3	This thesis also incorporates the outcome of a joint research undertaken in collaboration with Milad Rezamand under the supervision of Dr. Rupp Carriveau, Dr. David S-K. Ting, Mr. J.J. Davis, and Dr. Marianne Rodgers. In all cases, the key ideas, primary contributions, data analysis and interpretation, were performed by the author, and the contribution of the co-author was primarily through the provision of proposing the experimental methodology.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

This thesis includes an original paper that has been previously submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication	Publication Status
Chapter 2	Jones Shen, Ryan Francis, Lindsay Miller, Rupp Carriveau, David S.K. Ting, Marianne Rodgers, and J.J. Davis, “GIS Visualization of Wind Farm Operational Data to Inform Maintenance and Planning Discussions,” Wind Engineering	Under Review

Thesis Chapter	Publication	Publication Status
Chapter 3	Jones Shen, Milad Rezamand, Rupp Carriveau, David S.K. Ting, J.J. Davis, and Marianne Rodgers, "Improved Condition Monitoring of Wind Turbine Blades through Unsupervised Machine Learning," Wind Engineering	Submitted
Chapter 4	Jones Shen, Rupp Carriveau, David S.K. Ting, "Condition Monitoring of Wind Turbine Blades through k-means Clustering," Journal of Solar Energy Engineering	Submitted

I certify that I have obtained a written permission from the copyright owner(s) to include the above submitted material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material iv that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

As wind turbines continue to age, wind farm operators face the challenge of optimizing maintenance scheduling to reduce the associated operation and maintenance (O&M) costs. Wind farm operators typically use conservative maintenance scheduling in order to maximize the uptime of their wind turbines. In most cases however, maintenance may not be necessary and the components could operate for longer before repairs are required. This work presents three papers that collectively focus on providing potentially useful information to aid wind farm operators in making maintenance decisions. In the first paper, the utilization of Geographic Information Systems (GIS) to illustrate data trends across wind farms is introduced to better understand an operation's signature performance characteristics. It is followed by a paper that presents an improved condition monitoring system for the wind turbine blades through the use of the principal component analysis (PCA). The final paper introduces another condition monitoring system using a k-means clustering algorithm to determine the performance state of wind turbine blades.

DEDICATION

To my friends and family.

ACKNOWLEDGEMENTS

I would like to acknowledge the guidance, wisdom, and patience provided by my co-advisors, Dr. Rupp Carriveau, and Dr. David S-K Ting. The support they've provided these past two years have been essential in my personal development. I have truly enjoyed working with them in the Turbulence & Energy Laboratory, and I have found them to be accommodating and knowledgeable whenever I've sought their insight. I also wish to extend thanks to the remaining members of my committee, Dr. Tirupati Bolisetti and Dr. Beth-Anne Schuelke-Leech, for offering their expertise on my research. I would like to acknowledge the contributions of Mr. Ryan Francis, Ms. Carina Luo, Dr. Lindsay Miller who was of great help with the GIS process discussed in Chapter 2. Additionally, I would like to thank Mr. Milad Rezamand for his contributions to Chapter 3. This research would not be possible without the generous support from our industrial partners, and I wish to thank Mr. J.J. Davis, Dr. Marianne Rodgers, Mr. John Bridges for their efforts.

This work is part of the YR21 Investment Decision Support Program supported by progressive industrial partners, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Ontario Centres of Excellence (OCE).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

CHAPTER 2: GIS Visualization of Wind Farm Operational Data to Inform Maintenance and Planning Discussions

CHAPTER 3: Improved Condition Monitoring of Wind Turbine Blades through Unsupervised Machine Learning

CHAPTER 1

Introduction

1. Background

The global prospects for wind energy technology are very promising. The wind turbine market is rapidly advancing in terms of technology, increased power production, and decreased costs [1]. The annual Wind Technologies Market Report released for 2017 by the U.S. Department of Energy (DOE) highlighted that wind energy will continue to be a competitively priced source of energy and one of the most economic means of electricity generation [1]. Beyond economics, wind energy operation has practically zero carbon and water footprint. That said, the major challenge of reducing operation and maintenance (O&M) costs is ongoing. O&M costs for onshore and offshore wind turbines account for ~15% and ~35% of their total costs respectively [2], [3].

Due to the continuous aging of wind turbines, wind farm operators face the challenge of optimizing maintenance scheduling to reduce operating costs. Wind farm operators typically use conservative maintenance scheduling to maximize the uptime of their wind turbines. In some cases, however, maintenance may not be necessary, and the components could operate for longer before repairs are required. Equipping components within the turbine with specialized and invasive sensors that directly report on structural health can be prohibitively expensive for wind farms. Thus, the use of abundant and readily available supervisory control and data acquisition (SCADA) data is appealing to increase the understanding of wind turbines and their components throughout their operation. A more thorough understanding of the change in behaviour of the turbine components can aid wind farm operators in making informed maintenance decisions.

Therefore, the research objectives of this thesis are as follows:

- To create a simplified method for wind farm operational data analysis for improved understanding of turbine behaviour to aid in maintenance and planning discussions
- To understand the changes in wind turbine blade condition and behaviour throughout its operation by highlighting a method to infer when maintenance is required, and
- To interpret the changes in wind turbine blade condition and behaviour during operation and to determine when maintenance is required by using an alternative method that addresses weaknesses seen in objective 2.

A survey of available literature shows endless results for methods for analyzing wind farm operations data to characterize specific component health and conditioning. However, a number of these studies have been performed using more elaborate techniques ranging from neural networks to Bayesian estimation and other fuzzy methods to predict turbine component conditioning, performance, and their remaining useful life (RUL) [4–9]. These techniques tend to require specialized training or programs that are not easily implemented by owners and operators of wind turbines. For these reasons, a simplified process that can also be complementary to pre-existing methods may be valuable to the industry.

Although complicated, the aforementioned methods have proven to be successful in determining component behaviours to diagnose their operational health. Specifically, condition monitoring system methods typically use sensors for oil monitoring and vibration analysis to determine component health [10–12]. Their success is likely in no small part due to the value of the data provided by direct sensors. However, most wind farm blade populations lack the level of detailed and direct structural performance data provided by the types of specialized sensors. Wind turbine blades have been illustrated by Ciang et al [13] and Nivedh [14] that blade failures were among the most common components to fail within the entire turbine assembly. While these analysis methods, in general, have gained significant attention, publications that have applied these methods to wind turbine blades successfully are very rare [9].

These publications show the extent in which general condition monitoring systems based on elaborate neural network and fuzzy have been applied to. As the wind turbines continue to advance, higher reliability targets are emphasized which would require reduced O&M costs to remain sustainable. Higher reliability with lower operational costs will result in increases of incentives for investment in clean wind energy. This thesis attempts to address the issue of the operational state of wind turbine blades through the utilization of readily available SCADA data. It studies the application of a simplified geographic information (GIS) visualization method, an unsupervised learning method, specifically principal component analysis (PCA), and a k-means clustering algorithm to improve on condition monitoring by identifying the blade performance state to determine whether maintenance is necessary. While the PCA and k-means methods have been applied in many wind turbine components such as generators and gearbox subcomponents, such techniques have yet to be applied to turbine blades.

2. Methodology

In order to create a condition monitoring system for turbine blades, its main functions should first be established. Figure 1 shows a breakdown for determining the performance state of wind turbine blades using a typical condition monitor. The process is broken down into 2 main sections: conditioning prediction and blade failure modes. The conditioning prediction consists of forecasting specific faults before they occur for the blades as well as predicting its remaining useful life. This section requires very specific failure data which was not available at the time of this study within the selected wind farms so these processes will not be covered within this thesis. The blade failure modes cover the causes and effects for the deterioration of failure of the blades. The causes can include damaged blades or varying wind and environmental characteristics within the wind farms. Chapter 2 focuses on the examination of varying wind characteristics when compared with performance metrics within wind farms using a simplified ranking and GIS method to determine what can be inferred (i.e. are the components performing poorly due to high wear or from wake effects and other varying meteorological effects?) with this easily obtainable information. Chapters 3 and 4 explores the effects of failing blades which include reduced power production and rotor speeds as well as changes in the observed blade pitch angles and yaw. Using statistical methods such as PCA and k-means, the state of the blades can be inferred to determine the necessity for repairs where changes in observed performance variables can be detected visually.

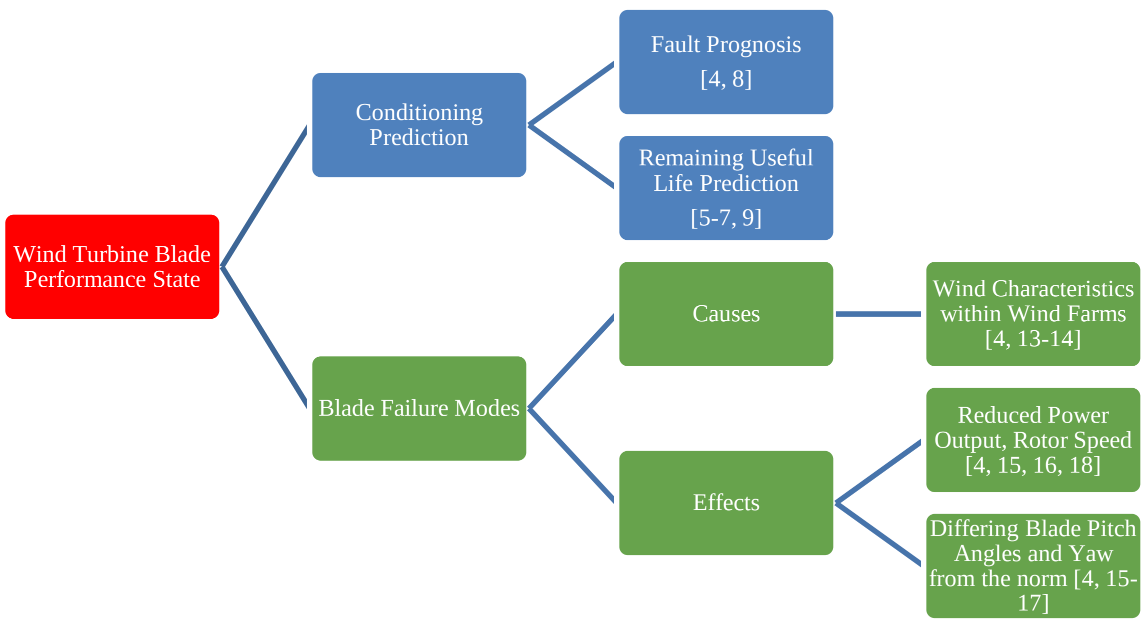


Figure 1. 1: A condition monitoring system's breakdown to determine the performance state of blades

To be able to develop an improved condition monitoring system for turbine blades, it is important to first understand its performance and behaviour under changing operating conditions. This is first introduced in Chapter 2 where the utilization of simple methods to analyze basic wind farm operations data are applied to provide high level insights into the usage history of critical wind turbine components. With the introduction of the simplified ranking method and GIS visualizations, results and findings from the application to real-world wind farms are presented. The process relied on 10-minute averaged SCADA data from a 3-year period acquired from three wind farms. Wind farm 1 and 2 are located in Southwestern Ontario and farm 3 is located near the northern coast of Prince Edward Island. The ranking method for the wind turbines based on usage history can help infer blade and other component wear alongside other pre-existing techniques. The utility of GIS visualizations was shown to be capable of providing additional long range, high level insight into the loading history of specific wind turbine components. This information can increase the level of information available to an operator when considering potential maintenance decisions.

Chapter 3 focused on the implementation of a SCADA based unsupervised machine learning method, particularly the principal component analysis (PCA) method for improved wind turbine blade condition monitoring. With the use of PCA, large measurement datasets of various SCADA tags can be reduced to new variables that act as linear combinations of select parent variables. The reduced set of variables (the principal components) can improve the ability to visualize the data. This technique is then utilized to infer the performance state of the blades to improve the lead time available to wind farm operators to make informed maintenance decisions.

Building on the research shown in Chapter 3, Chapter 4 sought to provide similar blade performance state results and inferences using a modified k-means clustering method. The integration of k-means allows for the clustering of similar data points together using their Euclidean distance between each point to identify whether the blades are operating normally or abnormally on a given day. The daily blade statuses were plotted over the time series of the data set to simulate a condition monitoring system. The results from the monitoring system were compared to the norm in order to define any deviations as performance abnormalities to determine whether blade maintenance was necessary.

The analysis performed in this thesis sets a foundation for development of future wind turbine blade condition monitoring systems and further research on improving their accuracy. The research presented herein can be improved by performing a sensitivity analysis to examine the influence of different parent variables on the derived principal components and the k-means clustering results. That is, which input variables among those available from the SCADA will be most descriptive of the performance state of the wind turbine blades? As the wind turbine technology continue to evolves, the demand for improved reliability will also aggressively be sought after. The analysis prescribed in this thesis serves to act as a stepping stone for future energy sustainability.

References

[1] R. Wiser and M. Bolinger, “2017 Wind Technologies Market Report,” Lawrence Berkeley National Laboratory, 2017.

[2] Y. Sinha, J. A. Steel, J. A. Andrawus, and K. Gibson, “A SMART Software Package for Maintenance Optimisation of Offshore Wind Turbines,” *Wind Eng.*, vol. 37, no. 6, pp. 569–577, 2013.

[3] Y. Sinha and J. A. Steel, “Failure Prognostic Schemes and Database Design of a Software Tool for Efficient Management of Wind Turbine Maintenance,” *Wind Eng.*, vol. 39, no. 4, pp. 453–478, 2015.

[4] C. S. Gray, F. Langmayr, N. Haselgruber, and S. J. Watson, “A Practical Approach to the Use of SCADA Data for Optimized Wind Turbine Condition Based Maintenance,” *EWEA Offshore Wind Amst.*, p. 10, 2011.

[5] Y. Zhao, D. Li, A. Dong, D. Kang, Q. Lv, and L. Shang, “Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data,” *2016 North Am. Power Symp. NAPS*, pp. 6–22, 2016.

[6] Y. Zhao, D. Li, A. Dong, J. Lin, D. Kang, and L. Shang, “Fault Prognosis of Wind Turbine Generator Using SCADA Data,” Tongji University, 2016.

[7] M. A. Djeziri, S. Benmoussa, and R. Sanchez, “Hybrid method for remaining useful life prediction in wind turbine systems,” *Renew. Energy*, pp. 1–15, 2017.

[8] D. Astolfi, F. Castellani, and L. Terzi, “Mathematical methods for SCADA data mining of onshore wind farms: Performance evaluation and wake analysis,” *Wind Eng.*, vol. 40, no. 1, pp. 69–85, 2016.

[9] J. S. Nielsen and J. D. Sørensen, “Bayesian Estimation of Remaining Useful Life for Wind Turbine Blades,” *energies*, vol. 10, no. 12, pp. 664–677, 2017.

[10] B. Morton, “Understanding Wind Turbine Condition Monitoring Systems.” Renewable NRG Systems, 2018.

[11] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, “Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges,” *energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[12] H. Luo, C. Hatch, M. Kalb, J. Hanna, A. Weiss, and S. Sheng, “Effective and accurate approaches for wind turbine gearbox condition monitoring,” *Wind Energy*, vol. 17, no. 5, pp. 715–728, 2013.

[13] C. C. Ciang, J.-R. Lee, and H.-J. Bang, “Structural health monitoring for a wind turbine system: a review of damage detection methods,” *Meas. Sci. Technol.*, vol. 19, no. 12, p. 20.

[14] B. S. Nivedh, “Major Failures in the Wind Turbine Components and the Importance of Periodic Inspections,” *Wind Insid.*, vol. 2014, p. 5, 2014.

[15] A. Sareen, C. A. Sapre, and M. S. Selig, “Effects of leading edge erosion on wind turbine blade performance,” Wind Energy, vol. 17, no. 10, pp. 1531–1542, 2014.

[16] N. Gaudern, “A practical study of the aerodynamic impact of wind turbine blade leading edge erosion,” J. Phys. Conf. Ser., vol. 524, p. 012031, 2014.

[17] C.-J. Bai and Y.-C. Shiah, “Power-Curve Corrections for Horizontal-Axis Wind Turbine by an Improved Blade Element Momentum Theory,” J. Mech., vol. 33, no. 03, pp. 341–349, 2016.

[18] P. W. Richards, D. T. Griffth, and D. H. Hodges, “Smart Loads Management for Damaged Offshore Wind Turbine Blades,” Wind Eng., vol. 39, no. 4, pp. 419–436, 2015.

CHAPTER 2

GIS Visualization of Wind Farm Operational Data to Inform Maintenance and Planning Discussions

Jones Shen[1], Ryan Francis[1], Lindsay Miller[1], Rupp Carriveau[1], David S.K. Ting[1], Marianne Rodgers[2], J.J. Davis[3]

[1] Turbulence and Energy Laboratory, University of Windsor, Windsor, Ontario, Canada
[2] Wind Energy Institute of Canada, North Cape, Prince Edward Island, Canada.
[3] Kruger Energy, 3285 Chemin Bedford, Montréal, Quebec, Canada

1. Introduction

The reliability of wind turbine components has been of major recent interest to the renewable energy sector owing to their substantial repair costs and earlier than anticipated failures [1]. While some causes may be traced back to design or manufacturing defects, many failures are caused by operational environments that accelerate wear. Where possible, it is desirable to engage prognostics to help prevent catastrophic failure and the accompanying financial loss to wind farm operations. To help combat component failures and excessive maintenance costs, many wind farms implement condition monitoring (CM) systems that use sensors to determine component health such as oil debris or vibration-based systems to detect faults before they cause further damage [2, 3]. While capable, sophisticated CM systems are not often originally installed in older operations and may be cost prohibitive to retrofit for others. It is here proposed that basic and potentially useful insights into turbine health may be gained through simple analysis of turbine Supervisory Control and Data Acquisition (SCADA) data. These techniques may provide potentially earlier and complementary warnings to that of conventional CM systems that are geared to report after faults have already occurred [3]. A number of previous studies have been performed for various wind turbine components using more elaborate techniques ranging from neural networks to

Bayesian estimation and other fuzzy methods to predict their conditioning, performance, and their remaining useful life (RUL) [4–9]. These techniques tend to require specialized training or programs and are not easily implemented by owners and operators of wind turbines. For these reasons, a simplified process that can also be complementary to pre-existing methods may be valuable to the industry.

Wind farm owners and operators are seeking guidance for how to optimize day to day decisions regarding farm assets. When maintenance or repair situations occur, they are often tasked with making quick decisions with the goal of minimizing downtime to meet generation targets. The optimal policies for making these decisions are very complex as they depend on many factors including the present physical condition of the assets, the replacement technology, insurance policies, social climates and on uncertain future regulatory and market conditions. The YR21 project, led by the Universities of Windsor and Western Ontario, aims to consider all factors to develop an investment decision support system to shepherd owners through the dynamic operating climate ahead. The advantage of this model lies in connecting reliability modelling and farm valuation.

As a first step to towards reliability modelling, the study presented here uses a SCADA data ranking and GIS visualization approach to inform maintenance and planning discussions. Without more sophisticated, resource intensive options, ranking and GIS visualizations can assist operators to highlight machines/components to prioritize and/or focus on when assessing maintenance requirements or administering maintenance budgets. Utilizing the most commonly accessible data tags from farm SCADA systems, we seek the most appropriate SCADA tags to act as a metaphorical odometer for wind turbine components. SCADA data contains valuable information about the performance and load history of the turbines and has been successfully applied as a tool to optimize condition-based maintenance on a wind farm [4] and to detect anomalies in operations [6]. SCADA variables of interest such as rotor speed and power output are incorporated with wind speed to illustrate variations across a wind farm. As farms progress through their operating lifetime and as downtimes and failures accumulate, this study will be combined with major component downtime records to seek to identify usage indicators (UI's); a simplified approach for improving insight into the usage, and potential wear, experienced by individual turbines on a wind farm. Such indicators may provide high level insights or starting points for farm operators tasked with making maintenance decisions.

The GIS visualization approach presented here provides quick reference for the farm and can inform operators which machines have produced the most power, which are subjected to the highest winds, as well as which are located downwind of others or along coastal regions. These parameters can be important in understanding operations and prioritizing maintenance. The visualization approach can also be used to support wake loss research, which is another important challenge surrounding wind developments. Turbines that are downwind of the front row machines can be severely impacted by the wake from those upstream. Losses typically range from 10 – 30% [10–12] and can be as high as 70% [13].

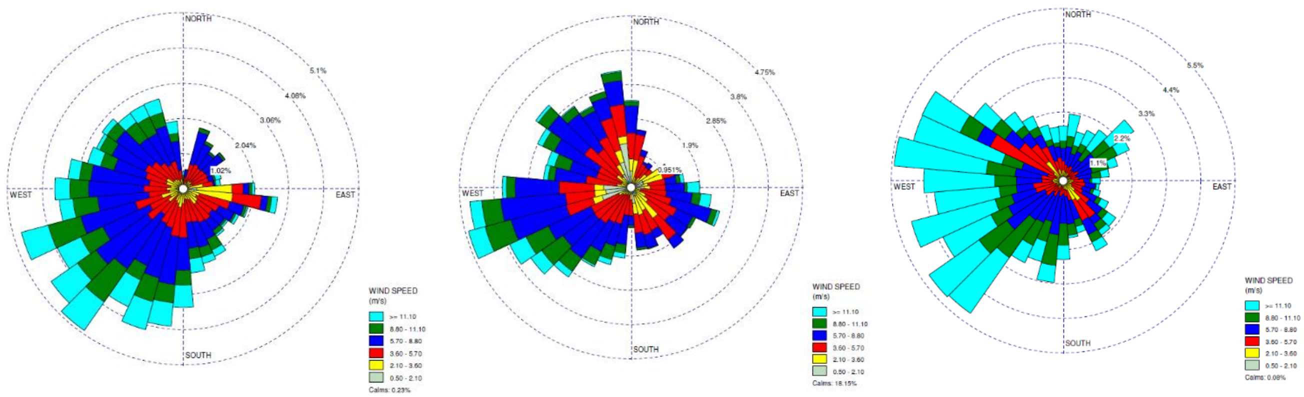
The aim of this work was to investigate simple methods to analyze basic wind farm operations data to provide high level insights into the usage history of critical wind turbine components. There have been many studies using neural networks and other more complex techniques to predict the remaining useful life (RUL) of specific wind turbine components with high accuracy, the approaches here are more blunt; but also more accessible. Specifically, we point to the utility of GIS visualizations in providing additional long range, high level insight into the loading history of specific wind turbine components. The ranking of wind turbines based on usage history can help infer blade and other component wear alongside other pre-existing techniques. This information can increase the level of information available to an operator when considering potential maintenance decisions.

2. Methodology

As a first step to inform operations discussions, the turbines were ranked based on selected SCADA variables. Figure 2.1 provides an overview of the ranking process. The process relies on 10-minute averaged SCADA data from a 3-year period acquired from three wind farms. Wind farm 1 and 2 are located in Southwestern Ontario and farm 3 is located near the northern coast of Prince Edward Island.

Figure 2. 1: Ranking methodology.

The wind farms' respective wind roses are shown in Figure 2.2 and their features are summarized in Table 2.1. The wind roses were produced using hourly wind speeds and wind directional data retrieved from the wind measurement (MET) towers located on each of the wind farms. This data acts as an input in the software WRPLOT to display the relative frequencies of wind speeds and directions within each wind farm. Wind farms located in Southwestern Ontario have dominant winds from the southwest while the winds at the PEI site largely originated from the west.



(a) (b) (c)

Figure 2. 2: Wind rose for (a) wind farm 1, (b) wind farm 2, and (c) wind farm 3 measured by on-site meteorological mast.

Table 2. 1: Wind farm characteristics

Feature	Wind farm 1		Wind farm 2	Wind farm 3
Number of turbines	48	40	43	5
Rotor diameter (m)	93	101	101	93
Hub height (m)	80	80	80	80
Rated power (MW)	2.3	2.3	2.3	2.0
Data set size	8,955,429	7,373,443	8,693,562	593,428

The SCADA data consists of operational tags of interest which include wind speed, rotor speed, pitch control, and power output. These tags were chosen for the study because of their relation to the performance of wind turbines and their potential to act as health indicators [4]. The data is then pre-processed to filter outliers. In this study, certain data points were labeled as outliers. These included any curtailed power outputs as well as instances of no power output when the wind speed is below the rated cut-in speed as these reductions in power for a given a wind speed are not related to the degradation of the wind turbine components [3].

In some regions it is common for system operators to curtail wind energy production at times of low demand. This can be done by adjusting the blade pitch angle, resulting in a reduction of power production for a given wind speed. This reduction in power could be misinterpreted as result of turbine degradation [3, 14], and therefore, it is important that the input SCADA data is filtered during instances where output power is being curtailed.

The processed SCADA variables are then totaled or averaged for each tag from the period of 2013 to 2016. These SCADA values are then compiled into a pool consisting of all wind turbines where each tag is ranked from greatest to least for inspection.

To visualize the more notable trends between SCADA variables, maps were created using inverse distance weighted (IDW) interpolation with the help of Carina Luo from Leddy Library of the University of Windsor. IDW interpolation is a deterministic method for multivariate interpolation that assigns values to unknown points by calculating a weighted average of the values available at known points. With this in mind, the main author provided wind speeds and power outputs measured at each individual wind turbine from the SCADA as it can then be utilized to make predictions for other unmeasured locations within the same wind farm, resulting in geographic information system (GIS) maps with a continuous wind speed prediction surface or contour covering the entire study area. This would allow for the main author to analyze the underlying relationships between the wind speeds and power production throughout the wind farm study area.

3. Results and Discussion

3.1. Ranking and Mapping of SCADA data

The turbines were ranked from highest to lowest based on the selected SCADA variables. For example, the ranking of SCADA variables is visually presented for wind farm 1 in Figure 2.3. Average wind speeds, total power production, and total blade rotations are depicted for selected turbines across the farm to provide a spatially representative sample. Preliminary observation-based evaluation shows that the turbines that are first to face the prevailing winds, in this case, South westerlies, produce higher total power on average than those downwind. For example, turbine 54, 46, 47, 58, and 50 are ranked 1, 2, 5, 4, 8 respectively for total power whereas turbine 44, 42, and 30 are ranked 88, 85, and 80 respectively. This confirms the general findings from wake loss models that have demonstrated the impacts of wakes from upstream turbines on the performance of those downwind [15].

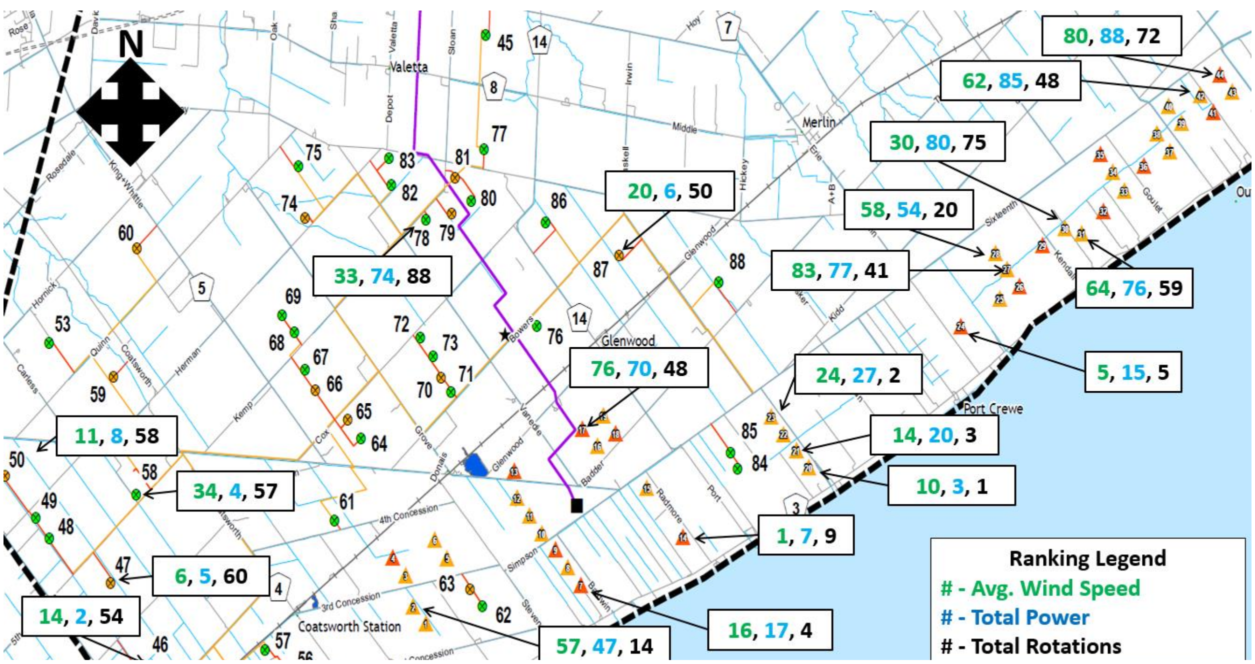


Figure 2. 3: Ranking of wind speed and select SCADA variables (from highest to lowest) for select turbines at farm 1.

Figure 2.4 presents spatially interpolated GIS maps of wind farms 1 to illustrate the corresponding average wind speeds experienced and average power produced.

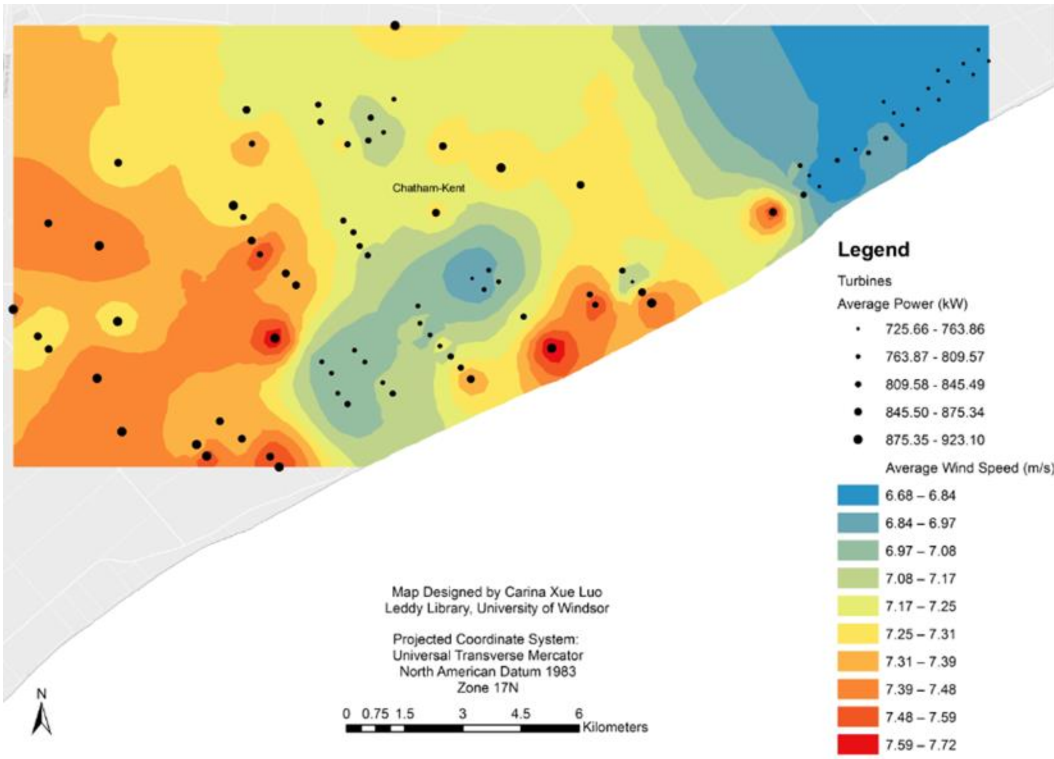


Figure 2. 4: Average wind speed and average power produced for wind farm 1.

The contours represent the average wind speed experienced throughout the farm and the dots represent the location of each individual turbine with the size of the dot corresponding to the average power production. The GIS map illustrates that higher power production is spatially correlated with higher wind speeds. These higher values were focused around the southwestern region of the wind farm which was also in good agreement with the wind rose of the wind farm shown in Figure 2.2a. For average power produced, the highest ranked turbine produces 1.15 times (101m rotor) and 1.12 times (93m rotor) that that of the lowest ranked turbines. These values are consistent with recent literature stating that the turbines downwind typically produce 10 – 20% less than the front-row machines [10, 11].

Figure 2.5 depicts the relationship between the average wind speeds experienced and the average power production on wind farm 2. The map displays similar trends to those observed in wind farm 1, specifically that higher power production is spatially correlated with higher wind speeds. For farm 3, the ratio between the highest and lowest producing turbine is 1.25, higher than that at farm 1.

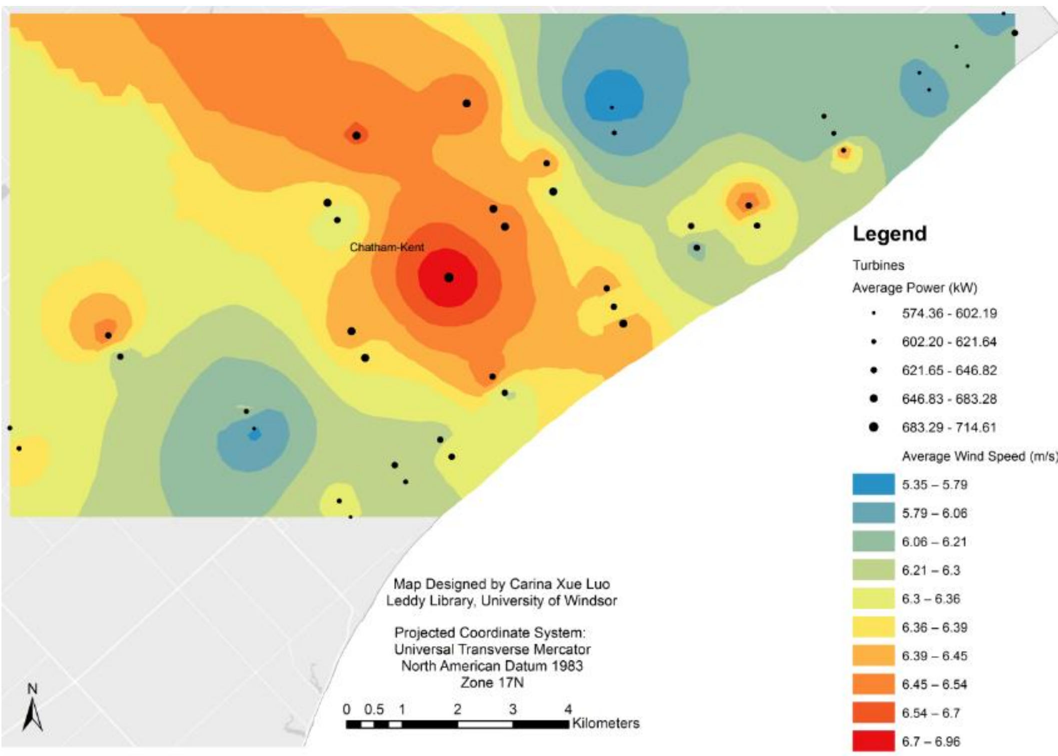


Figure 2. 5: Average wind speed and average power produced for wind farm 2.

The average wind speed and average power produced in wind farm 3 is presented in Figure 2.6. The relation between the average wind speed and the average power produced was not clearly discernable. This is most clearly represented by the wind turbine located in the centre of wind farm 3 as shown in Figure 2.6, where the turbine experiencing the highest average wind speed within the wind farm produced the lowest average power in the wind farm. The reason for this is unknown, and is likely due to a combination of factors, including increased occurrence of faults for this turbine. Although the sample size is small, the ratio of the highest to lowest producing machine is 1.06.

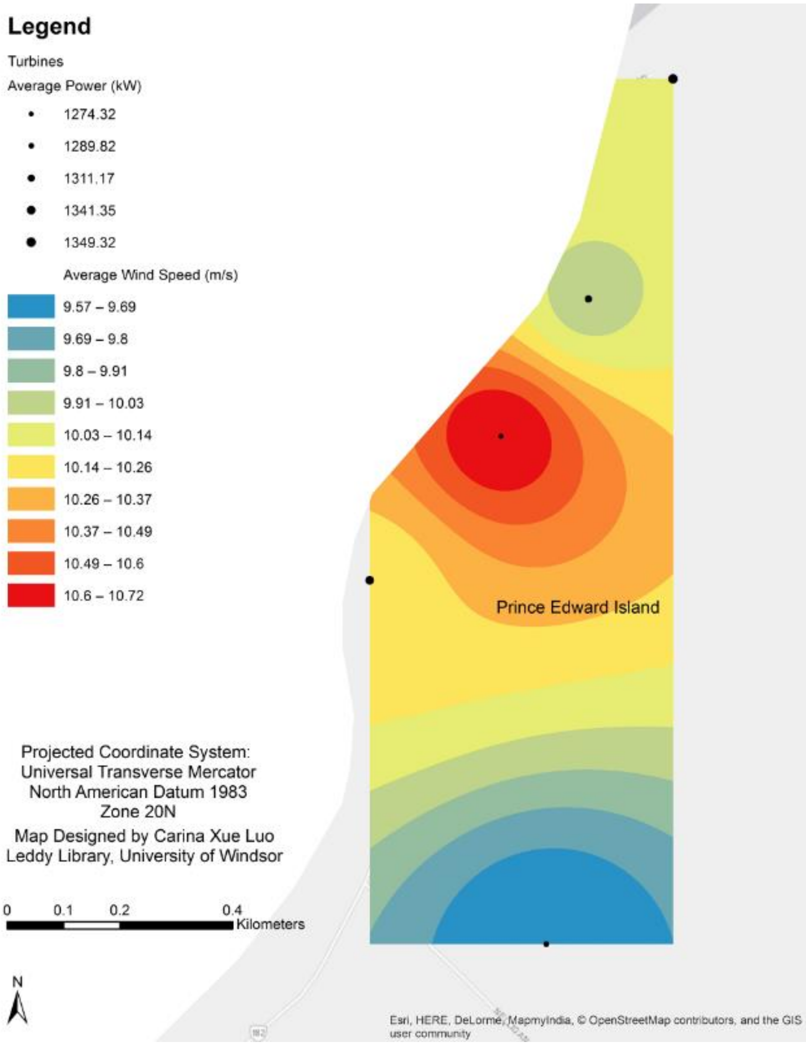


Figure 2. 6: Average wind speed and average power produced for wind farm 3.

Taken together, the GIS maps illustrate confirmation that high winds produce higher power. This is not unexpected, however, the range of speeds and average powers and variation across the farms were not previously established. This analysis also uncovered the importance of consideration of other factors. The simple wind speed vs. power output does not hold up for comparisons when the turbine downtime is considered. This could apply to other differences in turbine characteristics. The maps are also consistent with previous wake loss model research.

3.2. Utility of GIS Visualizations

Informing maintenance discussions and planning

The GIS maps serve to inform maintenance discussions and planning. When presented to the farm operators, many discussions were stimulated. The visual representations provide easily interpretable information to any audience and can assist in understanding turbine and farm behaviours. For example, if a subset of turbines are experiencing a common issue, such as blade deterioration, the maps can point to potential influencers such as exposure to high winds or environmental conditions such as being located close to a body of water. Although this approach cannot conclusively determine the cause, it may serve as a starting point to direct further investigation. For example, at farm 3, two of the turbines experienced major component failures at year 7. The map provided in Figure 2.6 started a conversation around what potential variables contributed to their failure. The turbines that experienced the failures were the highest producing turbines on the farm and experienced the highest average RPM. These alone may not explain the failure, but could potentially inform future maintenance and planning.

As mentioned, wind farm operators can be required to curtail their production during times of low demand. The maps can serve as a quick reference to determine which machines have been producing more or experiencing higher winds and could therefore inform the decision of which machines to curtail. This strategy could extend the useful life of the machines and ultimately, the profitability of the farm.

Lastly, it is common for operators to share experiences with other operators for the purpose of uncovering causes of unexpected downtime or repair situations. The maps can also be used to compare farm characteristics and to identify turbines that are exposed to similar conditions between wind farms.

Future site planning

The maps can also be used for future site planning. Although wake models are available, actual demonstrated cases of wake impacts and turbine layout can validate or improve models or potentially uncover additional variables for consideration. This could ultimately improve wind farm layout optimization. The GIS visualizations can also inform extension or repower planning at an existing wind farm. Familiarity with the machines, their layout, and operating characteritics can assist operators with selecting which machines might be suitable for retrofits or repowering and which may be best to decommission.

4. Conclusion and areas for future research

A simplified yet versatile ranking method for inferring usage and providing high level insight into the loading history of wind turbines has been presented. The ranking method can improve wind farm operators' understanding by inferring usage which is valuable as it can increase the amount of information available when considering potential maintenance decisions. Future work should focus on evaluating the correlations between the SCADA variables and major component repairs or downtime. At this point in the YR21 study, the participating farms did not have substantial downtime hours or repairs, and therefore, the statistical power was too low to provide clear correlation analysis. If correlations between SCADA variable rankings and undesirable outcomes, such as blade downtime, can be uncovered, a usage indicator could be developed to provide a simple, yet easily accessible tool to infer wear and inform O&M discussions.

The GIS visualizations also confirmed the results of wake loss models. The row of turbines first to face predominant winds consistently produced more power (between 12 – 25%) than those that were downstream of these machines. This effect was not well observed at farm 3 due to the small number of turbines on the farm although a 6% variation was still observed between the highest and lowest producing turbines. These findings support the importance of optimizing turbine layout to achieve maximum production for the farm.

The next step to this research work is to collaborate with the wind farms to produce detailed and specific blade downtime hour logs to support the development of a usage indicator (UI). Future works will also include the incorporation of other SCADA parameters relating to the health of major components. Of course, a potential UI cannot be applied to serial defects and may not be broadly application across farms (i.e. the SCADA variables that correlate with downtime at one farm, may not correlate at another). Environmental operating conditions will also influence the relationship between usage and wear and will need to be considered for broader applicability. These limitations will be carefully considered in the development of an informative and robust UI.

GIS visualizations of wind speed and average power were produced for the three farms. The maps present a simple compilation of information for comparison including impact of spacing, rotor size, coatings, wind speed, power production, layout, and proximity to water bodies. Although some of the relationships are intuitive, for example, higher winds produce higher average power, the demonstration of variation across the farms is a unique contribution.

Future work will also use the GIS maps to focus on the impacts of land use features including neighbouring wind farms, trees, and other nearby uses such as golf courses. The impact of lake breezes, which could possibly enhance fluctuations in wind speed and direction as well as convection in the region, leading to accelerate wear on components, should also be considered. Previous research has demonstrated wind turbine wear and damage caused by sudden changes in wind speeds and direction experienced by coastal turbines [16]. All three farms are exposed to water bodies which could present an opportunity to study this effect further.

The SCADA data ranking and GIS visualizations, when presented to the farm operators, resulted in increased knowledge and familiarity of their turbines and the operating and environmental characteristics of their farm. In the absence of elaborate and costly specialized systems, these tools can provide insight into maintenance and planning discussions that could ultimately improve operations and drive profitability.

Acknowledgement

This work was supported in part by Carina (Xue) Luo of the Leddy Library, University of Windsor. The authors would also like to acknowledge that this work is part of the YR21 Investment Decision Support Program supported by progressive industrial partners, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Ontario Centres of Excellence.

References

[1] R. Wiser and M. Bolinger, “2016 Wind Technologies Market Report,” U.S. Department of Energy (DOE), 2016.

[2] B. Morton, “Understanding Wind Turbine Condition Monitoring Systems.” Renewable NRG Systems, 2018.

[3] E. Lapira, D. Brisset, H. D. Ardakani, D. Siegel, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renew. Energy*, vol. 45, pp. 86–95, 2012.

[4] C. S. Gray, F. Langmayr, N. Haselgruber, and S. J. Watson, “A Practical Approach to the Use of SCADA Data for Optimized Wind Turbine Condition Based Maintenance,” *EWEA Offshore Wind Amst.*, p. 10, 2011.

[5] Y. Zhao, D. Li, A. Dong, D. Kang, Q. Lv, and L. Shang, “Fault Prediction and Diagnosis ofWind Turbine Generators Using SCADA Data,” *2016 North Am. Power Symp. NAPS*, pp. 6–22, 2016.

[6] Y. Zhao, D. Li, A. Dong, J. Lin, D. Kang, and L. Shang, “Fault Prognosis of Wind Turbine Generator Using SCADA Data,” Tongji University, 2016.

[7] M. A. Djeziri, S. Benmoussa, and R. Sanchez, “Hybrid method for remaining useful life prediction in wind turbine systems,” *Renew. Energy*, pp. 1–15, 2017.

[8] D. Astolfi, F. Castellani, and L. Terzi, “Mathematical methods for SCADA data mining of onshore wind farms: Performance evaluation and wake analysis,” *Wind Eng.*, vol. 40, no. 1, pp. 69–85, 2016.

[9] J. S. Nielsen and J. D. Sørensen, “Bayesian Estimation of Remaining Useful Life for Wind Turbine Blades,” *energies*, vol. 10, no. 12, pp. 664–677, 2017.

[10] P. Sørensen and T. Nielsen, “Recalibrating wind turbine wake model parameters – validating the wake model performance for large offshore wind farms,” *Eur. Wind Energy Assoc. EWEA Conf. Exhib.*, 2006.

[11] R. J. Barthelmie *et al.*, “Modelling and measuring flow and wind turbine wakes in large wind farms offshore,” *Wind Energy*, vol. 12, no. 5, pp. 431–444, 2009.

[12] P. McKay, R. Carriveau, and D. Ting, “Wake impacts on downstream wind turbine performance and yaw alignment,” *Wind Energy*, vol. 16, no. 2, pp. 221–234, 2012.

[13] J.-Å. Dahlberg, “Assessment of the Lillgrund Windfarm,” *Vattenfall*, 2009. [Online]. Available: https://corporate.vattenfall.se/globalassets/ sverige/om-vattenfall/om-oss/var-verksamhet/vindkraft/lillgrund/assessment. pdf.

[14] D. McLaughlin, P. Clive, and J. McKenzie, “Staying Ahead of the Wind Power Curve,” *Renew. Energy World*, vol. 13, no. 2, 2010.

[15] C. L. Archer *et al.*, “Review and evaluation of wake loss models for wind energy applications,” *Appl. Energy*, vol. 226, pp. 1187–1207, 2018.

[16] M. Shafiee, M. Finkelstein, and C. Bérenguer, “An opportunistic condition-based maintenance policy for offshore wind turbine blades subjected to degradation and environmental shocks,” *Reliab. Eng. Syst. Saf.*, vol. 142, pp. 463–471, 2015.

CHAPTER 3

Improved Condition Monitoring of Wind Turbine Blades through Unsupervised Machine Learning

Jones Shen[1], Milad Rezamand[1], Rupp Carriveau[1], David S.K. Ting[1], J.J. Davis[2], Marianne Rodgers[3]

[1] Turbulence and Energy Laboratory, University of Windsor, Windsor, Ontario, Canada
[2] Kruger Energy, 3285 Chemin Bedford, Montréal, QC, Canada
[3] Wind Energy Institute of Canada, 21741 PE-12, Tignish, PE, Canada

1. Introduction

With the wind turbine technology rapidly evolving year after year [1], wind farm operators face the challenge of upkeeping their machines in order for them to compete against newer models in terms of performance. However, by being very conservative with maintenance to maximize their turbine uptimes, there may be instances where the machines may not need repairs. With unnecessary maintenance, it needlessly increases operation and maintenance (O&M) costs [2], [3]. The biggest culprit for the O&M costs tend to be blades where they have the greatest repair costs within the turbine (up to 15-20% of the total turbine cost) [4]. To reduce this cost, the modes of blade failures as well as methods in determining the need for repairs are examined to create a possible solution.

Ciang et al [4] and Nivedh [5] showed that wind turbine blade failures were among the most common components to fail within the entire turbine assembly. Many wind farm operators are observing blade deterioration sooner than original manufacturer specifications. It can be difficult for manufacturers to anticipate all environmental challenges a blade will operate under, including meteorological conditions, pollution, varying precipitations and saline air. These conditions can lead to accelerated wear and leading edge erosion where surface defects, cracks, and structural discontinuities may arise, reducing life expectancy [6]. It was observed by Sareen et al [7], [8] and Gaudern [9] that eroded blades can cause severe lift reductions and drag increases. Sareen et al's results showed a 6-500% drag increase and a loss in annual energy production of up to 25% depending on the level of erosion present on the blade. Beyond this, there are catastrophic failure mechanisms in blades which have led to the collapse of entire machines [4]. With the concept of a CM system introduced in Chapter 2, a form of predictive maintenance can be deployed within wind farms to help improve wind turbine blade reliability and reduce O&M costs. These approaches can take advantage of the lead times provided by the predictions of faults being made in early stages of component health degradation [3].

This study introduces the use of a statistical approach for the condition monitoring of wind turbine blades. CM systems typically use sensors for oil monitoring and vibration analysis to determine component health [10]–[12]. Most often these components are drivetrain elements. Fortunately, the science of diagnosing the state health of bearings and gears is quite advanced. Its success is likely in no small part due to the value of the data provided by these direct sensors. Luo et al [12] and Siegel et al [13] were able to utilize vibrational data from subcomponents of their test cases to develop effective condition monitoring algorithms. However, most wind farm blade populations lack the level of detailed and direct structural performance data provided by the types of specialized sensors. Therefore, vibrational data for blades that may have a more potential in diagnosing its state is mostly not available. Subsequently, farms are looking for ways to leverage abundant and readily available supervisory control and data acquisition (SCADA) data. While SCADA data is collected at nearly all farms, in many cases it is not being used to its full potential.

Gray and Watson [14] proposed the use of SCADA performance data combined with an applied knowledge of failure physics to determine the theoretical damage accumulation and therefore, the risk of failure. With the application of the physics of failure (PoF) approach, a detailed assessment of the wind turbine system can be made by identifying potential failures modes, the acting loads and modelling of the damage kinetics in order to highlight critical operating conditions. The PoF is capable of performing rapid calculations due to its simplicity and is also able to account for variability of load capacity, this methodology opens channels for real time damage calculations. Bangalore et al [15] introduces an anomaly detection method using an artificial neural network (ANN) based condition monitoring system. Their work improves the data preprocessing and post-processing from previous publications to improve the confidence in the SCADA-based CM system process. Adaptive neuro-fuzzy interference system (ANFIS) models were employed to determine if specific wind turbine components' condition can be considered healthy or unhealthy in the work of Schlechtingen and Santos [16]. The ANFIS model utilizes a combination of concepts from ANNs and fuzzy logic that enables anomaly detection. It also works towards determining a root cause for such anomalies based on simple If-Then statements. An unsupervised machine learning method to detect anomalies within generator performance metrics was applied successfully in China to perform generator fault prediction, remaining useful life (RUL) estimation and fault type diagnosis using SCADA data [17], [18]. By forecasting an anomaly operation index (AOI) using an ARIMA model, the proportion of unhealthy instances to all instances over a given period was determined which highlighted the probability of performance degradation over time. Methods based on principal component analysis (PCA) have also proven to be quite capable in developing wind turbine fault detection strategies and differentiating between healthy and unhealthy component conditions [19]–[21]. With the use of PCA, large measurement datasets of various SCADA tags can be reduced to new variables that act as linear combinations of select parent variables. The reduced set of variables (the principal components) can improve the ability to visualize the data. This is often achieved through the use of clustering methods that group similarly behaving points and scatters others. Most of these applications focused on components with available SCADA variables that could be associated with their performance (i.e., generator temperature, bearing vibrations, failure histories, etc.).

Following the studies cited in [19]–[21], a SCADA data based, PCA technique for wind turbine blade condition monitoring is developed. This technique is utilized to infer the performance state of the blades as there is no such data point or value that can quantify the damage or the state of the blades. Multiple instances of abnormalities or outliers in the data points will allow for the classification that the blades are in need of repairs. This can improve the lead time available to wind farm operators to make informed maintenance decisions.

2. Methodology

2.1 Data Collection

SCADA data was collected by the main author for a period of 12 months from an 88 turbine farm in Southwestern Ontario, Canada as shown in Table 3.1. Table 3.2 shows the 10-minute averaged SCADA data variables of interest provided by each of the wind turbines.

Table 3. 1: Wind farm characteristics.

Feature	**Wind farm**	
Number of turbines	48	40
Rotor diameter (m)	93	101
Hub height (m)	80	80
Rated power (MW)	2.3	2.3
Data set size	2,522,880	2,102,400

Table 3. 2: SCADA data variables of interest.

Number	Sensor Type	Units
1	Active Power	kW
2	Wind Speed	m/s
3	Rotor Speed	RPM
4	Yaw	deg
5	Blade A Pitch Angle	deg
6	Blade B Pitch Angle	deg
7	Blade C Pitch Angle	deg

2.2 Variable Selection

Successful representative dimensionality reduction is influenced by the strength of the relationship between the principal components and the original parent variables. Subsequently, the selection of appropriate parent variables is critical to the PCA process. The SCADA data variables chosen for this study were active power output, wind speed, rotor speed, yaw, and blade pitch. The turbine active power, wind speed, and rotor speed were chosen as changes in their SCADA values can be observed due to the decaying geometry of the blades caused by leading edge erosion or other types of structural discontinuities [8], [9], [22]. The yaw and blade pitch angles also experience changes as their optimal angles for maximum power output will differ due to the eroding blade geometries [8], [9], [22], [23]. The use of derating to limit loads on damaged blades will also alter the rotor speed and blade pitch at the cost of reduced power production [24]. With these variables, they may serve as features that may reflect the blade's performance health state which will allow for the classification of their operating conditions.

2.3 Data Pre-Processing

Preprocessing required that the data be filtered for errors and curtailment values. SCADA records are often impacted by periodic instrument error. Further, within the region of this study, it is common for wind farm operators to curtail wind energy production when the demand is low. Curtailment can be performed by adjusting the blade pitch angles which results in a decrease of electrical power production for a given wind speed. [25], [26] highlights that this loss of power output could be misunderstood as being a result of blade performance abnormalities or degradation, which therefore, confirms the importance of filtering these instances from the data.

The 10-minute averaged SCADA data was then averaged across a day to form a daily average. Thus, each data point corresponds to a daily average of each SCADA variable for each wind turbine.

2.4 Principal Component Analysis (PCA) Model

PCA is popular when dimensionality reduction and feature extraction is required [27], [28]. It can extract a small number of latent variables, namely the scores $T\ (t_1, t_2, \ldots, t_P)$. These are positioned to best explain the greatest variance in the initial input data matrix X with dimensions of $(N\ \times P)$. Here, N is the number of samples and P is the number of variables or features. The transformed data is linearly independent and are referred to as the principal components which represents a combination of all the features within the input dataset [21]. Before implementing the PCA, it is vital to first normalize and standardize the dataset. This is done by subtracting the mean value of each feature from the input data while scaling each dimension so that they are in the same range and equally weighted.

The principal components are typically obtained by computing eigenvectors U through the use of Singular Value Decomposition (SVD) of the covariance matrix $S\left(S = \frac{1}{m}\sum_{i=1}^{n}\left(x^{(i)}\right)\left(x^{(i)}\right)^{T}\right)$. One must first calculate the eigenvalues λ $(\lambda_1, \lambda_2, \ldots, \lambda_P)$ of the covariance matrix S which can be solved using the equation,

$$|S - \lambda I| = 0 \tag{1}$$

where the eigenvalues λ represent the variances of each principal component and the sum of the variances equals the total sum of the variance of the original variables. With the eigenvalues determined, the following equation can be used to calculate the corresponding eigenvectors $U = \{u_i\}$:

$$U'SU = \lambda \tag{2}$$

The eigenvectors U represent the loadings which indicate the influence (i.e., load or weight) of each of the original variables in forming the principal components, T, as well as explaining how the variables are correlated with the principal components. To calculate the principal components T using the original dataset X $(N \times P)$ and eigenvectors U, the following relationship can be used:

$$T = XU \tag{3}$$

Since the PCA has the effect of concentrating most of the original dataset within the first few principal components, it is ideal to eliminate later components as they may be dominated by noise [19].

In this application, our interest is in capturing the health degradation of turbine blades by finding the optimal combinations (i.e., PCs) of the SCADA variables of interest. The goal of these components is to summarize as much of the initial information as possible with a smaller variable set for ease of interpretation. With the assumption made that the majority of the blades function mostly in a normal state and that faults rarely occur, the largest and heavily dense cluster may represent blades operating in a normal state [17]. Therefore, instances deviating away and located outside of this dense cluster may represent blades operating in an abnormal or "unhealthy" condition which is deemed as a blade failure for this study. The inferences of the blade performance states are hindcasted with historical blade maintenance reports to determine its accuracy. Abnormality in the blade performance state can be observed from the gradual accumulation of leading edge erosion, lightning strike damage, molding surfaces due to trapped precipitation, icing, bird strikes, etc. Boundaries for the normal and abnormal clusters are assigned to create a consistent method of differentiating between the two states. According to a similar method applied to generators [17], blade state changes from normal to abnormal is considered a progressive process rather than instantaneous. Thus, a singular instance of an abnormality may not represent degradation as anomalous environmental operating conditions can force components outside their normal performance specification. Repetitive instances of these abnormal states for a given set of wind turbine blades could however be declared as operation in an abnormal state. However, being classified as abnormal may not necessarily be a blade failure, but rather a deviation from the design performance characteristics so farm operators can then choose to investigate further into the identified abnormality.

The selected SCADA variables have shown to be interdependent with one another where they are inherently correlated with each other. This is not a surprise as the performance of various assemblies within the turbine are at least somewhat dependent on one another. For this study, each of the selected variables are vital as they go through observable changes when the blades deteriorate from the norm [8], [9], [22], [23], [24]. The wind speeds will likely be the only variable to remain constant, but it can be used as a reference point to compare normal and abnormal instances (i.e. expected outputs for a given wind speed might not be produced when blades are operating abnormally). Performing the PCA will create the most compact representation that can reconstruct these observable changes. However, there may be merit in removing some variables thought to be measuring the same latent aspects as the group of variables. The potential redundancy can cause the PCA to overemphasize the variables' contributions. Because of this, no linear regression was used to identify the state of the blades since an important condition for regression is that independent variables should be used. The state of the blades were therefore, not predicted, but rather inferred by statistically clustering blade performance instances.

Typically, when a given wind speed is measured, there is, for example, an expected power output by the turbine. However, as the state of the blades deteriorate, a substantial decrease in power output can be observed given the same wind speed for normal operating blades. If the ranking and GIS method in Chapter 2 was applied to this data, it can highlight the lower power produced by that specific turbine compared to other machines in similar operating conditions without being able to pinpoint whether the cause was due to damage or environmental conditions. With the PCA method, it can be inferred as blade damage if consistent abnormal data points are observed whereas vary environmental anomalies may appear as singular abnormal instances. The method's steps for the model are summarized as follows:

Step 1: Cluster all instances within the top two principal components.

Step 2: Designate boundaries to differentiate normal and abnormal instances.

Step 3: Based on frequency of abnormal instances, classify blade state.

Algorithm 1 presents the detailed procedure for the inference of the wind turbine blade state:

Algorithm 1

---

1. **Setting a Training set**: $x^{(1)}, x^{(2)}, \dots, x^{(m)}$

2. **Standardizing** (Normal Standard Distribution):

   - Calculating mean: $\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)}$

   - Calculating standard deviation: $\sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} - \mu_j)^2}$

   - Replacing each $x_j^{(i)}$ with $\frac{x_j^{(i)}-\mu_j}{\sigma_j}$

3. **PCA** (reducing data from n-dimensions to k-dimensions):

   - Computing covariance matrix: $S = \frac{1}{m}\sum_{i=1}^{n}(x^{(i)})(x^{(i)})^T$

   - Computing eigenvectors of covariance matrix by using SVD (Singular Value Decomposition) algorithm: $[U, S, V] = svd(S)$

   - Employing Alternating Least Squares (ALS) algorithm to find the best rank-k approximation

   PCA algorithm returns the principal component coefficients for the N by P data matrix *X*. Rows of *X* correspond to observations and columns to variables. Each column of COEFF contains coefficients for one principal component. The columns are in descending order in terms of component variance (LATENT).

4. **Assigning criteria for an alarm signal to determine if blades are normal/abnormal**

---

3. Results and Discussion

In this section, the proposed PCA blade performance state inference method is applied and evaluated.

Figure 3.1 represents the cumulative variance explained for each additional principal component added to the study. With two principal components, approximately 85% of the initial input data is explained. Figure 3.2 plots the first and second principal components against one another. Each point is representative of one day over one year, per turbine. In this case, this amounts to 365 days * 88 turbines or 32,120 points. Figure 3.2 appears to have two distinct regions in the plot: one region being the large and dense cluster of points, and the other being points deviated or located outside of this cluster. Given the previous assumptions, the densely packed cluster could be representative of instances where blades are operating in a normal state. The majority of data points reside in this cluster. Conversely, the points outside of the cluster could be representative of operational states outside of the majority norm, here deemed as abnormal.

Figure 3.3 plots the same principal components with only the first and last instances of the dataset shown. Each data point corresponds to a daily average of the principal components for each wind turbine (i.e., Wind Turbine 6, Day 365), but only Day 1 and Day 365 for each turbine are displayed in this plot. The motivation for taking the daily average was to avoid 10-minute averaged outliers from the dense cluster that may cause confusion in the condition of the blades. In Figure 3.3, it is shown that the majority of the wind turbine blades began and ended their operations in the normal cluster. This may imply that some wind turbine blades might not be as damaged as other blades or maintenance was performed during the year. There were also cases where some blades started normal, but ended in an abnormal performance state, this could have been due to accumulated damage. Two wind turbines appeared to start outside of the normal region in an abnormal state and ended in a normal state. This potentially suggests that their blades suffered from performance degradation prior to the start of the data record and were repaired during the given sampling period.

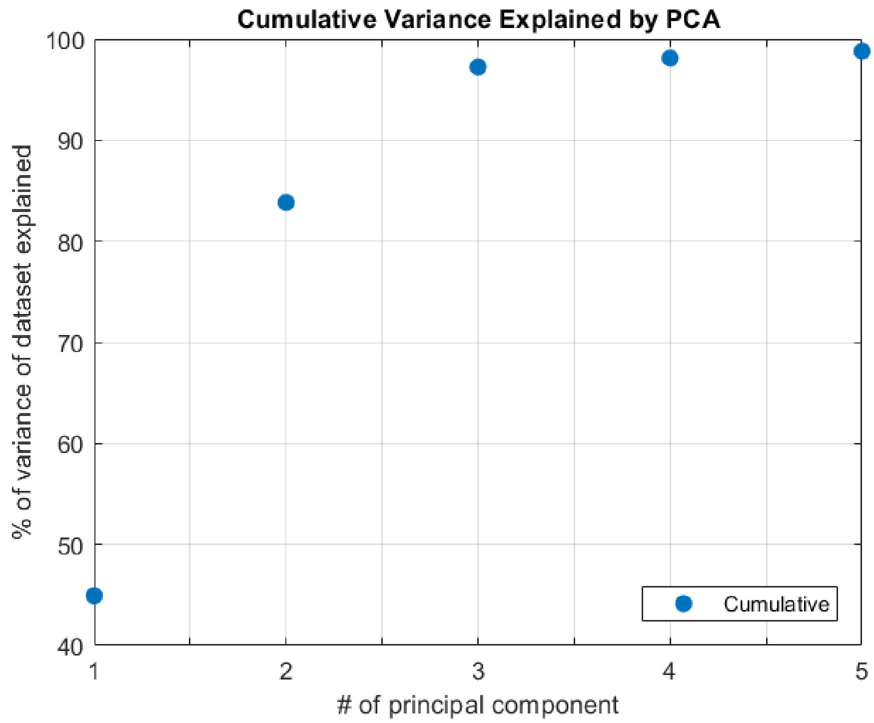


Figure 3. 1: Percentage of initial input dataset explained with each additional principal component.

To determine the performance state of the blades, boundaries were assigned to facilitate more objective assessments. With the assumption made previously that faults and failures are rare occurrences, clusters that contain a large portion of the data points would be classified as normal conditions. For this study the normal cluster contains 98% of the data points. Figure 3.4 demonstrates the execution of the proposed method where the green and white regions represent normal and abnormal blade states respectively. Here, the green line represents the path taken by each wind turbine's blades from the start of the year to the end. The path is significant to this study as it aids in visualizing the deviation of the principal components from the norm as the blades operate over time. The observed changes in behaviour of the principal components can then be used to infer the state of the blades. As seen in Figure 3.4a, there appeared to be multiple instances experienced by wind turbine 38 that fell outside of the normal region. This may indicate that the blades' state was deteriorating from normal operating conditions and repair was required. Hindcasting with historical blade maintenance data to validate these inferences showed that the blades for wind turbine 38 were indeed repaired for leading edge erosion. Figure 3.4b shows a different scenario where its path remained within the normal region aside from the single instance of abnormality. Since blade state change from normal to abnormal is a progressive process, a singular instance of abnormal behaviour may not represent degradation of the blades. Such a singularity could stem from anomalous environmental operating conditions which may not necessarily equate to a change in the health state of the blades.

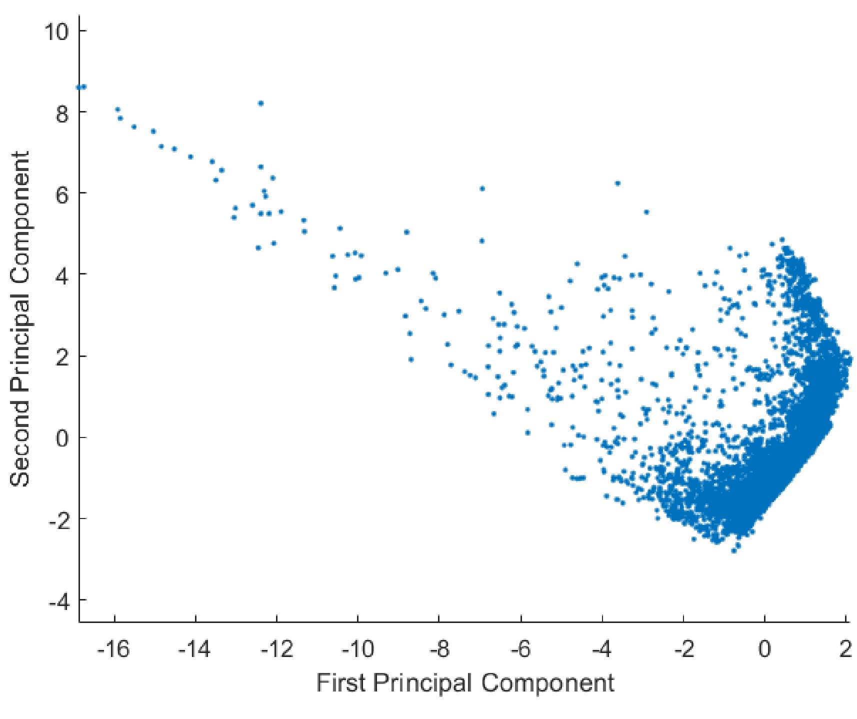


Figure 3. 2: First and second principal components for each of the 88 wind turbines are plotted against each other.

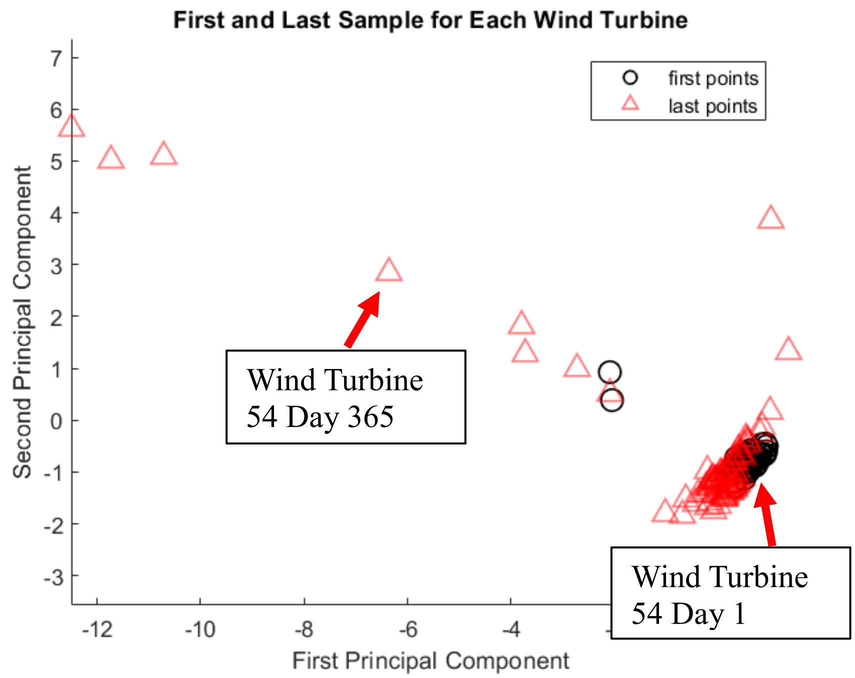


Figure 3. 3: First and last instances of each of the 88 wind turbines' principal components.

Subsequently, Figure 3.4b should indicate that wind turbine 78 was operating normally and did not require maintenance. Comparing this assessment with historical blade reports confirmed that the blades did not require maintenance. In Figure 3.4c, the blades for wind turbine 1 experienced high frequencies of abnormal instances similar to Figure 3.4a. As a result, the performance of the blades of wind turbine 1 were deemed abnormal and were identified as candidate for repair. Figure 3.4d indicated that the blades for wind turbine 68 appeared to be operating mostly normal. Historical blade data showed that the blades from turbine 1 were repaired for leading edge erosion; whereas for turbine 68, no blade repairs were performed. Thus, multiple instances of path fluctuations travelling outside of the normal region appear to relate to changes in blade performance state. Here we arbitrarily classify blade performance as abnormal, when more than a singular instance of abnormality (i.e., at least 2 instances) is observed.

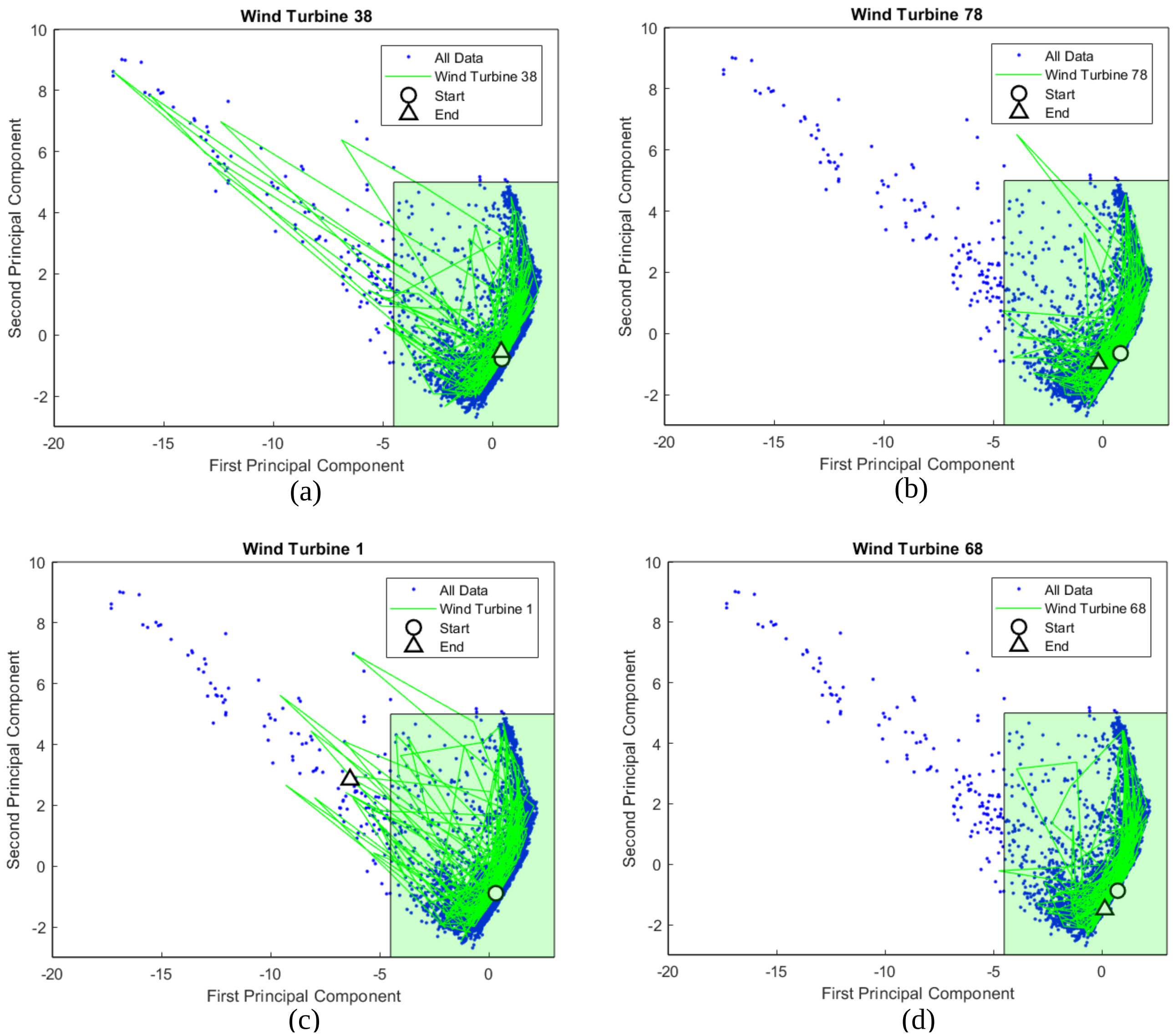


Figure 3. 4: The path that the blades of wind turbines (a) 38, (b) 78, (c) 1, and (d) 68 take respectively from the start of the dataset to the end.

Figure 3.5 features a comparison of power curves for wind turbine 1 on a day when it was classified as abnormal versus when it is classified as normal. The results indicate that for the abnormal day, there are more points on its power curve that deviate from the expected power curve compared to the normal day. This deviation shows that with the blades experiencing higher winds, it is producing much lower than the anticipated power output. This result possibly suggests that the blades are not operating normally during that day, confirming the observations made for Figure 3.4c.

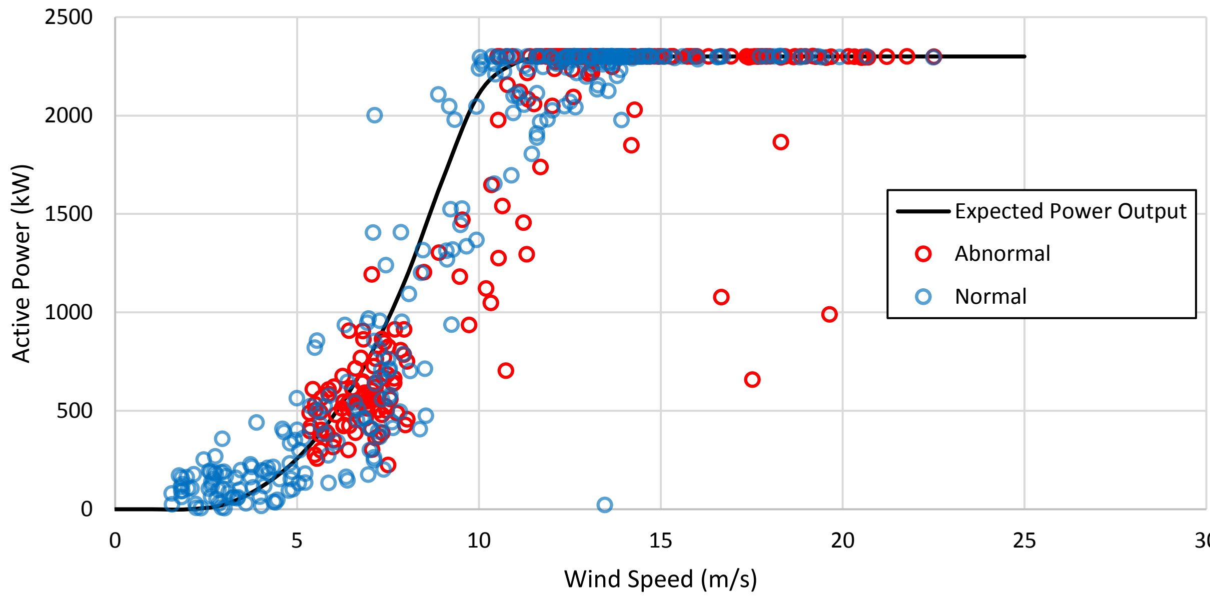


Figure 3. 5: Power curve comparison for wind turbine 1.

Figure 3.6 highlights a case where the data path shown suggests that the blades were operating normally. However, maintenance records show that repairs were performed for the blades of wind turbine 66. The results of the algorithm suggest that the blades were operating normally and repairs may not have been necessary. This could have been a case where wind farm operators were being overly conservative in their maintenance to maximize uptime and to mitigate any avoidable faults and failures. While this approach is proactive, it may not be resource optimal.

Inspection of Figure 3.7 gives insight into how this PCA approach might illustrate the impact of blade repairs on the technique's characterization of blade performance. The blades for wind turbine 17 showed a high number of instances where it operated abnormally, indicative of a need for blade maintenance. After repairs were performed, the data path appears to jump back from the abnormal region to the normal region. After this jump, the path remained in the normal region until the end of the recorded dataset. This result highlights the potential promise of the technique's capability in detecting changes of performance state. Conversely, if the blades were to require repair, the data path may not necessarily remain outside of the normal region as the blades and tower may pitch and yaw accordingly in order to perform closer to the norm. Because of this, we arbitrarily classified blades as being abnormal when more than a singular instance of abnormality is observed.

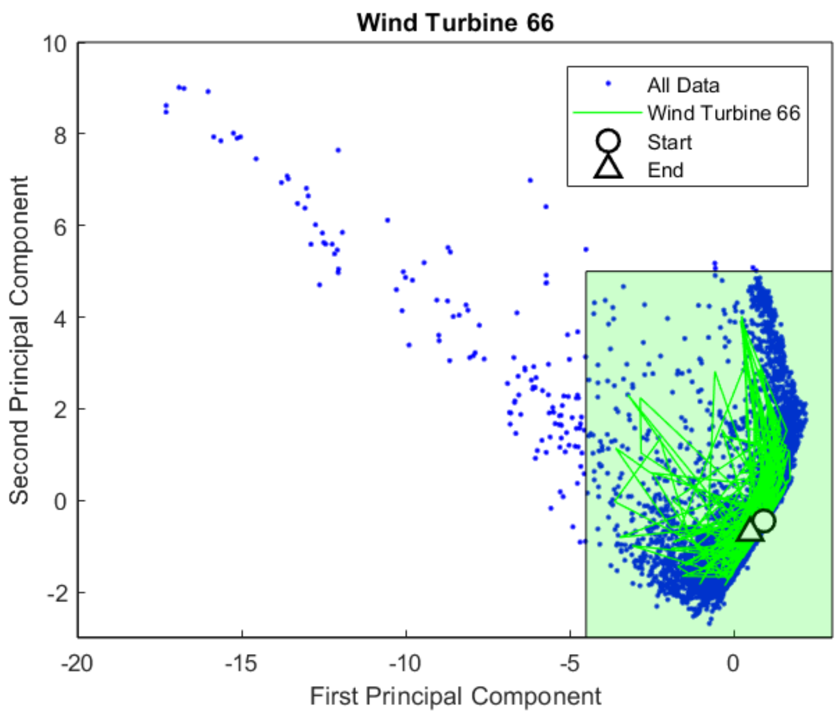


Figure 3. 6: The path taken by wind turbine 66 that appears to be operating normally, but repairs were still performed.

This strategy was applied to all 88 wind turbines within the wind farm and was evaluated using the historical blade maintenance data to validate the performance state classifications. The technique correctly identified 82% of past blade performance issues. The remaining 18% appeared to be cases where blades were classified as operating abnormally (other than the case highlighted in Figure 3.6). With the abnormal classification, these blades were recommended for repairs, but the historical blade data showed no indication of repairs. This may suggest that some blade repairs were being prioritized over others due to budget constraints. For the case of Figure 3.6, wind farm operators noted that repairs are sometimes performed not due to abnormal performance, but rather for blade integrity management (i.e. lightning or other damage) to avoid catastrophic failure or significant water absorption that could lead to larger lightning damages later on.

Results of the study were presented to the farm operators. The PCA inferences served to inform and motivate maintenance discussions. The graphic representations of blade performance states were relatable for a wider stakeholder audience. This increased stakeholder understanding can increase buy-in from budget designers and improve appreciation of maintenance requirements in general.

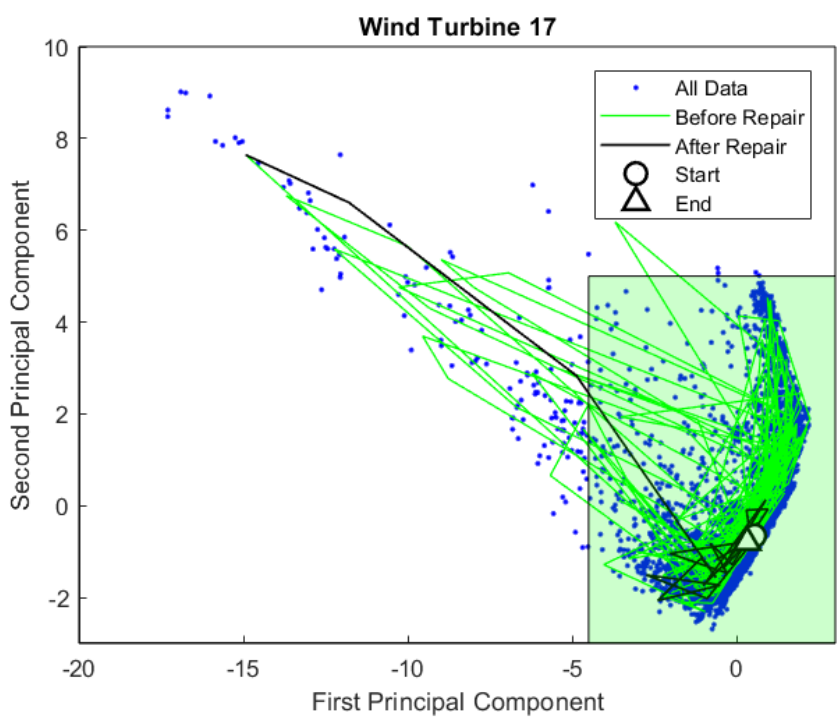


Figure 3. 7: The path taken by wind turbine 17 blades before and after repairs were performed.

4. Conclusion and areas for future research

This study presents the use of the principal component analysis (PCA) for the condition monitoring of wind turbine blades. Without the availability of specialized and invasive sensor equipment, the PCA technique showed promise in its ability to classify the performance state of blades as normal or abnormal. This approach correctly identified 82% of past blade performance issues. The proposed method is cost-efficient as it's underpinned by SCADA data that is already accessible to most wind farms. The very visual results can provide strategic insights into maintenance discussions which could ultimately improve wind farm profitability and reliability.

Future work includes a sensitivity analysis to examine the influence of different parent variables on the derived principal components. That is, which input variables among those available from the SCADA will be most descriptive of the performance state of the wind turbine blades? Further future work includes the incorporation of a third principal component to the study to determine the influence on assessment accuracy. According to Figure 3.2, with three principal components, approximately 97% of the input dataset can be explained which is a significant increase from two principal components. It is noted however that this also increases the risk of additional noise in the results. Finally, the potential for inclusion of a forecasted anomaly operating index (AOI) for this dataset will be considered. This could extend the method for prospective application in remaining useful life (RUL) predictions that are less dependent on additional specialized sensors and equipment.

Acknowledgement

The authors would like to acknowledge that this work is part of the YR21 Investment Decision Support Program supported by progressive industrial partners, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Ontario Centres of Excellence (OCE).

References

[1] R. Wiser and M. Bolinger, "2017 Wind Technologies Market Report," Lawrence Berkeley National Laboratory, 2017.

[2] Y. Sinha, J. A. Steel, J. A. Andrawus, and K. Gibson, "A SMART Software Package for Maintenance Optimisation of Offshore Wind Turbines," *Wind Eng.*, vol. 37, no. 6, pp. 569–577, 2013.

[3] Y. Sinha and J. A. Steel, "Failure Prognostic Schemes and Database Design of a Software Tool for Efficient Management of Wind Turbine Maintenance," *Wind Eng.*, vol. 39, no. 4, pp. 453–478, 2015.

[4] C. C. Ciang, J.-R. Lee, and H.-J. Bang, "Structural health monitoring for a wind turbine system: a review of damage detection methods," *Meas. Sci. Technol.*, vol. 19, no. 12, p. 20.

[5] B. S. Nivedh, "Major Failures in the Wind Turbine Components and the Importance of Periodic Inspections," *Wind Insid.*, vol. 2014, p. 5, 2014.

[6] J. A. Andrawus and L. Mackay, "Offshore Wind Turbine Blade Coating Deterioration Maintenance Model," *Wind Eng.*, vol. 35, no. 5, pp. 551–560, 2011.

[7] A. Sareen, C. A. Sapre, and M. S. Selig, "Effects of Leading-Edge Protection Tape on Wind Turbine Blade Performance," *Wind Eng.*, vol. 36, no. 5, pp. 525–534, 2012.

[8] A. Sareen, C. A. Sapre, and M. S. Selig, "Effects of leading edge erosion on wind turbine blade performance," *Wind Energy*, vol. 17, no. 10, pp. 1531–1542, 2014.

[9] N. Gaudern, "A practical study of the aerodynamic impact of wind turbine blade leading edge erosion," *J. Phys. Conf. Ser.*, vol. 524, p. 012031, 2014.

[10] B. Morton, "Understanding Wind Turbine Condition Monitoring Systems." Renewable NRG Systems, 2018.

[11] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges," *energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[12] H. Luo, C. Hatch, M. Kalb, J. Hanna, A. Weiss, and S. Sheng, "Effective and accurate approaches for wind turbine gearbox condition monitoring," *Wind Energy*, vol. 17, no. 5, pp. 715–728, 2013.

[13] D. Siegel, W. Zhao, E. Lapira, M. AbuAli, and J. Lee, “A comparative study on vibration-based condition monitoring algorithms for wind turbine drive trains,” *Wind Energy*, vol. 17, no. 5, pp. 695–714, 2014.

[14] C. S. Gray and S. J. Watson, “Physics of failure approach to wind turbine condition based maintenance,” *Wind Energy*, vol. 13, no. 5, pp. 395–405, 2010.

[15] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, “An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox,” *Wind Energy*, vol. 20, no. 8, pp. 1421–1438, 2017.

[16] M. Schlechtingen and I. F. Santos, “Condition monitoring with wind turbine scada data using neuro-fuzzy normal behavior models,” *Appl. Soft Comput.*, vol. 13, no. 1, pp. 259–270, 2013.

[17] Y. Zhao, D. Li, A. Dong, D. Kang, Q. Lv, and L. Shang, “Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data,” *energies*, vol. 10, no. 8, p. 1210, 2016.

[18] Y. Zhao, D. Li, A. Dong, J. Lin, D. Kang, and L. Shang, “Fault Prognosis of Wind Turbine Generator Using SCADA Data,” Tongji University, 2016.

[19] F. Pozo, Y. Vidal, and Ó. Salgado, “Wind Turbine Condition Monitoring Strategy through Multiway PCA and Multivariate Inference,” *energies*, vol. 11, no. 4, p. 749, 2018.

[20] F. Pozo and Y. Vidal, “Wind Turbine Fault Detection through Principal Component Analysis and Statistical Hypothesis Testing,” *energies*, vol. 9, no. 1, p. 3, 2016.

[21] Y. Wang, X. Ma, and P. Qian, “Wind Turbine Fault Detection and Identification Through PCA-Based Optimal Variable Selection,” *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1627–1635, 2018.

[22] C. S. Gray, F. Langmayr, N. Haselgruber, and S. J. Watson, “A Practical Approach to the Use of SCADA Data for Optimized Wind Turbine Condition Based Maintenance,” *EWEA Offshore Wind Amst.*, p. 10, 2011.

[23] C.-J. Bai and Y.-C. Shiah, “Power-Curve Corrections for Horizontal-Axis Wind Turbine by an Improved Blade Element Momentum Theory,” *J. Mech.*, vol. 33, no. 03, pp. 341–349, 2016.

[24] P. W. Richards, D. T. Griffth, and D. H. Hodges, “Smart Loads Management for Damaged Offshore Wind Turbine Blades,” *Wind Eng.*, vol. 39, no. 4, pp. 419–436, 2015.

[25] E. Lapira, D. Brisset, H. D. Ardakani, D. Siegel, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renew. Energy*, vol. 45, pp. 86–95, 2012.

[26] D. McLaughlin, P. Clive, and J. McKenzie, “Staying Ahead of the Wind Power Curve,” *Renew. Energy World*, vol. 13, no. 2, 2010.

[27] M. D. Farrell and R. M. Mersereau, “On the impact of PCA dimension reduction for hyperspectral detection of difficult targets,” *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, 2005.

[28] A. Malhi and R. X. Gao, “PCA-Based Feature Selection Scheme for Machine Defect Classification,” *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, 2004.

CHAPTER 4

Condition Monitoring of Wind Turbine Blades through k-means Clustering

Jones Shen[1], Rupp Carriveau[1], David S.K. Ting[1]

[1] Turbulence and Energy Laboratory, University of Windsor, Windsor, Ontario, Canada

1. Introduction

Wind turbines have evolved at a drastic rate in terms of technology, increased power production, and decreased costs [1]. With its fast growth, the focus is shifted towards methods that improve reliability where studies [2], [3] show that utilizing predictive maintenance schedules can reduce operation & maintenance (O&M) costs. This is very important as O&M costs make up about 25-30% of energy generation costs where a large portion of this percentage is caused by unforeseen failures [4]–[6]. Sinha and Steel [7] illustrates the high initial capital cost for onshore wind farms causing higher O&M costs and therefore, prolonging the payback period and increasing the cost of wind energy.

Wind turbine blades play a huge role in O&M costs due to their operation in extreme atmospheric conditions [8 – 11]. Because of this, the blades are susceptible to many sources of damages such as lightning, bird strikes, ice, sand, rain, precipitation, UV rays and extreme winds [12]. This can inadvertently create overload and in-service damages for the blades and other components in the turbine. If the damages were to be left unchecked, it can create an imbalance on the blade and rotor assembly which can put the tower and power train under unwarranted dynamic loads [13], even leading to the collapse of entire machines in severe cases [14]. With the help of a CM system, a predictive maintenance strategy can be implemented within wind farms to help improve wind turbine blade reliability and reduce O&M costs by capitalizing on the lead times created by the predictions of the blade performance state being made early to avoid major breakdowns and failures as well as minimizing any associated downtimes [7].

A CM system typically reports on specific wind turbine component health [15]–[19]. These components also happen to be mainly drivetrain elements [20]–[23] where the science of diagnosing their state health is quite advanced. However, most wind farm blade populations lack the level of detailed and direct structural performance data provided by specialized sensors due to their prohibitively high costs. Thus, the use of abundant and readily available supervisory control and data acquisition (SCADA) data is appealing. While SCADA data is collected at nearly all farms, in many cases it is not being used to its fullest capabilities.

Yang et al [24] proposes an unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring which involves the detection of anomalies where a spatiotemporal pattern network (STPN), is applied to extract spatial and temporal features between the unlabeled variables in the SCADA system, and an energy-based model is used to capture the system-wide patterns for condition monitoring. Another unsupervised machine learning method to detect anomalies within generator performance metrics was proven to be successful in performing generator fault prediction, remaining useful life (RUL) estimation and fault type diagnosis using SCADA data [25], [26]. An anomaly operation index (AOI) is forecasted using an ARIMA model where a higher AOI would represent a lower remaining useful life (RUL). The AOI describes the proportion of unhealthy instances to all instances during a given time period can be determined which would highlight the probability of performance degradation over time. Qiu et al [27] goes in a different direction with available SCADA data by incorporating underutilized SCADA alarms into their study due to their large number of occurrences in a 10-minute SCADA period. First, the origins of the SCADA alarms are determined, followed by applying a methodology adopted from an oil and gas industry standard that prioritizes alarms to show the seriousness of the alarm data volume. Finally, two methods of alarm analysis, time-sequence and probability-based, are proposed and demonstrated on the data which show the potential for providing valuable fault detection, diagnosis and prognosis. A vibration based structural health monitoring (SHM) system was applied to wind turbine blades where mechanical energy was introduced by means of an electromechanical actuator mounted inside the blade [28]. The actuator is used to induce vibrations that propagate along the blade and are measured by an array of accelerometers. Unsupervised learning is applied to the data and vibration patterns corresponding to the undamaged blade are used to create a statistical model of the reference state. When the vibration is introduced, its current vibration pattern is compared with the reference state, and the differences can then be associated with damage. Zhang and Kusiak [29] introduces a method for detecting abnormalities of wind turbine vibrations using readily available SCADA data. An unsupervised k-means clustering algorithm was applied to detect abnormal values experienced in the wind turbine drivetrain and tower acceleration parameters to differentiate between normal and faulty wind turbines.

In this work, the proposed methodology loosely follows the studies cited in [28], [29] where a SCADA data based, k-means clustering algorithm is developed. The motivation for using k-means lies in its advantages of being able to systematically cluster data points while having a low computational run time. These advantages addresses disadvantages observed from utilizing the PCA method described in Chapter 3. The k-means technique can then be utilized to infer the performance health state of the blades to determine if they are operating in a normal or abnormal condition. These cluster classifications are then plotted against a given time series to simulate vibrational data similar to [28]. The purpose for this is to visualize and infer the performance state of the blades through the fluctuations observed by classifying differences from the normal observations as abnormal behaviours. To consistently identify the blade's performance state, a moving window is implemented on the vibrational data. The moving window has an assigned threshold such that if a certain number of instances were to exceed this boundary within a timeframe, the blades are deemed abnormal and will require repairs. Being able to infer the performance state will improve the lead time available to wind farm operators to make informed maintenance decisions.

2. Methodology

2.1 Data Collection

The SCADA data used for this study was collected over a period of 12 months from a real-world wind farm consisting of 88 wind turbines located in Southwestern Ontario, Canada as shown in Table 4.1. For each wind turbine, two separate sets of data was provided: 10-minute averaged SCADA data and the maintenance reports specifically for the blades which provide information on what types of damages and repairs were experienced as well as the times of the occurences. Table 4.2 shows the 10-minute averaged SCADA data variables of interest extracted from each of the wind turbines.

Table 4. 1: Wind farm characteristics.

Feature	**Wind farm**	
Number of turbines	48	40
Rotor diameter (m)	93	101
Hub height (m)	80	80
Rated power (MW)	2.3	2.3
Data set size	2,522,880	2,102,400

Table 4. 2: SCADA data variables of interest.

Number	Sensor Type	Units
1	Active Power	kW
2	Wind Speed	m/s
3	Rotor Speed	RPM
4	Yaw	deg
5	Blade A Pitch Angle	deg
6	Blade B Pitch Angle	deg
7	Blade C Pitch Angle	deg

2.2 Variable Selection

The selection of appropriate variables is critical to the k-means algorithm in order to produce valuable information for blades. The SCADA chosen were active power output, wind speed, rotor speed, yaw, and blade pitch control. The turbine active power, wind speed, and rotor speed were chosen as changes in their SCADA values can be observed due to the deteriorating geometry of the blades caused by leading edge erosion or other types of structural discontinuities [30]–[33]. Changes in the yaw and blade pitch angles can also be detected as their optimal angles for maximum power output will also differ due to their deteriorating blade geometries [30]–[33]. The practice of derating damaged blades to limit loads will also alter the rotor speed and blade pitch at the cost of reduced power production [34]. These variables may serve as features that may reflect the blade's performance health state which will allow for the identification of blades in normal or abnormal conditions.

2.3 Data Pre-Processing

Collected SCADA data can be riddled with errors due to faults in the data collection system as well as curtailment values. Thus, it is essential to filter these values which can be identified as either missing, inappropriately positioned, and contradicting values. Within the region of this study, it is quite common for wind farm operators to curtail energy production when the demand for it is low. Curtailment can be performed by adjusting the blade pitch angles, resulting in decreased electrical power production for a given wind speed. [35], [36] draws attention to the fact that the losses of power output could be misinterpreted as being a result of blade performance abnormalities or degradation, which therefore, shows the importance of filtering these instances within the SCADA data.

2.4 k-means Clustering Algorithm

The k-means clustering algorithm was used to develop a simplified model for monitoring the vibration of the wind turbine blades and inferring their performance health states. The k-means method is a type of algorithm that clusters similar data points together where similarity is measured by the Euclidean distance between each point. The input dataset $x \left\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\right\}$ is grouped into a k number of clusters. The intuition behind k-means is an iterative process that begins with initial centroids of the clusters which is refined by repeatedly assigning data points to their closest centroids and then recomputing the centroids based on the data point assignments until convergence is met on a final set of centroids [37].

To start off, the parameters of the dataset x are normalized to [0, 1] before implement the k-means algorithm. The algorithm then assigns each data point example $x^{(i)}$ to its closest centroid given the positions of the current centroids. For every example i, k-means assigns each point $x^{(i)}$

to the cluster whose mean has the least squared Euclidean distance or essentially, the "nearest" mean:

$$c^{(i)} = \left\{x^{(i)}: \left\|x^{(i)} - \mu_j\right\|^2 \leq \left\|x^{(i)} - \mu_{j^*}\right\|^2, j^* = 1, 2, \dots, k\right\} \tag{4}$$

Where $c^{(i)}$ represents the index of the cluster's centroid that is closest to the dataset value $x^{(i)}$, and μ_j is the current position of the j'th centroid. After each dataset value $x^{(i)}$ is assigned to a centroid, the algorithm recomputes the new means to be the centroids of the clusters. For every centroid *k*,

$$\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} x^{(i)} \tag{5}$$

Where $C_k$is the number of data points that are assigned to centroid *k*. For example, if two points $x^{(1)}$ and $x^{(8)}$ are assigned to centroid 5, then the new mean should be calculated as $\mu_5 = \frac{1}{2}\left(x^{(1)} + x^{(8)}\right)$. The algorithm continues to run until it has converged when the assignments no longer change.

Once convergence is met, the final cluster indexes $c^{(i)}$ for each dataset value $x^{(i)}$ are plotted for the given time domain to visualize the vibrations experienced by the wind turbine blades based off of the SCADA variables of interest. To determine if the blades are operating abnormally, a moving window is implemented on the vibration plot where a threshold is set such that if a certain number of abnormal instances experienced exceeds the threshold within the moving window's timeframe, the blades are deemed abnormal and will require repairs given the set criterion.

Overall, the key steps for the k-means method in order to assess the wind turbine blade state are summarized as follows:

Step 1: Assign each dataset value $x^{(i)}$, to its closest centroid

Step 2: Recompute the mean of each centroid using the dataset values assigned to it.

Step 3: Plot all the classified centroid indexes $c^{(i)}$ for each dataset value $x^{(i)}$ against the specified time domain.

Step 4: Implement moving window across the time domain of the plotted indexes to determine if the blades are normal or abnormal, therefore dictating whether repairs are required.

3. Results and Discussion

In this section, the proposed k-means blade performance health monitoring method is applied and evaluated.

Table 4.3 summarizes the clustering results produced by the k-means algorithm. The clusters are indexed from 1 to 2 where in this study. The cluster number, the centroid values of each of the clusters, the number of data points assigned to each cluster, and the percentage of these data points are in each cluster are displayed.

As shown in Table 4.3, the clusters represent the performance health state of the wind turbine blades. Looking more closely at the centroid values of each of the clusters, cluster 1 appears to be associated with blades operating in normal conditions since the corresponding centroid values appear to resemble SCADA values observed from a wind turbine operating at a somewhat regular capacity. On the contrary, cluster 2 appears to contain data points that may be representative of blades that are operating in abnormal conditions as the centroid values appear to be fairly low compared to the values from cluster 1 which may indicate periods of faults or maintenance being performed.

Table 4. 3: Summary of clustering results.

Cluster Number	Rotor Speed (RPM)	Blade A Pitch (deg)	Blade B Pitch (deg)	Blade C Pitch (deg)	Wind Speed (m/s)	Yaw (deg)	Active Power (kW)	Number of Points	Percentage (%)
1	14.2	2.93	2.93	2.94	9.82	205.8	1553	7731	33.8
2	9.59	2.30	2.30	2.29	4.95	178.8	435.0	15150	66.2

Figure 4.1 shows the clustering results from the use of the k-means algorithm. The horizontal axis represents the values of the rotor speed and the vertical axis shows the values of blade A's pitch angle. Each point is representative of one day over one year, per turbine. In this case, this amounts to 365 days * 88 turbines or 32,120 points. The assignments of the data points in the 2 clusters are labelled to differentiate between what is considered a normal or abnormal operating state. The plot shows that there is an overwhelming amount of points that fall under the abnormal cluster, specifically when the rotor speed drop below 14 RPM and when the blade pitch begins to increase. This is also confirmed in Table 3 where the percentage of data points in cluster 2 suggests that the blades are operating abnormally at an alarming amount of 66% of the year.

Figure 4.2 shows the same clustering results, but with wind speed as the horizontal axis and active power produced in the vertical axis. This visual depiction of the cluster assignments suggests that high wind speeds and power produced are indicative of normal conditions whereas the low wind speeds and power produced resemble abnormal conditions. This may imply that the lower wind speed and power values could be representative of blades that are experiencing faults or operating in a deteriorated state. A lower power output may also be an indication of derating to limit the loads experienced by deteriorated blades which would come at a cost of reduced power production.

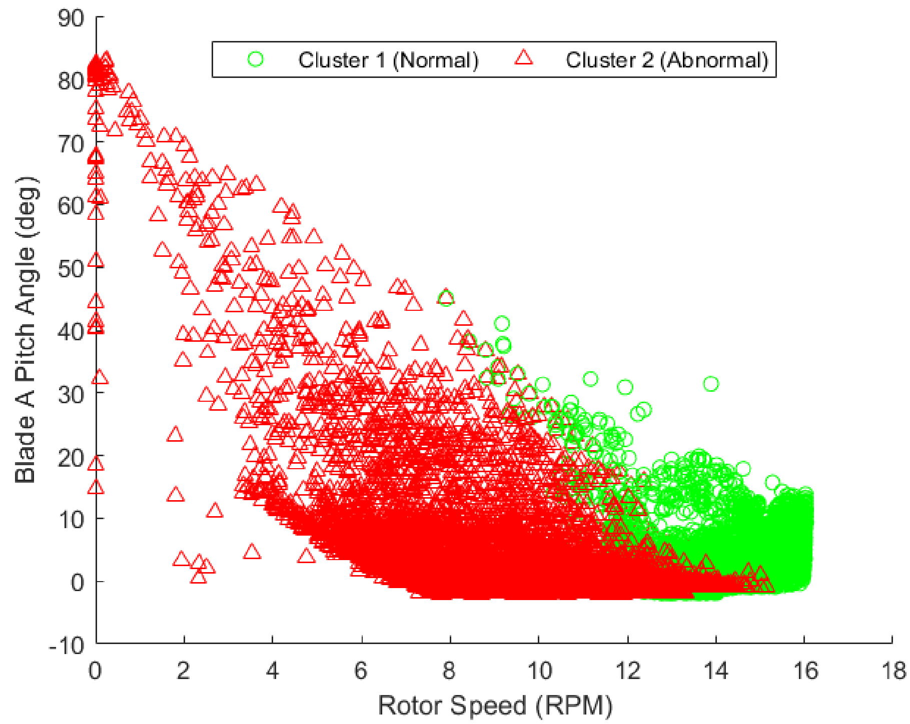


Figure 4. 1: Clustering results from monitoring blade SCADA parameters with rotor speed and blade pitch shown.

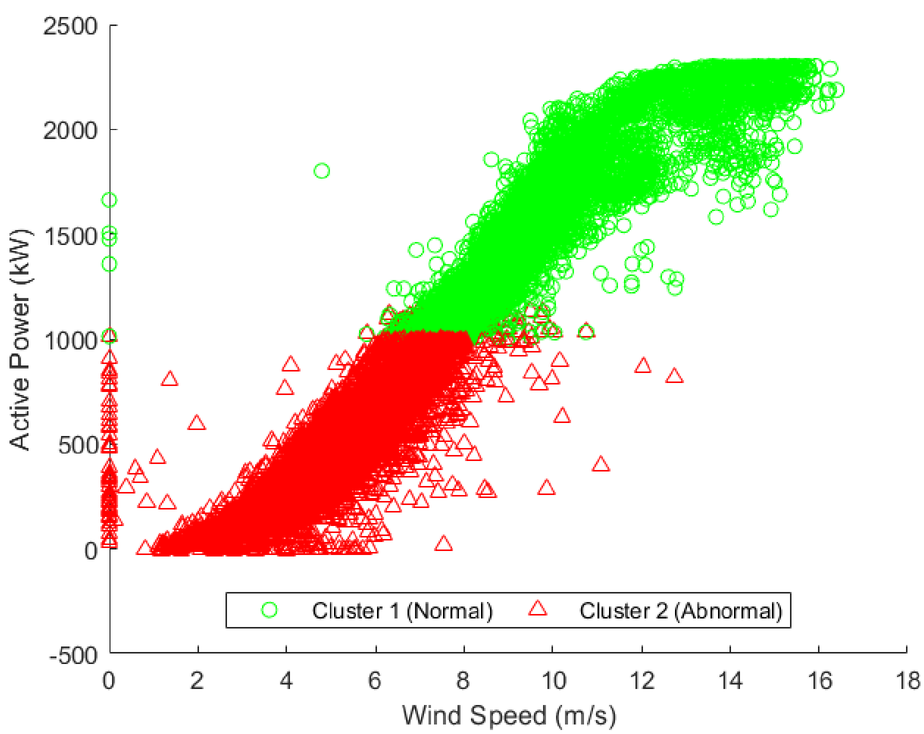


Figure 4. 2: Clustering results from monitoring blade SCADA parameters with wind speed and power shown.

To visualize these cluster classifications for each turbines' set of blades, the cluster assignments were plotted against the time domain of the dataset. The 10-minute averaged SCADA was compressed into daily averages in order to summarize the cluster indexes versus the day. The motivation behind this visualization was to determine if these classifications can identify which blades are in need of repairs before they were ever performed. The first 260 days are plotted instead of the full year as repairs were performed between day 261 and 365 for the blades in the wind farm. Figure 4.3 is an example of this where turbine 38's blade state fluctuates between normal and abnormal throughout the year. However, by around day 175, the blades appeared to remain in the abnormal cluster which suggests that repairs may be required. Historical blade maintenance data showed that the blades for turbine 38 suffered from leading edge erosion during the time of this dataset and repairs were performed. Subsequently, Figure 4.4 shows the opposite case for turbine 47 blades where the fluctuations between the normal and abnormal clusters continue throughout the 260 days. Although delays were observed between each fluctuation near the end of the observed dataset, the fluctuations remained. Compared to Figure 4.3, Figure 4.4 represented a case where no repairs were performed according to the historical blade data. This may suggest that the fluctuations between cluster 1 and cluster 2 could indicate that the blades are operating in a normal state as these fluctuations are consistent in Figure 4.4 where repairs were not performed.

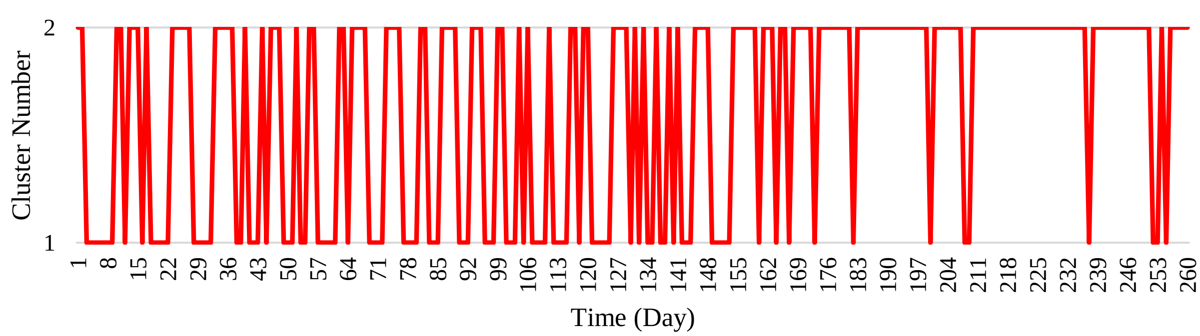


Figure 4. 3: Assigned clusters for each day for the first 260 days of the dataset for wind turbine 38.

It also appears that the delays between fluctuations seem to increase over time until the points remain in cluster 2 which may signify the gradual deviation from the norm. Therefore, any staggering differences observed from the normal fluctuation may be associated with abnormal behaviours and may need repairs.

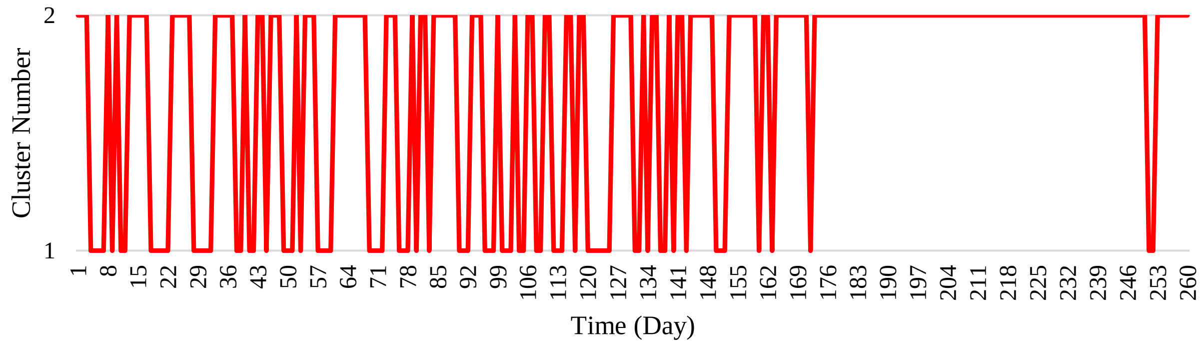


Figure 4. 4: Assigned clusters for each day for the first 260 days of the dataset for wind turbine 47.

In order to consistently identify the state of blades to determine if repairs are required, a moving window was implemented. The window enforced is 50 days longs (i.e., Day 1 to Day 50) where it starts on day 1 of the dataset and advances one day at a time. The window has an assigned threshold such that if instances are assigned to cluster 2 more than 48 out of 50 times during that timeframe, we arbitrarily classify the blade performance as abnormal. In the case of Figure 4.5, this threshold was exceeded in the window that covers Day 169 to Day 219.

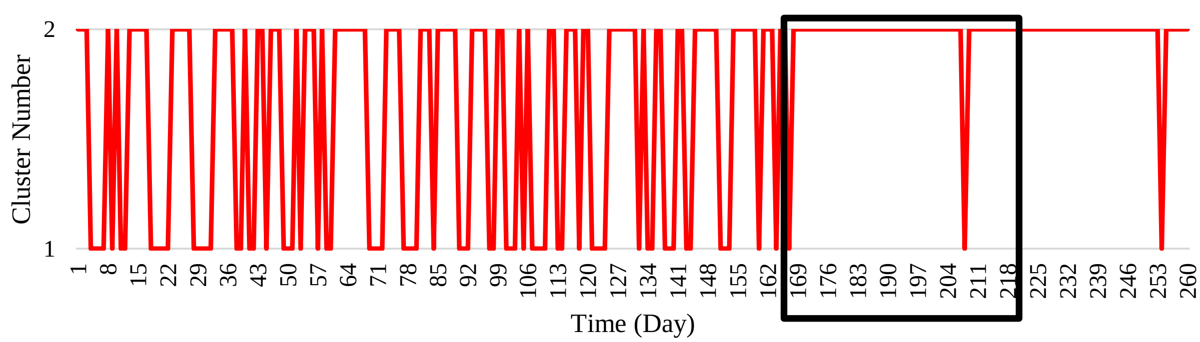


Figure 4. 5: Assigned clusters for each day for the first 260 days of the dataset for wind turbine 42 with a 50-day moving window.

Thus, the blades for wind turbine 42 were classified as abnormal and repairs were recommended. Validating this identification with historical blade data showed that the blades were in fact repaired right after the time that this data recording took place. In addition, given that this observation can only be made after Day 219, this method was still able to identify that the blades needed repairs over 2 months prior to the actual date of the real-life repairs. Figure 4.6 shows an example where blades are identified as operating in normal conditions using the moving window method. The window's threshold of cluster 2 instances being greater than 48 out of 50 total instances was never exceeded. There were times however, where this threshold was eclipsed a few times where 47 out of the 50 total instances in the window were classified as abnormal occurrences. This may indicate that although action was not required to be taken at that time, the impending deviation from the normal to abnormal classification was near and repairs could be required soon.

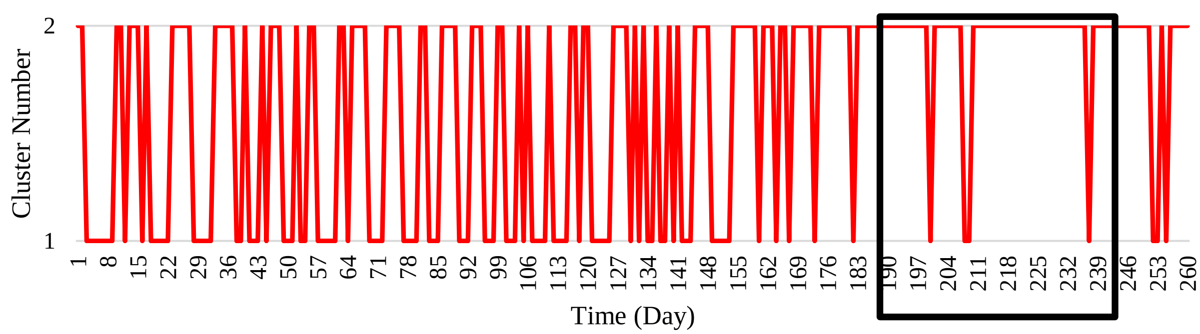


Figure 4. 6: Assigned clusters for each day for the first 260 days of the dataset for wind turbine 47 with a 50-day moving window.

Applying this moving window method to all 88 wind turbines showed that it was able to correctly identify the condition of the blades 75% of past blade performance issues. The remaining 25% consisted of the method classifying blades as operating abnormally and action was required, but no blade repairs were actually performed according to historical reports. This could imply that some repairs for specific blades may have been prioritized over others due to the wind farm operator's restriction of not having unlimited resources for repairs. There were also a couple cases observed where the algorithm identified some blades were operating normally since the window's threshold was never exceeded whereas the historical data showed that repairs were performed. This could suggest that wind farm operators were overly conservative with their repairs to mitigate any preventable downtimes and failures to their blades. Operators of the wind farm confirmed this point as they noted that repairs are sometimes performed not due to abnormal performance, but rather for blade integrity management (i.e. lightning or other damage) to avoid catastrophic failure or significant water absorption that could lead to larger lightning damages later on.

The k-means clustering algorithm served to inform and generate maintenance discussions. With the results presented, it shows great potential in enhancing the farm operator's strategic planning for performing blade inspections and repairs accordingly to optimize the use of resources and to increase uptime of the wind turbines.

4. Conclusion and areas for future research

In this study, the use of the unsupervised learning method, the k-means clustering algorithm is implemented to identify the performance health state of the blades to determine if they are in a normal or abnormal condition. The k-means algorithm grouped data into either cluster 1 or cluster 2 according to their similarity (Euclidean distance) where the clusters represented normal and abnormal operating conditions respectively. Plotting the cluster indexes over a time series allowed for the visualization of the performance state of the turbine blades throughout the time domain of the dataset. Using the moving window with the threshold assigned, the technique correctly identified 75% of past blade performance issues with hindcasted historical blade data. With its dependence on readily available SCADA data rather than specialized and invasive sensors, this method has a low capital cost overhead, and once established, is relatively inexpensive to manage.

The results illustrate the potential to provide major insights into maintenance discussions which could ultimately increase incentives for investment in clean wind energy.

The next step to this research is evaluate the clustering cost to determine if there are more clusters that can detect abnormal instances for a more detailed analysis due to the high volume of SCADA parameters being used. Another approach could be to utilize principal component analysis (PCA) to reduce the dimensionality of the input data into new variables that act as linear combinations of select parent variables. The reduced set of variables (the principal components) can improve the ability to visualize the data. The incorporation of these methods offer a promising research direction.

Acknowledgement

This work was supported in part by Carina (Xue) Luo of the Leddy Library, University of Windsor. The authors would also like to acknowledge that this work is part of the YR21 Investment Decision Support Program supported by progressive industrial partners, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Ontario Centres of Excellence.

References

[1] R. Wiser and M. Bolinger, “2017 Wind Technologies Market Report,” Lawrence Berkeley National Laboratory, 2017.

[2] P. Tavner, J. Xiang, and F. Spinato, “Reliability Analysis for Wind Turbines,” *Wind Energy*, vol. 10, no. 1, pp. 1–18, 2007.

[3] F. Spinato, P. Tavner, G. van Bussel, and E. Koutoulakos, “Reliability of wind turbine subassemblies,” *IET Renew. Power Gener.*, vol. 3, no. 4, pp. 387–401, 2009.

[4] E. Wiggelinkhuizen, T. Verbruggen, H. Braam, L. Rademakers, J. Xiang, and S. Watson, "Assessment of Condition Monitoring Techniques for Offshore Wind Farms," *J. Sol. Energy Eng.*, vol. 130, no. 3, pp. 031004-1–9, 2008.

[5] X. Ma, P. Cross, and P. Qian, "A condition monitoring system for an early warning of developing faults in wind turbine electrical systems," *Insight Non-Destr. Test. Cond. Monit.*, vol. 58, no. 12, pp. 663–670, 2016.

[6] J. Zhou, S. Roshanmanesh, F. Hayati, V. Jantara Junior, T. Wang, S. Hajiabady, X. Y. Li, H. Basoalto, H. Dong, and M. Papaelias, "Improving the reliability of industrial multi-MW wind turbines," *Insight Non-Destr. Test. Cond. Monit.*, vol. 59, no. 4, pp. 189–195, 2017.

[7] Y. Sinha and J. A. Steel, "Failure Prognostic Schemes and Database Design of a Software Tool for Efficient Management of Wind Turbine Maintenance," *Wind Eng.*, vol. 39, no. 4, pp. 453–478, 2015.

[8] M. Khan, M. Iqbal, and F. Khan, "Reliability and condition monitoring of a wind turbine," *18th Ann Can. Conf Electr. Comput. Eng.*, pp. 1978–1981, 2005.

[9] B. S. Nivedh, "Major Failures in the Wind Turbine Components and the Importance of Periodic Inspections," *Wind Insid.*, vol. 2014, p. 5, 2014.

[10] M. Flemming and S. Troels, "New lightning qualification test procedure for large wind turbine blades," *Int Conf Light. Static Electr.*, pp. 36–46, 2003.

[11] W. Yang, "Condition Monitoring the Drive Train of a Direct Drive Permanent Magnet Wind Turbine Using Generator Electrical Signals," *J. Sol. Energy Eng.*, vol. 136, no. 2, pp. 021008-1–8, 2014.

[12] U. I. K. Galappaththi, A. K. M. De Silva, M. Macdonald, and O. R. Adewale, "Review of inspection and quality control techniques for composite wind turbine blades," *Insight Non-Destr. Test. Cond. Monit.*, vol. 54, no. 2, pp. 82–85, 2012.

[13] R. Ramlau and J. Niebsch, "Imbalance Estimation Without Test Masses for Wind Turbines," *J. Sol. Energy Eng.*, vol. 131, no. 1, pp. 011010-1–7, 2009.

[14] C. C. Ciang, J.-R. Lee, and H.-J. Bang, "Structural health monitoring for a wind turbine system: a review of damage detection methods," *Meas. Sci. Technol.*, vol. 19, no. 12, p. 20.

[15] Z. Tian, T. Jin, B. Wu, and F. Ding, "Condition based maintenance optimization for wind power generation systems under continuous monitoring," *Renew. Energy*, vol. 36, pp. 1502–1509, 2011.

[16] B. Morton, “Understanding Wind Turbine Condition Monitoring Systems.” Renewable NRG Systems, 2018.

[17] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, “Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges,” *energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[18] H. Luo, C. Hatch, M. Kalb, J. Hanna, A. Weiss, and S. Sheng, “Effective and accurate approaches for wind turbine gearbox condition monitoring,” *Wind Energy*, vol. 17, no. 5, pp. 715–728, 2013.

[19] W. Yang, C. H. Ng, and J. Jiang, “Wind Turbine Condition and Power Quality Monitoring by the Approach of Fast Individual Harmonic Extraction,” *J. Sol. Energy Eng.*, vol. 135, no. 3, pp. 034504-1–5, 2013.

[20] A. F. Hills, Z. Q. Lang, S. Soua, T.-H. Gan, A. Perera, P. van Lieshout, and B. Grainger, “A novel baseline model-based technique for condition monitoring of wind turbine components,” *Insight Non-Destr. Test. Cond. Monit.*, vol. 53, no. 8, pp. 434–438, 2011.

[21] E. Artigao, J. L. Ferrando, T.-H. Gan, B. Wang, V. Kappatos, C. Selcuk, and S. Soua, “Development of a baseline as a novel strategy for a condition monitoring system for the assessment of wind turbine generators,” *Insight Non-Destr. Test. Cond. Monit.*, vol. 56, no. 8, pp. 417–421, 2014.

[22] W. Yang, R. Court, and J. Jiang, “Wind turbine condition monitoring by the approach of SCADA data analysis,” *Renew. Energy*, vol. 53, pp. 365–376, 2013.

[23] A. Romero, S. Soua, T.-H. Gan, and B. Wang, “Condition monitoring of a wind turbine drive train based on its power dependant vibrations,” *Renew. Energy*, vol. 123, pp. 817–827, 2018.

[24] W. Yang, C. Liu, and D. Jiang, “An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring,” *Renew. Energy*, vol. 127, pp. 230–241, 2018.

[25] Y. Zhao, D. Li, A. Dong, D. Kang, Q. Lv, and L. Shang, “Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data,” *energies*, vol. 10, no. 8, p. 1210, 2016.

[26] Y. Zhao, D. Li, A. Dong, J. Lin, D. Kang, and L. Shang, “Fault Prognosis of Wind Turbine Generator Using SCADA Data,” Tongji University, 2016.

[27] Y. Qiu, Y. Feng, P. Tavner, P. Richardson, G. Erdos, and B. Chen, “Wind turbine SCADA alarm analysis for improving reliability,” *Wind Energy*, vol. 15, no. 8, pp. 951–966, 2012.

[28] D. Tcherniak and L. Mølgaard, “Vibration-based SHM System: Application to Wind Turbine Blades,” *J. Phys. Conf. Ser.*, vol. 628, p. 012072, 2015.

[29] Z. Zhang and A. Kusiak, “Monitoring Wind Turbine Vibration Based on SCADA Data,” *J. Sol. Energy Eng.*, vol. 134, no. 2, pp. 021004-1–12, 2012.

[30] A. Sareen, C. A. Sapre, and M. S. Selig, “Effects of leading edge erosion on wind turbine blade performance,” *Wind Energy*, vol. 17, no. 10, pp. 1531–1542, 2014.

[31] N. Gaudern, “A practical study of the aerodynamic impact of wind turbine blade leading edge erosion,” *J. Phys. Conf. Ser.*, vol. 524, p. 012031, 2014.

[32] C. S. Gray, F. Langmayr, N. Haselgruber, and S. J. Watson, “A Practical Approach to the Use of SCADA Data for Optimized Wind Turbine Condition Based Maintenance,” *EWEA Offshore Wind Amst.*, p. 10, 2011.

[33] C.-J. Bai and Y.-C. Shiah, “Power-Curve Corrections for Horizontal-Axis Wind Turbine by an Improved Blade Element Momentum Theory,” *J. Mech.*, vol. 33, no. 03, pp. 341–349, 2016.

[34] P. W. Richards, D. T. Griffth, and D. H. Hodges, “Smart Loads Management for Damaged Offshore Wind Turbine Blades,” *Wind Eng.*, vol. 39, no. 4, pp. 419–436, 2015.

[35] E. Lapira, D. Brisset, H. D. Ardakani, D. Siegel, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renew. Energy*, vol. 45, pp. 86–95, 2012.

[36] D. McLaughlin, P. Clive, and J. McKenzie, “Staying Ahead of the Wind Power Curve,” *Renew. Energy World*, vol. 13, no. 2, 2010.

[37] D. MacKay, “Chapter 20. An Example Inference Task: Clustering,” in *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003, pp. 284–292.

CHAPTER 5

Conclusions and Recommendations

1. Summary and Conclusions

It is vital for condition monitoring systems to be relatable to a wider stakeholder audience as well as being efficient and budget friendly. As such, studies on determining the performance states of wind turbine blades have been presented.

In Chapter 2, a simplified yet versatile ranking and GIS visualization method for inferring usage and providing high level insight into the loading history and behaviour of wind turbines has been presented. Rather than the implementation of currently existing complicated methods, this simplified ranking method can be used by a wider audience rather than just skilled and trained professionals. By doing so, it can improve wind farm operators' understanding in their machines by inferring turbine usage which is valuable as it can increase the amount of information available when considering potential maintenance decisions. The results also showed that the GIS visualizations was capable of reaching similar findings to that of wake loss models. These findings support the importance of optimizing turbine layout to achieve maximum production for the farm.

Chapter 3 demonstrated the use of the principal component analysis (PCA) for the condition monitoring of wind turbine blades to better understand blade behaviour throughout its operation. The PCA showcased its great capability in reducing the large SCADA dataset which makes it very easy to visualize the results. With the use of interdependent variables in this chapter, the systematic prediction of the major dependent variable, blade performance state, cannot be performed using regression. Thus, the PCA technique is instead used to infer the performance state of blades as being normal or abnormal through the statistical clustering of the principal components. This approach correctly identified 82% of past blade performance issues. The application of the PCA showed that it needs an improved method for the clustering of data points rather than the use of arbitrary classifications. A limitation in this technique lies in its inability to identify the specific cause for the deterioration of the performance state. A potential solution for this can be to implement a diagnosis model that can classify future faults using historical blade status data.

An alternative condition monitoring method was implemented in Chapter 4 using a unique k-means clustering algorithm to interpret the change in blade behaviour during operation. The k-means algorithm grouped data into either cluster 1 or cluster 2 according to their similarity (Euclidean distance) where the clusters represented normal and abnormal operating conditions respectively. With this method, it addresses the weakness of the PCA where an improved clustering method was needed. However, when plotting the clustering results, only 2-3 variables can be visualized at a time which may be difficult for the audience to fully grasp the analysis. The results therefore, highlight a need for a method to reduce the variable set to a smaller dimensional size. By plotting the cluster indexes over the dataset time series, it aided in the visualization of the performance state of the turbine blades during operation. With the integration of a moving window with an assigned threshold, the technique correctly identified 75% of past blade performance issues with hindcasted historical blade data.

The contributions of this thesis to the wind turbine research area are summarized below:

1) Introduced a simplified ranking and GIS approach in order to draw simple information about wind turbines within a wind farm such as performance metric comparisons

2) Utilized PCA to infer performance state of turbine blades whereas many pre-existing studies focused on other components in the turbine assembly
3) Applied a k-means algorithm to show its capabilities in assessing performance state of blades

2. Recommendations

The results of these studies performed mainly serve as recommendations to the development of future wind turbine blade condition monitoring systems. In-field blade condition monitoring systems are rare due to the difficultly to implement pre-existing elaborate algorithms as well as the inability to anticipate all environmental challenges a blade will operate under. That said, the information presented in this thesis provides valuable insights. While the analyses presented highlight the inner works and changing behaviours and the blade condition monitoring systems and wind turbine blades respectively, a few recommendations can be made to further enhance the completed analyses.

In general, the studies performed can be further enhanced by performing a sensitivity analysis to examine the influence of different parent variables on the derived results. That is, which input variables among those available from the SCADA will be most descriptive of the performance state of the wind turbine blades?

The simplified ranking and GIS method discussed in Chapter 2 can be improved by focusing on evaluation of the correlations between the SCADA variables and major component repairs or downtime. The participating farms did not have substantial downtime hours or repairs, and therefore, the statistical power was too low to provide clear correlation analysis. If correlations between SCADA variable rankings and undesirable outcomes, such as blade downtime, can be uncovered, a usage indicator could be developed to provide a simple, yet easily accessible tool to infer wear and inform O&M discussions.

Further future work for the PCA method presented in Chapter 3 includes the incorporation of a third principal component to the study to determine the influence on prediction accuracy. According to Figure 3.2, with three principal components, approximately 97% of the input dataset can be explained which is a significant increase from two principal components. It is noted however that this also increases the risk of additional noise in the results. A potential for inclusion of a forecasted anomaly operating index (AOI) for this dataset should also be considered. This could extend the method for prospective application in remaining useful life (RUL) predictions that are less dependent on additional specialized sensors and equipment.

The next step for the k-means algorithm performed in Chapter 4 should include an evaluation for the clustering cost to determine if there are more clusters that can detect abnormal instances for a more detailed analysis due to the high volume of SCADA parameters being used. For the cluster indexes plotted over the blades' operations, it would be interesting to examine the fluctuations more closely to understand what the cause is and to determine whether it can be used to forecast potential blade state classifications. Finally, the application of principal component analysis (PCA) to reduce the dimensionality of the input data into principal components can improve the visualization of the results (i.e. from a 7-dimensional analysis to 3-dimensional analysis). The incorporation of these methods can offer a promising research direction in a growing wind energy market.

VITA AUCTORIS

NAME:	Jones Paul Shen
PLACE OF BIRTH:	Chatham, ON
YEAR OF BIRTH:	1994
EDUCATION:	Vincent Massey Secondary School, Windsor, ON, 2012
	University of Windsor, B.A.Sc., Windsor, ON, 2016
	University of Windsor, M.A.Sc., Windsor, ON, 2018