An Integrated Prognostics Condition-Based Maintenance Approach to Reduce Premature Failure of Wind Turbines' Gearbox Components

by Ahmed Issahak

B.S. in Civil/Structural Engineering, August 2007, Addis Ababa University
M.S. in Structural Engineering, May 2012, Norwich University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

January 8, 2021

Praxis directed by

Timothy Blackburn
Professorial Lecturer of Engineering Management and Systems Engineering

The School of Engineering and Applied Science of The George Washington University certifies that Ahmed Issahak has passed the Final Examination for the degree of Doctor of Engineering as of November 12, 2020. This is the final and approved form of the Praxis.

An Integrated Prognostics Condition-Based Maintenance Approach to Reduce Premature Failure of Wind Turbines' Gearbox Components

Ahmed Issahak

Praxis Research Committee:

Amir Etemadi, Assistant Professor of Engineering and Applied Science, Praxis Director

Timothy Blackburn, Professorial Lecturer of Engineering Management and Systems Engineering, Praxis Co-Director

Aristos Dimitriou, Professorial Lecturer of Engineering Management and Systems Engineering, Committee Member.

© Copyright 2020 by Ahmed Issahak
All rights reserved

Dedication

This research is dedicated to my wife, Ifrah. Your patience, perseverance, encouragement, and most importantly, your constant support during the challenging time of balancing work, life, and school was incredible and unimaginable. Furthermore, I wish to dedicate this intense writing to my three beautiful daughters, Leyan Ahmed (6 years old), Samah & Samal Ahmed (4 years old twins). I know in the past couple years of your life, there are times when I seem mentally or emotionally absent.Taking time from your precious age you ought to enjoy with me was not easy. Yet, you indeed have made me stronger, and better than I could have ever imagined. Thank you for being patient with me. Eventhough, you are not old enough to understand the content of this writing, consider this is daddy making the world a better future for you by contributing, and promoting clean energy, and meanwhile raising the bar as a role model on the level of education you should strive for in return.

Acknowledgements

I wish to acknowledge my employer for their support and sponsorship. Thank you for being willing to provide the requested historical data; it would have been impossible to conduct the research without it. Taking this continuous leadership development opportunity and working toward the achievement of the 2050 100% carbon free goals is the best decision I have made in my professional journey.

Furthermore, I would like to express my sincere gratitude to my advisor Dr. Timothy Blackburn for his guidance, immense knowledge, and encouragement. I could not have imagined finishing this praxis without him. Thank you.

Last but not least, I would like to acknowledge Dr. Mehadi Hassen (Ph.D.) for his support in troubleshooting and helping me code the sliding window algorithm to feature engineer the raw dataset to make it readable to the machine learning tool. Thank you for your time and support in helping me succeed in my journey. Thank you.

Abstract of Praxis

An Integrated Prognostics Condition-Based Maintenance Approach to Reduce Premature Failure of Wind Turbines' Gearbox Components

This research proposed an integrated prognostics condition-based maintenance (CBM) strategy to reduce the premature sudden failure of three-bladed horizontal axis wind turbine (HAWT) gearbox. It is apparent that the maintenance of wind turbine components in the growing renewable energy industry significantly affects the levelized cost of energy (LCOE) [8]. More than half of wind turbine failures stem from control systems and electric components, but these failures have low downtimes compared to the premature failure of critical components such as blades, gearboxes and generators, which accounts for 95% of the production downtime [9].

The integrated approach predicts and reduces the premature failure and the Remaining Useful Life (RUL) distribution of the gearbox to spot upcoming failure so maintenance can proactively be scheduled. It also utilizes instantaneously fluctuating variables encompassing environmental conditions such as temperature, wind speed and direction along with the parameters from the gearbox and rotor system. The method combines the data-driven predictive model with the physics-based degradation process of the gearbox to account for the turbine components' complexity and lack of sufficient industry-wide failure data. Ultimately, this research will improve operation and maintenance (O&M) cost, performance and system reliability of wind turbines.

Table of Contents

Dedication iv

Acknowledgements v

Abstract of Praxis vi

List of Figures xi

List of Tables xiii

List of Symbols xiv

List of Acronyms xv

Glossary of Terms xiii

Chapter 1—Introduction 1

1.1 Background 1

1.2 Research Motivation 3

1.3 Problem Statement 3

1.4 Thesis Statement 4

1.5 Research Objectives 5

1.6 Research Questions and Hypotheses 5

1.7 Scope of Research 6

1.8 Research Limitations 6

1.9 Organization of Praxis 6

Chapter 2—Literature Review 8

2.1 Introduction to Wind Turbine Component Systems 8

2.1.1 Rotor system 9

2.1.2 Drivetrain system……12

2.2 Reliability of the Wind Turbine Gearbox 15

2.3 Failure Modes, Physics of Fracture, and Gear-Crack Propagation 19

2.3.1 Failure modes……..19

2.3.2 Physics of fracture and gear-crack propagation……..20

2.3.2.1 Westergaard's solution..……21

2.3.2.2 Irwin's near crack tip solution……...22

2.3.2.3 Fatigue crack propagation…….24

2.4 Condition-Based Monitoring and Maintenance Systems.…….…….…25

2.4.1 Vibration-based monitoring……..27

2.4.2 Oil-based monitoring……28

2.5 Summary and Conclusion……..29

Chapter 3—Methodology 31

3.1 Introduction 31

3.2 Integrated prognostics data-driven method 30

3.2.1 Data description, collection and trend…….……...……33

3.2.2 Physics based analysis……35

3.2.3 Data Diagnostics and Imputation……...38

3.2.4 Prepare views of dataset…….……40

3.2.5 Proposed data analysis method……...40

3.2.5.1 Regression algorithms…….41

3.2.5.2 Regression model validation procedure……43

3.2.5.3 Paired T-test…….43

Chapter 4—Results 45

4.1 Introduction 45

4.2 Input variable trend analysis 45

4.2.1 Environmental factors……45

4.2.2 Rotor system factors……..48

4.2.3 Gearbox factors……..51

4.3 Stress intensity and finite element analysis results……..54

4.4 Regression machine learning steps and analysis results……..57

4.4.1 Analyzed data…….57

4.4.2 Views of the dataset……...58

4.4.3 Regression algorithms result evaluation……59

4.5 Model analysis results and interpretation…….60

4.5.1 Reduced model output……...61

4.5.2 Model interpreation…..……...62

4.6 Test validation and results…….……63

4.7 Proposed approaches and maintenance actions…….67

Chapter 5—Discussion, Conclusions, and Recommendations..... 69

5.1 Discussion..... 69

5.2 Conclusions..... 69

5.3 Contributions to Body of Knowledge..... 70

5.4 Recommendations for Future Research..... 71

References..... 73

Appendix A..... 81

Appendix B……...83

Appendix C……...87

Appendix D……...91

Appendix E……...96

List of Figures

Figure 1. A typical modern wind turbine. 9

Figure 2. A typical modern HAWT blade with aerofoil profiles. 11

Figure 3. A typical modern rotor blade cross-section……11

Figure 4. Configuration of modular drivetrain wind turbine components……..13

Figure 5. 2-Megawatt WT gearbox……..15

Figure 6. Failure intensity curve—bathtub curve……...17

Figure 7. WT reliability expectations and vision……18

Figure 8. WT Gearbox Reliability Database: Damage distribution……19

Figure 9. Three types of loading on a cracked body..……20

Figure 10. Stress plot along the crack plane..……22

Figure 11. Stress field in an infinite plate……..23

Figure 12. Influence of CBM strategy on asset condition…….26

Figure 13. Integrated prognostics CBM model & flowchart…….32

Figure 14. Determination of stress intensity factor (SIF)……..36

Figure 15. Moment and force components in cartesian coordinates…….37

Figure 16. Hypothetical location of the crack in the tooth roots……...39

Figure 17. Element representation of the observed crack……..39

Figure 18. Raw data trend plot - Wind speed (X_1) ……..……..46

Figure 19. Raw data trend plot - Ambient temperature (X_2) ……..…….47

Figure 20. Raw data trend plot - Wind Direction (X_3)……...…….48

Figure 21. Raw data trend plot - Turbine Nacelle Position (X_4)……..49

Figure 22. Raw data trend plot - Rotor Sys. Blade Pitch Angle (X_5)……..…….49

Figure 23. Raw data trend plot - Rotor Sys. RPM (Torque) (X_6)…….……..49

Figure 24. Raw data trend plot - Blade A Pitch Angle (X_7)……..……..50

Figure 25. Raw data trend plot - Blade B Pitch Angle (X_8)……..……..……..50

Figure 26. Raw data trend plot - Blade C Pitch Angle (X_9)…….…….…..51

Figure 27. Raw data trend plot - Nacelle Temperature (X_{10})……..……..……..51

Figure 28. Raw data trend plot - Gear Bearing Temperature (X_{11})……...……..52

Figure 29. Raw data trend plot - Gear Oil Temperature (X_{12})……...52

Figure 30. Raw data trend plot - Tower Acc. X Dir. (X_{13})……..……..53

Figure 31. Raw data trend plot - Tower Acc. Y Dir. (X_{14})……..……..54

Figure 32. Raw data trend plot - Turbine Generation. (X_{16})……...……..54

Figure 33. Raw data trend plot - Static & Dynamic load distribution ……..55

Figure 34. Raw data trend plot - Stress Intensity Factor. (X_{15})……...……..56

Figure 35. Historgram of difference…….....……...66

Figure 36. Proposed Maintenance Concept with multi-level Actions……...68

List of Tables

Table 1. Gear material property ..... …35

Table 2. Sample data – O&M failure tracker database ..... .39

Table 3. Response and predictor variables……41

Table 4. Rotor System Factors: Summary ……48

Table 5. Gearbox Factors: Summary ……53

Table 6. Stress intensity: Summary ……..……56

Table 7. Input variables of unreduced model ……...58

Table 8. Linear regression algorithm: Summary……..…….59

Table 9. Support Vector regression algorithm: Summary….……...…….59

Table 10. Input variable significance distribution……...…….……60

Table 11. Variable significance……...…….……..….61

Table 12. WEKA Linear Regression Analysis Result……..…….61

Table 13. Paired two sample t-test results ……..…….……65

List of Symbols

L_{nm}	Bearing life	[*million revolutions*]
a_1	Life modification factor for reliability	
a_{ISO}	Life modification factor	
L_{10}	Basic rating life	[*million revolutions*]
f	Dynamic factor	
e_c	Contamination factor	
c_u	Fatigue load limit	[*newton, N*]
P, P_r	Dynamic equivalent load	[*newton, N*]
K	Viscosity ratio	
R	Reliability	
SSR	Strict series components	
n	Number of components	
T	Number of years	
i , j, t	Index for components of years	
F	Failure probability	
σ_{xx}	Normal stress in x-x plane	[*Mega pascal, MPa*]
σ_{yy}	Normal stress in y-y plane	[*Mega pascal, MPa*]
τ_{xy}	Shear stresses in xy plane	[*Mega pascal, MPa*]

List of Acronyms

O&M	Operation and maintenance
WT	Wind turbine
LCOE	Levelized cost of energy
kWh	kilo-Watt hour
IRENA	International Renewable Energy Agency
OEM	Original equipment manufacturer
HAWT	Horizontal axis wind turbine
CBM	Condition-based maintenance
MW	Mega watts
RUL	Remaining useful life
TTF	Time to failure
NREL	National renewable energy laboratory
ISO	International Organization for Standardization
FEA	Finite element analysis
PLP	Power law process
HPP	Homogeneous Poisson process
SIF	Stress intensity factor

Glossary of Terms

S-N Curves	log-log plots with allowable or nominal stress range on the vertical axis and allowable number of cycle on the horizontal axis.
OSIsoft PI	OSIsoft is an organization that provide simplicity , from asset and process intelligence to operational intelligence with the PI system.

Chapter 1—Introduction

1.1 Background

Wind energy, a growing source of renewable energy, has been developed aggressively over the past few decades, and continues being deployed to help realize the goal of achieving carbon-free emissions by the year 2050. However, the increasing number of deployments and the constant changes in turbine technology have created concerns about wind power's reliability and availability. The operation and maintenance (O&M) cost of wind turbines (WTs) is one of the driving factors in determining the levelized cost of energy (LCOE) for wind power. LCOE, often refer to as cost of energy per kilowatt-hour (kWh), varies depending on the wind resource and project costs. Within the International Renewable Energy Agency (IRENA) cost analysis, in 2010 alone, O&M cost accounted between 11 – 30% of the project's LCOE for most onshore wind projects [7]. The need for increasing turbine rotor diameter, size, height and the everchanging weather conditions, pose significant challenges that most of the wind turbines' original equipment manufacturers (OEMs) are facing as they work to overcome the large number of turbine component failures and frequent downtimes [7].

Horizontal axis wind turbine (HAWT) components comprises of the hub, support structure, rotor blades, mechanical brake, drive train, gearbox, generator, yaw system, sensors, hydraulic system, electrical system, and control system. The critical turbine components include the support structure, rotor blades, drive train, gearbox, and

generator. Even though more than 50% of turbine components failures are due to electrical and control systems, an overall survey showed very low downtimes for these issues compared to the failures of the critical components. Failure of turbine blades, gearboxes and generators accounts for 95% of the production downtime, of which 80% of critical component failure is due to failure of the turbine gearbox . This research therefore focuses on gearbox failure, as it causes the greatest downtime in most of the wind farms in United States [10].

Three categories of maintenance actions are utilized in the wind industry to mitigate all types of system and sub-system components failures, including gearbox failure: failure-based or corrective maintenance, time-based maintenance, and condition-based maintenance (CBM). Failure-based maintenance action is scheduled after a system component fails. It tends to be the costliest strategy because it causes prolonged system downtime, but currently it is the most widely used type of maintenance category. Time-based maintenance action is scheduled prior to the failure of a system component to prevent upcoming failure by performing preventive maintenance on a regular time basis (e.g., once per quarter or once per year) and conducting regular age-based maintenance. Frequent time-based maintenance would typically incur a high cost but lower the chances of component failure. Finally, the CBM strategy is scheduled according to the health condition of the turbine system, which is analyzed based on data on collected from a condition monitoring system.

CBM constitutes three broad areas: a) diagnostics, which involves detecting,

isolating, and identifying faults when they occur, b) Prognostics, which involves predicting future failures to avoid unnecessary maintenance, and c) maintenance decision-making, which entails scheduling maintenance based on detected early faults and predicted remaining useful life. In most industries, CBM serves as a highly data-driven method; however, due to the data limitations of the wind industry, a physics-based degradation process will be integrated into the prognostics method applied in this research. This approach will help overcome the data limitations and the complexity of the turbine gearbox physical model.

1.2 Research Motivation

The increased demand of wind energy and significant room for improvement in its reliability are the key elements motivating this research in the optimization of maintenance actions to mitigate turbine gearbox component failures.

The primary maintenance the wind industry has adopted to date includes failure-based or corrective maintenance and the time-based maintenance. A few newer wind farms with latest sensor technology are currently rolling over to CBM. However, there is an opportunity for saving O&M costs by implementing preventive maintenance that integrates both historical failure data and physics-based degradation model of turbine gearbox to better predict and reduce the premature failure of turbine components.

1.3 Problem Statement

Premature wear of wind turbine gearbox components has resulted in wind turbines failures, which leads to losses of millions of dollars in revenue.

The overall life of a wind turbine components ranges from 20 – 25 years.

However, in some wind farms, 50% of the gearbox components prematurely wear out in the span of a few years, causing turbine failure and production downtime. The gearbox is one of the costliest elements of the turbine drivetrain, requiring replacement from original equipment manufacturer (OEM), transportation, installation and commissioning could easily take up to six months or more. Depending on the type of turbine production, the financial loss could easily be in the millions. For instance, for two- megawatts (MW) turbine, production loss can approximately run up to $12,500 per week. If ten gearboxes failed per year at a different windfarm, it could easily cost around $30 million dollars in production loss alone without including installation, labor, and material costs [3].

1.4 Thesis Statement

An Integrated prognostics condition-based maintenance (CBM) solution will be implemented to predict and assess remaining useful life (RUL) distribution to avoid premature future gearbox failure and unnecessary downtime.

Prognostics CBM works to predict the time to failure (TTF) of the gearbox component by utilizing historical failure data derived from condition monitoring data and integrate it with a fault degradation model from a finite element analysis of the gearbox component to assess RUL distribution in order to minimize unwanted turbine downtime. The gearbox component is under time-varying external load from factor such as wind speed, temperature, and the yawing of the rotor system with the wind direction. The finite element physics-based degradation model result add a layer to the model utilized in this study, considering the instantaneously time-varying load condition to predict the gear tooth crack propagation.

1.5 Research Objectives

This praxis focuses on outlining a predictive model using an integrated prognostics approach for gearbox WT components and employing a CBM action to enhance the O&M of the wind power systems. The methods developed will help significantly reduce the failure of gearboxes, increase the reliability of wind turbines, promote a clean and sustainable energy future worldwide by lowering the LCOE.

1.6 Research Questions and Hypotheses

1.6.1 Research Questions

RQ1: How do environmental factors influence the failure of turbine gearbox components?

RQ2: Which variables are most significantly influencing the turbine gearbox components failure?

RQ3: How do rotor systems affect the failure of the gearbox components?

RQ4: What maintenance actions are needed to mitigate premature gearbox failure?

1.6.2 Hypotheses

H1: Time to failure (TTF) of the gearbox turbine component is affected by varying environmental conditions, rotor systems, and gearbox factors.

H2: If rotor systems and gearbox integrated CBM actions are implemented, it will reduce premature failure of wind turbine gearboxes.

1.7 Scope of Research

The scope of this research encompasses predicting the TTF of future WT gearbox components. It supplements prognostics CBM approach with physics-based degradation model to generate a predictive model and maintenance actions. The key predictors include environmental, rotor system and gearbox factors predicting the gearbox time to failure (TTF). Statistical and finite element analysis tools such as Minitab, WEKA, ABAQUS and ANSYS or their equivalent will be implemented. Based on the result obtained, maintenance actions will be suggested to mitigate future failures.

1.8 Research Limitations

The National Renewable Energy Laboratory (NREL) is being challenged by receiving a large volume of data from various wind farms. The limitation of this research include consideration of only wind farms in the Midwest climatic region, as well as consideration of the gearbox OEM to certain suppliers and models.

1.9 Organization of Praxis

This praxis is prepared in partial satisfaction of the requirements for doctoral degree of engineering at the George Washington University. The remainder of this praxis is organized as follows:

Chapter 2 presents a detailed literature reviews on WT component systems with the focus on gearboxes, reliability engineering, CBM prognostics approaches pertaining to wind energy systems. This chapter is also devotes to presenting literatures on statistical data analysis methodology and physical model of gear tooth meshing, fractures, stress intensity, and finite element analysis.

Chapter 3 presents the implementation of the research methodology, which entails a quantitative research approach that mainly employs regression or time series analysis using Minitab or an open source machine learning tool such as WEKA.

Chapter 4 presents the finite element analysis of the gearbox, along with implementation and validation of the statistical model and experimental results. It also details some cost saving on opportunistic maintenance methods for wind farms systems considering multiple maintenance approaches.

Chapter 5 concludes the approach discussed, generalize its applicability to other WTs components and to the body of knowledge in other related industry. It also suggests several futures directions for study and potential improvements that can be accomplished.

Chapter 2—Literature Review

This chapter embarks on a review of the literature pertaining to the configuration of wind turbine (WT) component systems and the reliability of turbine component systems, with a focus on the gearbox, modes of failure, physics of fracture, gear-crack propagation, condition-based maintenance (CBM), and monitoring systems.

Even though this study focuses on gearbox component failures, understanding how the whole WT system works is essential to comprehending the covariates of the critical components and to seeing whether the data analysis results during prognostics make sense. Gearboxes in WTs transform low-speed, high-torque loads exerted from the rotor system to a higher speed required by the generator in order to convert the mechanical power to electricity, and they are often considered the heart of the drivetrain system. The following section introduces and outlines the configuration of the typical 2-megawatt WT component system.

2.1 Introduction to Wind Turbine Component Systems

A wind turbine, unlike a home fan that uses electricity to generate air, is a machine that transforms the energy in the wind into electricity with the help of a rotor and drivetrain system. In today's world, the most common WT is the horizontal axis wind turbine (HAWT), in which the axis of rotation is parallel to the ground. Typical configuration and components of HAWT comprise the foundation; the support-tower leg; the rotor system, which constitutes the blades and the supporting hub; and the drivetrain, which includes the shafts, gearbox, coupling, mechanical brake, control system, and generator. The drivetrain components are located inside the nacelle and main frame,

including the WT housing, bedplate, and yaw system. To better understand how the drivetrain (including the gearbox subcomponent) works, it is pertinent to understand the rotor system, including the composition of the blade engineering. The section below introduces how the rotor system of a typical 2-MW wind turbine configuration works.[13]

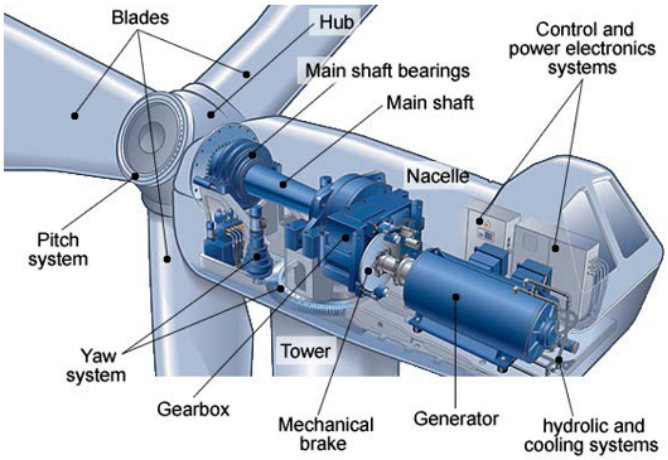


Figure 1 – A typical modern WT. [14]

2.1.1 Rotor System

The rotor system consists of the blades and the supporting hub of the WT, which are considered the turbine's most important components, driving performance, power generation, and cost. Most HAWTs have upwind rotors with three blades, which is the type of rotor blade considered in this research. In the past, downwind rotors and a few designs with one- and two-blade turbines were also built, but they are no longer in production due to their lower efficiencies and performance in driving the levelized cost of energy.[15]

Rotor blades are typically 80–140 meters in diameter and are meant to be slender, light, thin-walled, and hollow—and thus, flexible. Their aerodynamic design requires the blade material to be as light and thin as possible, whereas structural considerations might require it to be strengthened in certain regions due to large loads from the environment in the form of elements such as wind, ice, and temperature. Most rotor blades are therefore constructed from fiber-reinforced polymers (FRP) made of continuous glass-fiber composites to meet the needs for high strength, high stiffness-to-density ratio, and high strength-to-density ratio. As blades become longer and slenderer, carbon fibers are more cost-efficient for large WT structures [15].

A WT blade cross-section and structural component are divided into load-carrying parts and geometrically optimized for an aerodynamic shape. They are predominantly loaded in the flapwise and edgewise direction (refer to figures 2 and 3 for cross-section and direction). Rotor blades become stressed mainly due to aerodynamic wind contact, gravitational and centrifugal forces, ice buildup, ultraviolet radiation, lightning strikes, gyroscopic force, and other operational loads. Cracks arising on the FRP material from fatigue and other contributing loads weaken the fiber bond and delaminate the girder holding the glass or carbon fibers, leading to failures of the blades, supporting tower and other subsequent turbine components. [17]

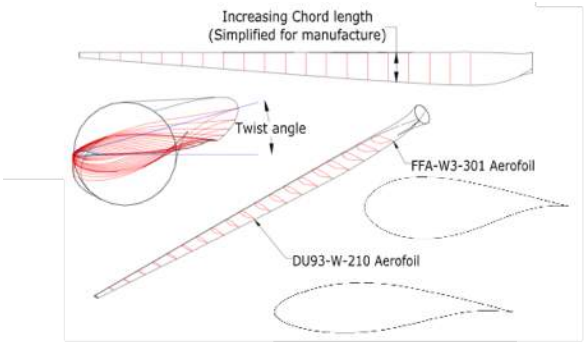


Figure 2 – A typical modern HAWT blade with aerofoil profiles. [15]

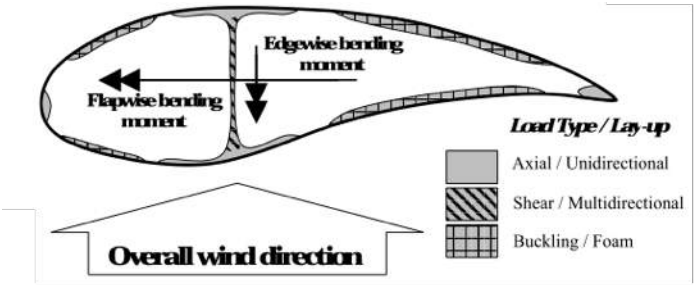


Figure 3 – A typical modern rotor blade cross-section. [21]

For reasons of efficiency, noise, shadow flickers, control, and aesthetics, the modern wind turbine is characterized by a three-rotor-blade design, with the use of a yaw-and-pitch control system, to increase its capacity to survive and operate under varying load conditions. A modern rotor blade is equipped with pitch control system of its own. Yaw systems are a very important component of WT that plays a significant role in increasing power efficiency. As the wind direction changes, yaw systems sometimes cease and start frequently to keep the turbine aligned with the wind direction, increasing the wind angle of attack. [15] These operational conditions and yaw error create rapid-rotor torque fluctuations and resistance torque variations in response to changes in the yaw angle and wind speed, which generates negative load fluctuations in the yaw system. These load variations result in further speed fluctuations, which affect the rotation of low-

speed shaft, vibrations of the blades and tower, while resonating through the whole drivetrain including the gearbox components compromising the gear tooth—potentially causing them to fracture and crack.[31]

The blade pitch control is another critical subsystem within the WT control system situated within the hub. It serves the important function of preventing operation of the blade to regulate rotor speed during emergencies and mitigating blade-root flapwise bending loads for above-rated wind speed. The WT main controller calculates the required pitch angle from wind speed, generator speed, and power production. If the actual angle is not the same as the set point, the pitch system will direct power to the electric motor or fluid to the cylinder in order to make the actuator move the blade to the required angle. [21] The rotor system feeds high torque and low speed through the low-speed shaft to engage the drivetrain system, which is discussed briefly in the next section.

2.1.2 Drivetrain System

The drivetrain system of a WT comprises the main bearing, low-speed shaft (LSS), gearbox, brake system, high-speed shaft (HSS), and generator (see Figure 4.). It mainly converts the low-speed, high-torque rotation from the turbine rotor system (blades and hub assembly) into electricity.

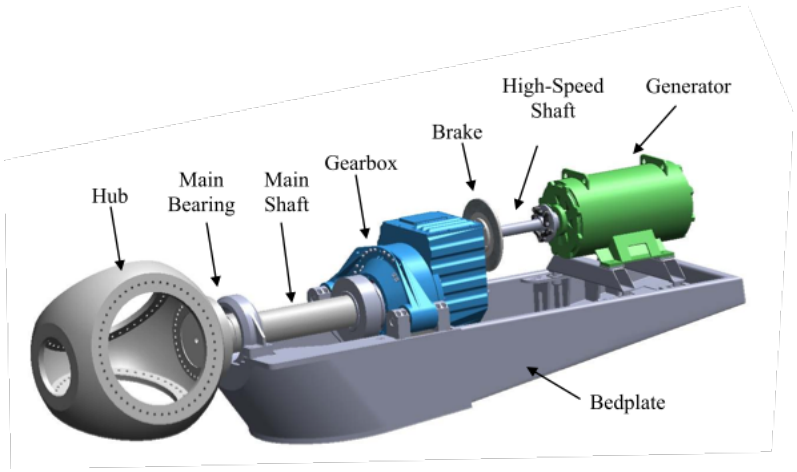


Figure 4 – Configuration of modular drivetrain wind turbine components. [14]

The main bearing is typically a spherical roller mounted in the bearing housing bolted to the mainframe of the hub. Like other WT components, most main bearings fail to achieve the design life of 20 years. According to the study completed by the National Renewable Energy Laboratory (NREL) of the United States, the health of the main bearing affects the reliability of the wind turbine drivetrain.[26] Unsoundness of the main bearing can even cause failure of the gearbox as a whole. Poor lubrication, wear, tear, pitting, and deformation of the rolling elements may also cause bearing failures. [26]

The main shaft, also known as the LSS, drives the low speed, high torque to the gearbox component. The finite element analysis (FEA) results found by the University of Beijing School of Mechanical Engineering have demonstrated that a severe stress concentration leads to the formation of initial cracks at the variable inner diameters of the main shaft. [27] It is important to note that a WT drivetrain is a series connection system wherein any malfunction of the rotor system, main bearing, or main shaft could lead to either lower production or failure of the subsequent critical component such as the gearbox, HSS, or generator.

For the drivetrain to operate the generator directly with the power from the main shaft, and to connect a 50 Hz alternating-current three-phase grid, an extremely high-speed turbine with between 1,000 and 3,000 revolutions per minute (rpm) will be required to maintain the generator's rotational speed. However, for a directly driven generator, the mass of the rotor must handle the torque and moment, which will make the drivetrain expensive and complex from a structural integrity perspective. [30] The typical 2-MW WT gearbox converts the low speed, high torque of 20 rpm to around 1,440 rpm. It requires more than one gear stage to achieve an overall gear ratio of up to 1:12 to outperform a parallel shaft gear over the planetary stage.[25] Therefore, the implementation of the WT gearbox to the drivetrain system is to ensure the generator meets its generation requirement.

The generator transforms mechanical energy into electrical energy. The kinetic energy from the rotor system and wind transforms into a rotational energy using the gearbox, and high rotational speed generates mechanical energy from the gearbox, which is then converted to electrical energy by the generator once the required revolutions per minute are met.[42] Depending on the type of drivetrain it has, the generator produces either alternating current or direct current, and the range of power output varies depending on the type and size of the WT.[42] Its rating mainly depends on the size of the turbine rotor blades; larger blades and a larger tower produce more power but entail more system complexity. Efficiency from both the rotor and drivetrain system is required to maintain the reliability of wind turbines in general, which drives the levelized cost of energy. As numerous premature failures of gearboxes occur, the reliability of the entire WT system is compromised.[46] Thus, this research addresses the reliability of a WT in

terms of the gearbox component and how it is modeled.

2.2 Reliability of the Wind Turbine Gearbox

Modern WTs are becoming larger and more complex aerodynamic mechanical and electrical machines that include sophisticated control systems. An increasing number of wind turbines are being added to the U.S. grid system. In some states, they are overwhelmingly becoming a part of permanent grid networks, and they therefore affect the overall system's performance and reliability.[12]The dependability of a typical 2 MW gearbox that includes the first planetary stage and the following two parallel stages (see Figure 5) has been evaluated for a number of cases such as changes in oil cleanliness in terms of particle content and oil viscosity at a specified nominal load, oil temperature, and number of full-load hours and vibration, assuming a 20-year design life cycle [12].

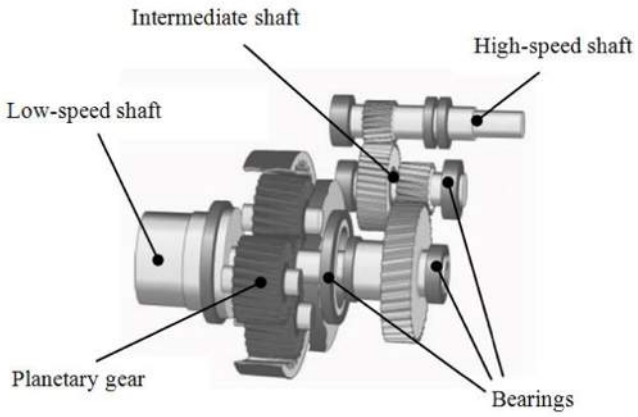


Figure 5 – 2-MW WT gearbox. [22]

Since the inception of larger-geared WTs, several tools or methods have been developed to evaluate gearbox reliability. The "strict series reliability" approach, which is becoming more and more common, has been introduced. It addresses the design life of

the gearbox subsystem, primarily in regard to the gears and bearings. Equation 1 evaluates the component according to bearing life standard as per the International Organization for Standardization's (ISO) 281:2007 standard, which calculates both the basic dynamic load and rating life. [34]

$$L_{nm} = a_1 a_{ISO} L_{10},\ a_{ISO} = f\left(\frac{ecCu}{P}L10, K\right) \quad (1)$$

A component's reliability in reaching the desired lifespan is also determined by equation 2, as follows:

$$R = e^{\left(\ln(100) - \left(\frac{a1}{3.85}\right)^{\frac{5}{3}}\right)} \quad (2)$$

A system's reliability in using a strict series approach wherein every gear and bearing is essential for the gearbox is expressed by equation 3, which calculates the reliability, *R*, for each service year by adjusting the a_1 factor with *t*/*T*, where *t* is the actual service years and *T* is the calculated full lifetime. [34]

$$R_{SSR,j} = \prod_{1,1}^{n,T} R_{i,j} \quad (3)$$

The model presented above is objective and considers only failure modes that are included in current ISO 281 bearing lifetime calculations. Other failure modes such as cracking, macropitting, scuffing, corroding, becoming misaligned, and having poor oil quality are also covered in other studies. [34] The method using equations 1–3 is described in detail in [34], along with few examples. [34]

WT subassemblies are considered repairable, and thus, Travner, Xiang, and Spinato use the turbine and subcomponent reliability results, plus a mathematical model such as power law process (PLP) or homogeneous Poisson process (HPP), to interpret the data.[12]

PLP is the model most typically used in reliability analysis of complex repairable system components. The intensity function, $\lambda(t)$, describes the failure rate of the subassemblies of the WT, and is calculated as:

$$\lambda(t) = \frac{\beta}{\theta} * \left(\frac{t}{\theta}\right)^{\beta-1} \quad (4)$$

where β is a parameter that describes the shape of the intensity function, θ is the scale parameter and has dimensions of time, and $\theta > 0$ for $t \geq 0$. For early failures, $\beta < 1$; for constant failure $\beta = 1$; and for deterioration, $\beta > 1$. When $\beta = 1$, equation 4 will reduce to $\lambda(t) = \frac{1}{\theta}$ and the PLP model will become the HPP—a special case of PLP for which time between failures (TBFs) is independent. Figure 6 below shows the typical failure-intensity curve, referred to as the bathtub curve, whose three regions are described by Equation 4. [12]

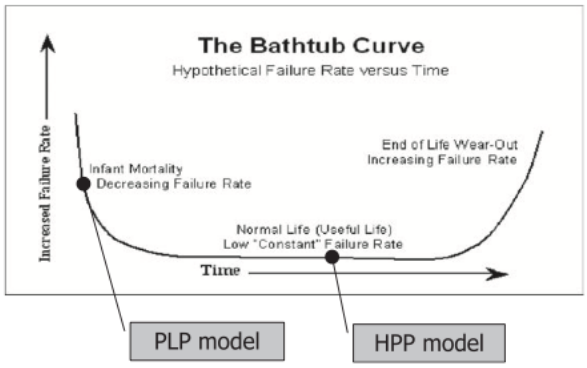


Figure 6 – Failure intensity curve—bathtub curve. [12]

Thus, θ becomes the mean time between failures (MTBF) of the turbine, and the likelihood estimate (MLE) of θ is

$$\theta = \frac{1}{\lambda} = \frac{\sum_{i=1}^{I} Ti}{\sum_{i=0}^{I} ni} \quad (5)$$

A more traditional life data analysis such as the Weibull reliability measure evaluates these components and their drivetrain. In reality wind turbine gearbox failure is accompanied by a catastrophic gear tooth cracks and breaks, which is considered a non-repairable component; thus, MTBF is interchangeable with mean time to failure (MTTF). MTTF describes the expected time to failure for a non-repairable system or component. Assuming Weibull distribution, the MTTF can be mathematically calculated by $MTTF = \int_0^\infty tf(t)dt = \int_0^\infty R(t)dt$. For the purpose of this research, historical time to failure (TTF) of each failed gearbox turbine component is used as a response variable. Figure 7 illustrates a comparison between the expectation of failure rate and today's reality of the failure rate in general. The industry seeks to maintain the reliability of the drivetrain in a bathtub-shaped curve, shown in Figure 7, for the 20-year design life cycle.[53] Understanding the failure modes and studying the physical nature of the gearbox components will aid in better predicting a more robust model and lead to a more reliable system. The next section introduces physics-based fracture mechanics and crack propagation to assess the remaining useful life of a cracked component.Understanding it's failure modes and fracture of the material will better explain the crack propagation.

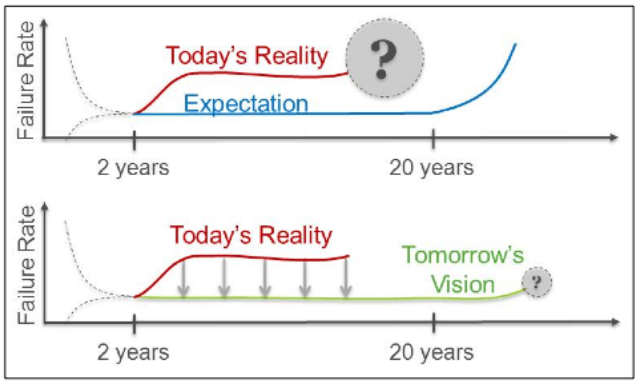


Figure 7 – WT reliability expectations and vision. [53]

2.3 Failure Modes, Physics of Fracture, and Gear-Crack Propagation

2.3.1 Failure modes

Well-proven approaches such as failure modes, effects, and criticality analysis (FMECA) and Six Sigma are used to ensure quality in WT product design and development. Gearboxes can fail in different ways—a 2015 report from the NREL Wind Turbine Gearbox Reliability Database shows that 76% of gearbox failures stem from bearing issues, 17% are due to gears, and 7% occur for other reasons.[46] Most historical failures observations have also shown that both bearing and gear failures are concentrated in the parallel stages, in which top failure mode was recorded in the high-speed shaft (HSS) or intermediate-speed shaft (IMS) bearing axial cracks.[46] Furthermore, the 2018 failure database shows that high- and intermediate-speed-stage bearings contribute to 62% of total drivetrain failures and that planet-bearing failures lead to the highest O&M cost. [46]

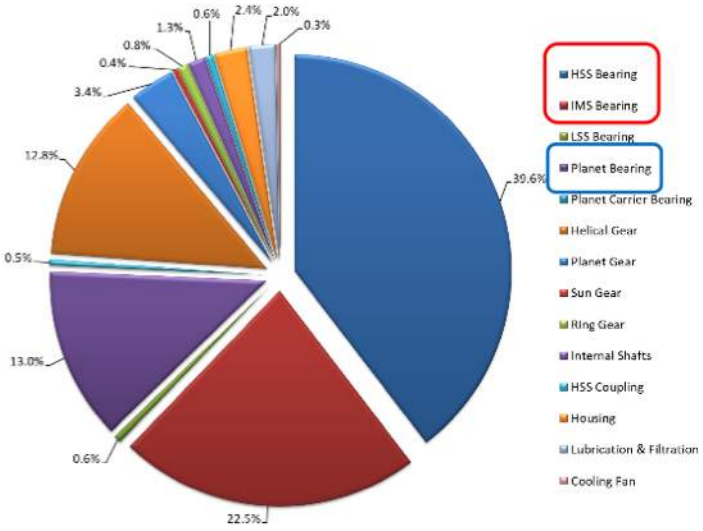


Figure 8 – Wind Turbine Gearbox Reliability Database:

Damage distribution. [36]

There are many types of WT gear (e.g., spur gear, helical gear, bevel gear). Spur gear, the most common type, is easy to design compared to a cylindrically shaped gear, which has teeth that are parallel to the axis of the rotating shaft. In this research, we therefore focus more on the spur gear. WT gearboxes' most common structural faults stem from the gear tooth and bearing damage accompanied by broken shafts. Although bearing failures tend to be the root cause of the majority of WT gearbox failures, they occur mainly due to poor lubrication, wear, tear, micropitting, deformation of the rolling mechanism, scuffing, false brinelling, and corrosion. The first mode (Mode I) is the opening mode, which has a tensile stress that aligns with the plane of the crack. The second mode (Mode II) is sliding, shear stress that normally acts parallel to the plane of the crack and perpendicular to the crack front. The third mode (Mode III) of failure is the tearing mode, which involves continuation of shear stress from the second mode, but parallel to both the plane of the crack and the front of the crack (see Fig. 9 below). [48]

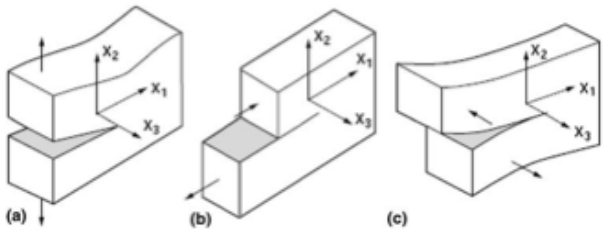


Figure 9 – Three types of loading on a cracked body:

(a) Mode I, (b) Mode II, and (c) Mode III. [48]

2.3.2 Physics of fracture and gear-crack propagation

The advancement of linear fracture mechanics and material science in general for

predicting the strength and life of cracked structures has increased the accuracy of prediction in failing components.[47] In reality, damages accumulate due to an irreversible change that takes place in the microstructure of a component exerted by certain external loading or environmental conditions. The sad fact is that these damages may not be detected before failure occurs. Cracks usually occur to the gearbox tooth root, where maximum bending stress is anticipated due to the varying cyclic loading.[48] A WT system with a gearbox tooth crack may be able to run for a while without the issue being noticed and causing downtime. The accumulated damage to the system is incremental, gradually growing with increased exposure and the amount of load cycles exerted. If the physics of fracture on a WT gearbox is studied well enough, then predicting the expected gearbox component's lifespan could be possible even with an instantaneously varying load and complex system.[47] The fracture mechanics portion of this research deals with the study of how a WT gearbox crack in a structure propagates under the given applied loading. This research basically correlates the numerical crack propagation using a fracture parameter called *stress intensity factor* (SIF) with the historical failure data and experimental results. The crack growth for each type of failure mode is estimated using SIF in the cracked region (refer to Irwin's expression below). For failure modes I, II, and III shown in Fig. 9, the corresponding SIFs K_I, $K_{II,}$ and K_{III} are designated throughout this literature. The SIF describes the stress state at a crack, and Westergaard and Irwin's solutions are introduced to establish failure criteria due to fracture and determine the expression for SIF.[48]

2.3.2.1 Westergaard's solution

Westergaard found an airy stress function of complex numbers that is the solution

for the stress field in an infinite plate containing a crack. The complete set of equations for the stress field is as follows:

$$\sigma_{xx} = \mathrm{Re}\, Z - y\, \mathrm{Im}\, Z' \qquad (6)$$

$$\sigma_{yy} = \mathrm{Re}\, Z - y\, \mathrm{Im}\, Z' \qquad (7)$$

$$\tau_{xy} = -\, y\, \mathrm{Re}\, Z \qquad (8)$$

where Z and its derivative, Z', are given by $Z(z) = \frac{\sigma \infty}{\sqrt{1-\left(\frac{a}{z}\right)^2}}$ and $Z'(z) = \frac{-\sigma \infty a^2}{z^3\left[1-\left(\frac{a}{z}\right)^2\right]^{3/2}}$,

a is a crack's length, and z equals $x + iy$. At $y = 0$, the equation is shown as $\sigma_{yy} = \frac{\sigma \infty}{\sqrt{1-\left(\frac{a}{x}\right)^2}}$

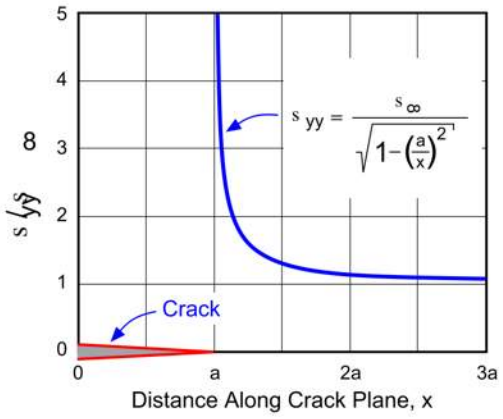


Figure 10 – Stress plot along the crack plane. [29]

As the plot above in Figure 10 shows, stress dropped quickly from infinity at the crack tip, $x = a$, to the far field value of σ_∞. Note that this is along the crack plane, $y = 0$.

2.3.2.2 Irwin's near crack tip solution

Irwin's solution is the simplified form of Westergaard's in the area immediately

surrounding the crack tip. Irwin substitutes the z value as the sum of two terms (the real and complex portion), $z = a + re^{i\theta}$, which results in $Z(z) = \frac{\sigma\infty}{\sqrt{1-\left(\frac{a}{a+re^{i\theta}}\right)^2}}$, $Re\ Z = \sigma_\infty \sqrt{\frac{a}{2r}}\cos\frac{\theta}{2}$ and $Z'(z) = \frac{\sigma\infty}{2r}\sqrt{\frac{a}{2r}}(\cos\frac{3\theta}{2} - i\sin\frac{3\theta}{2})$. [28]

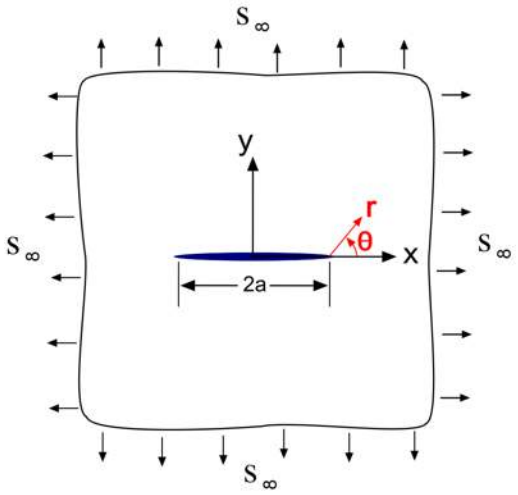


Figure 11 – Stress field in an infinite plate. [28]

The simplified version of Irwin's expressions for stress components is given by:

$$\sigma_{xx} = \frac{\sigma\infty\sqrt{\pi a}}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) \quad (9)$$

$$\sigma_{yy} = \frac{\sigma\infty\sqrt{\pi a}}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\left(1 + \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right) \quad (10)$$

$$\tau_{xy} = \frac{\sigma\infty\sqrt{\pi a}}{\sqrt{2\pi r}}\cos\frac{\theta}{2}\sin\frac{\theta}{2}\cos\frac{3\theta}{2} \quad (11)$$

Studies show that the difference between Westergaard's exact solution and Irwin's approximate solution is that Irwin's solution is more accurate at the crack tip, which is important in predicting how quickly the crack will grow, its direction, and its

severity or failure.[28] As you may have noticed in Irwin's three stress equations above, they all have the same expression, $\sigma_\infty \sqrt{\pi a}$, which means the combination of σ_∞ and a completely describes the severity of the stress state at the crack tip. This term is referred to as *stress intensity factor* (SIF) by Irwin, designated by $K = \sigma_\infty \sqrt{\pi a}$. It is also notable that all three expressions have a $\sqrt{r}$ in the denominator, which implies that the stress is dependent on the distance from the crack tip, and the remaining expression contains a function of θ alone. It is clear that if the value of K is known, it is possible to solve for all stress components as a function of θ and r. [28] The crack typically grows in the direction normal to the maximum tangential tensile stress from the crack tip, and the angle of propagation, θ, is given as:

$$\theta = 2 \tan^{-1} \left[\frac{1}{4} \left(\frac{kI}{KII} \mp \sqrt{\left(\frac{KI}{KII}\right)^2 + 8} \right) \right] \qquad (12)$$

2.3.2.3 Fatigue crack propagation

The crack propagation direction and the fatigue life of any mechanical component that exhibits a crack as a failure mode, including a WT gearbox, undergoes three stages: crack initiation, crack propagation, and crack acceleration. Most studies completed thus far have focused more on prediction of crack propagation stage to prevent premature and unprecedented catastrophic failures. The most famous crack propagation model is known as Paris' law. It mainly describes the linear correlation between the crack propagation with the SIF in a log-log scale. Equation (13) below describes *Paris and Erdogan's* lognormal and crack propagation model with the assumption of stationary load process, which is not the case for a WT gearbox system: [52]

$$\frac{da}{dN} = C(\Delta k)^n \quad (13)$$

where C and n are material constants and Δk is the range of stress intensity factor. In reality, a WT gearbox component experiences time-varying instantaneous loads from the environment through the drivetrain system, which makes it a non-stationary load. However, due to the lack of literature in this area specific to WT systems, some literature follows Paris' law to fit for the crack growth process under variable loading conditions. The current study ignores the random time-varying feature and assumes the random loading and crack growth rate as random variables.[52] *Palmgren-Miner's rule* was also reviewed, but since the prediction results are usually inaccurate or conservative if the S-N curve is not appropriately determined, for that reason Paris' law will be utilized throughout this research.[51] Fatigue crack propagation and RUL prediction will guide the operation team to actively schedule maintenance. The next section covers the available monitoring and maintenance systems that are most widely used in the wind industry.

2.4 Condition-Based Monitoring and Maintenance Systems

In the growing wind energy industry, mitigating the rise of turbine component failures has gained much attention, garnering interest in using CBM approaches over the failure-based maintenance and time-based preventive maintenance measures [54]. Modern WTs are growing larger and smarter, well equipped with more sensors and better control systems. The availability of health-monitoring data (i.e., sensing signals), has made a substantial contribution to the demand for CBM and opened up many research areas for the wind industry. [54]

Among the available maintenance approaches, CBM is the most advanced and effective in proactively tracking and monitoring the health of WT components and minimizing unnecessary downtime and cost overrun.[54] Figure 12 below illustrates the influence of different maintenance strategies on the condition of an asset. Condition-monitoring systems provide an early indication of component damage, allowing the operator to plan component repair prior to premature failure.[54] However, despite the age of the control system in the wind industry, there are numerous glitches in the collected warning signals and health-assessment data that do not reflect the reality of the WT system's operation and observed turbine component failure mode.[43]

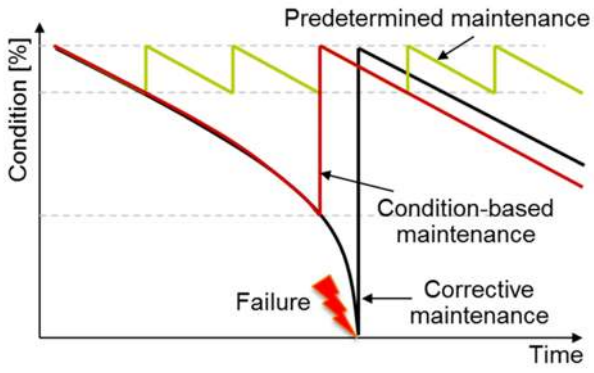


Figure 12 – Influence of CBM strategy on asset condition. [54]

The implementation of supervisory control and data acquisition (SCADA) on wind turbines is becoming more of an industry standard practice. In the past decades, design of a wind farm and turbine array was often supplemented with meteorological (MET) tower and aircraft detection lighting systems (ADLS), which feed data to the SCADA system through a condition-monitoring system (CMS) to better track the health

of the turbine component systems. The use of CMS for CBM is considered a three-step process, involving the following key elements: (1) *Data acquisition*: the collection and storage of raw data through SCADA and other data-acquisition systems. (2) *Data diagnostics and processing*: a prerequisite for making use of the large volume of collected data. Within this process, features from raw data are extracted to monitor the condition of a specific component. It is important to analyze and interpret the raw data, with the goal of detecting faults and locating where they are developing. (3) *Prognostics*: utilizing CMS data for maintenance, planning, and optimization by analyzing the remaining useful life (RUL) of the monitored component. The following subsection presents the two widely used vibration and oil-based state-of-the-art methods of WT gearbox condition monitoring and structural health monitoring that are available in the industry. [54]

2.4.1 Vibration-based monitoring

Vibration-based condition monitoring is the most broadly used assessment method for the whole drivetrain of the WT system. The premise of this approach is that damage in rotating mechanical machinery will lead to excessive vibration, which is caused by unbalanced loads or defects. [54] Instrumentation and sensors applied for vibration-based CMS include accelerometers, as well as velocity and displacement transducers installed in different positions in the drivetrain; however, most of the WT gearbox uses different types of accelerometers ranging from the low- to high-frequency range. Selection of the sensor instrument also depends on the size of the WT in addition to its dynamic range and sensitivity. [54] A gearbox typically requires 4–5 sensors per sub-component, measuring in radial and axial directions with a frequency range from 0.1

Hz – 10 *Hz* at the low- and high-speed shaft. The sensors should be mounted in a way that allows for a reliable frequencies measurement, and thus, gearboxes containing planetary stages will have one sensor per stage in the area of the ring gear and at the level of the sun. [54] Vibration-based data processing and diagnosis methods are divided into time-domain and frequency-domain analysis techniques. *Time-domain analysis* methods include statistical-parameter-based methods, trend analysis, time-synchronous averaging, the planet-separation method, and amplitude demodulation.[54] The *frequency-domain analysis* method includes fast-Fourier transform and spectrum analysis, order analysis, cepstrum analysis, envelope-curve analysis, spectral kurtosis filtering, and wavelet analysis. [54] Understanding each of the aforementioned methods plays an essential role in better diagnosing and analyzing observed faults in order to prevent premature failure of the gearbox.

2.4.2 Oil-based monitoring

Oil-condition monitoring (OCM) is also a typical industry practice applied to monitor the health of the WT gearbox component. OCM goals have two main objectives: 1) monitoring the quality of the lubricant oil, including the effectiveness of the filter system, and determining the necessity of an oil change; and 2) assessing wear-debris content in the oil particle, which can indicate faults in the mechanical components.[54] These two main objectives are accomplished in an online or offline setting, depending on the turbine life cycle, period of maintenance, and inspection requirements. Instrumentation and data-acquisition sensors required for OCM mainly include sensors measuring pressure, temperature, and particles. Parameters that could negatively affect the oil performance mainly include oil temperature, water ingress, air, particles, oil

oxidation, and wear debris.[54] Thus, appropriate sensors for each of the above parameters are utilized. Unlike vibration-based CMS, the data processing and analysis for the oil-based condition-monitoring technique is straightforward. Trend analysis is normally used to monitor the amount of the particles and other factors developed over time.[54] The choice of online and offline oil analysis depends on the operating condition during oil sampling, and industry practice follows ISO 4406 codes to interpret the results.

Consistent condition-monitoring data remain very limited due to recent evolution of the control system and the frequent control system downtimes. According to [10] it is apparent that more than half of wind turbine failures stem from control systems and electric components, which is affecting the diagnosis and prognostics portion of the CBM approaches intended to accurately predict failures of the critical components. Data-driven models alone are therefore insufficient to generate a more robust predictive model and prevent premature failures of the WT critical components.[54] The CBM approach is as good as its health-monitoring data. Some literature has modeled the degradation process as a continuous stochastic process (e.g., failure probability distributions), while others have considered degradation a discrete aging process. [54] To account for both the continuous degradation process and damage accumulation, the physical models and data-driven models must be integrated together.

2.5 Summary and Conclusion

In this chapter, literatures on wind turbine component reliability, fracture mechanics, gear-crack propagation, and condition-based maintenance and monitoring systems that pertain to the gearbox component are reviewed. The most common 2-MW wind turbine configuration and component systems are introduced. Existing industry

practice in the growing area of condition-based maintenance, in terms of diagnosis and prognostics, is discussed. It is apparent that the data-driven approach alone does not adequately predict the premature failure of the critical WT components, and thus, the integration of physics-based fracture and crack propagation knowledge could play a key role in better predicting the premature failure of the WT gearbox, minimizing unnecessary downtime, and increasing the reliability of the system in general.

Chapter 3—Methodology

3.1 Introduction

This chapter gives an overview on how the data are gathered and the procedures and steps taken to analyze them. An integrated prognostics condition-based maintenance model combines data-driven and physics-based methods to overcome the limitations of each in terms of the availability of gearbox-failure data and the physical complexity of the gearbox model itself. Historical data on wind turbines in the U.S. Midwest and gearboxes with unique gear tooth counts, planet gears, ring gears, sun pinions, and LSS and HSS pinions from five different original equipment manufacturer (OEM) products were evaluated.

Wind farms that suffered a lot of premature gearbox failures between two and four years after being commissioned were studied to address the research hypotheses. To re-iterate, these are Hypothesis 1: Time to failure (TTF) of the gearbox turbine component is affected by various environmental conditions, rotor systems, and gearbox factors; and Hypothesis 2: If rotor systems and gearbox-integrated CBM actions are implemented, it will reduce the premature failure of wind turbine gearboxes. In the next few sections, details of the steps taken to generate the integrated prognostics data-driven methods and assess the research hypotheses are described and found to have better predictive capability regarding the gearbox components prior to failure.

3.2 Integrated Prognostics Data-Driven Method

The integrated prognostics method simply fuses data-driven and physics-based approaches to analysis to account for data limitation and the complexity of physical gearbox models. The predictive model will help us generate a degradation model for the premature failure of the gearbox and the remaining useful life (RUL) distribution. Necessary maintenance actions will be scheduled on the basis of age, oil degradation, and predicted RUL.

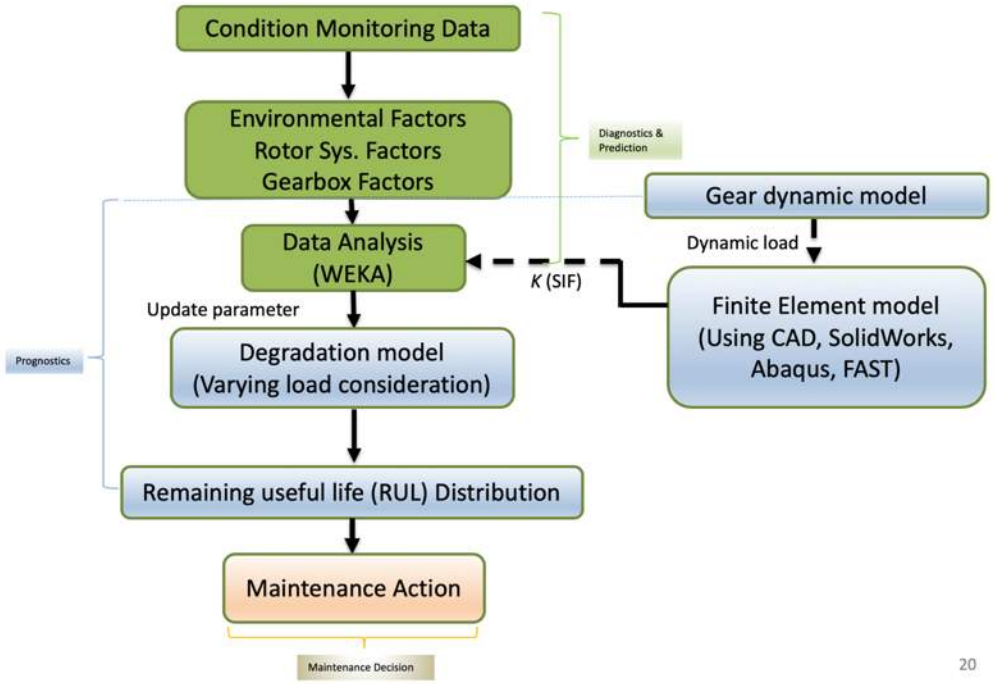


Figure 13: Integrated prognostics CBM model and flowchart.

This method primarily uses historical data on wind turbine failures resulting from gearbox failures at various points among the planetary gear, intermediate shaft (IMS) or

high speed shaft (HSS), and ring gear. Condition-monitoring data are collected from the supervisory control and data-acquisition (SCADA) system, which is integrated with the PI server and coupled with the operation and maintenance (O&M) database used by operation managers to track failures. Most available studies have used vibration or oil-based condition-monitoring systems to track the health of the gearbox system. However, most of those studies also failed to correlate these data to actual environmental conditions, such as the rotor or drivetrain systems in combination with the realtime stress intensity of the gearbox in early stages of the turbine's life cycle.

The predictive model uses three predictor variables, *environmental factors*, *rotor system factors*, and *gearbox factors*, which correlate to the response variable, *time to failure* (TTF) of the gearbox. The stress intensity factor (SIF) from the finite element analysis is also integrated as an input variable, as part of the gearbox factor. As the bathtub curve in figure 6 shows, the failure intensity, particularly in the early years of turbine in-service, is very high. Therefore, this predictive model examines mainly variables that can contribute significantly to the premature failure of the gearbox.

3.2.1 Data description, collection, and trends

The three categories of data considered are environmental factors, rotor system factors, and gearbox factors (see table 3 below for details). It worth studying the rotor and drivetrain system variables to determine whether they are the leading influences on the premature failure of wind turbine gearboxes. Most of the gearbox failures studied in this research happened between 2016 and 2019, two to four years after the turbine was installed. The historical 10-minute data point intervals for the three predictor variables are used in the data-driven analysis, with four or five wind farms consisting of 100–110

turbines each. The trend of the input variables was studied thouroughly to understand its regression over time and the distribution of the dataset before the gearbox failure.

Environmental factors are predictor variables that depend mainly on weather conditions. For the purposes of the predictive model, the variables considered as environmental factors are *wind speed*, *ambient temperature*, and *wind direction*. The main driving variable is the fluctuating torque of the rotor system induced by variation in wind speed, and careful examination of these variables is therefore extremely important.

Rotor system factors are predictor variables related to the rotor blades, pitches, angles, and yaw system. As many studies have shown, in operational conditions, yaw errors create rapid-rotor torque fluctuations and resistance torque variations in response to changes in yaw angle and wind speed, and this is highly likely to generate negative load fluctuations in the yaw system [31]. Therefore, it is important to consider these rotor system factors as input variables: *nacelle position*, *blade pitch angle*, *rotor system pitch angle*, and *torque from the rotor system*.

The third group of predictor variables are the gearbox factors, which include *gear bearing temperature*, *gear oil temperature*, and *tower acceleration*. These variables are tied directly to the gearbox and are the main condition-monitoring data used in the oil- and vibration-based condition monitoring of the gearbox and other turbine components.[54] Furthermore, as specified among the gearbox factors, finite element–generated *stress intensity factors* are also integrated as a predictor variable. The next section describes how the physics-based finite-element analysis of the spur gear will be conducted.

3.2.2 Physics-based analysis

The understanding of the basic features, processes, and mechanisms of the failure of WT gearboxes adopted in this research is that gear failures frequently result from excessive and unexpected loading or misalignment. Understanding of the root causes, involving fundamental design errors and manufacturing deficiencies, have come a long way, and lots of improvements have been made in those areas. For this reason, advanced computational and engineering analysis methods are useful for dealing with unprecedented and variable wind-loading conditions. Table 1 shows the material properties of the failed gear during operation, and used in the finite element modeling. Other type of gears from various OEM were also considered for comparison.

Gear Tooth No.	**Poisson's ratio**	**Pressure angle (deg.)**	**Pitch diameter (in^{-1})**	**Outer diameter (mm)**	**Young's modulus (Pa)**	**Module**	**Base circle radius (mm)**
47	0.3	20	8	79.5	2.068	3.2	72

Table 1: Gear material properties

The physics-based analysis involves the physical model of the spur gear itself and uses finite element (FE) analysis software to calculate the *stress intensity factor* (SIF) K, static displacement, and crack propagation. If K is known, it is possible to analyze all the stress components as functions of r and θ (see equation 12 and [28]). The flowchart in figure 14 shows the steps taken to predict crack length and propagation using physics and numerical modelling methods.

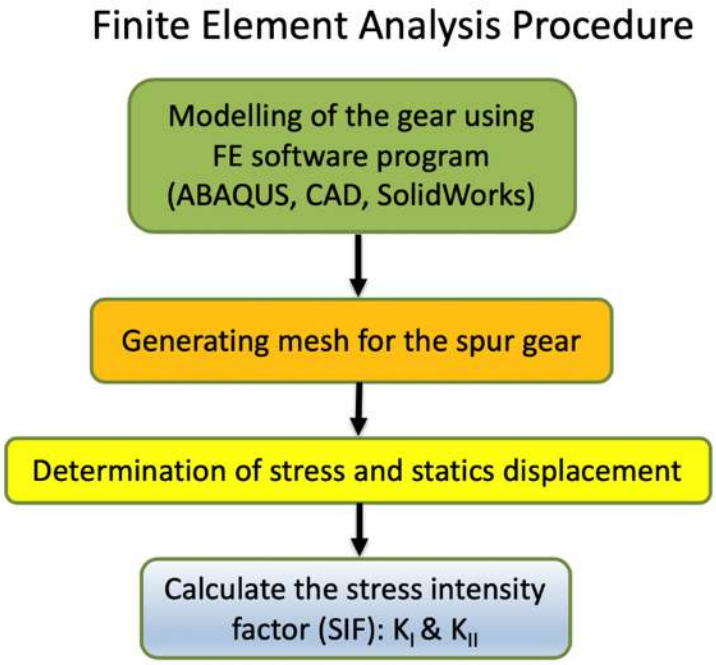


Figure 14: Determination of stress intensity factors (*SIF*).

The FE model is established using the programs CAD, Solidworks, and ABAQUS. We first draw the 2D and 3D geometries of the gear portions using CAD and then import these to FEA. Then external forces and moments of loading from the rotor system caused by wind gusts and abnormal operating conditions, such as the emergency stopping of the drivetrain system, are evaluated using a cartesian coordinate system (see figure 15). The wind turbine hub reaches heights into the hundreds of meters, and even normal wind produces greater acceleration and variation in velocity in the vertical direction. Historical data on wind speed, wind direction, and other variables are used to evaluate the loads exerted on these turbines prior to complete gearbox failures.

M_x refers to the driving torque applied to the gear drivetrain system due to the

rotation of the rotor blades; the *Mz* and *Fy* components of the wind flow are inclined to the rotor axis. F_x and F_y are the results of the axial pressure of the wind and the total weight of the hub assembly, and M_y is mostly caused by the difference in wind pressure distribution among the three rotor blades.[18] Following the load path, the driving torque M_x and vertical force F_z are clearly the most critical components of force and contribute significantly to the structural failure of the gearbox.[18]

In addition to the variety of OEM gear models, the major specifications of the gear components considered are the teeth, pitch-circle diameter (mm), tooth width (mm), pressure contact angle of 20°, rated input and output speeds, speed-up ratio, specific density of gears, Young's modulus, and Poisson's ratios. Each of the first- and second-stage planetary gearsets consists of a fixed ring gear, planetary gears, and a sun gear. Therefore, FEA was carried out to determine the stress distribution and crack propagation pathways of the gearbox structure.

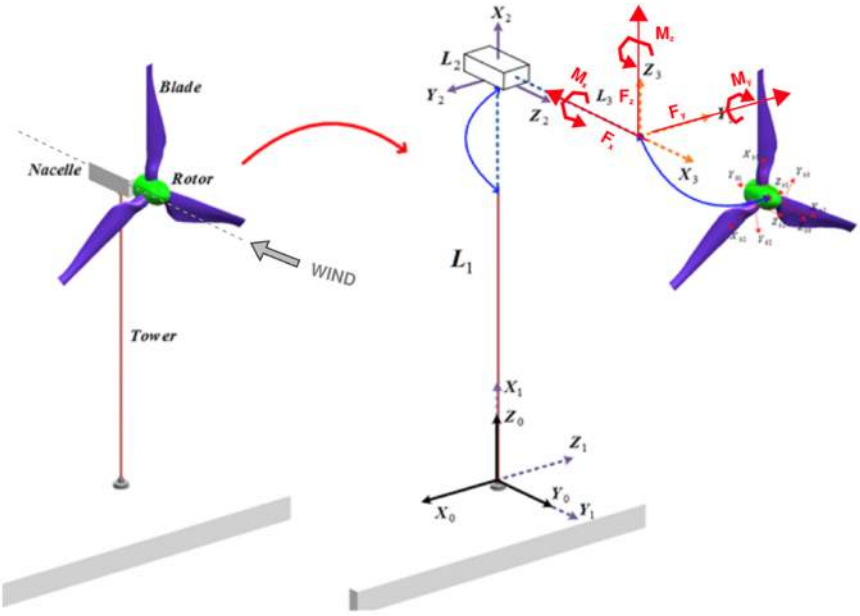


Figure 15: Moment and force components in cartesian coordinates.

Gear meshing was generated automatically by the selection of controlling areas. For structural components such as a WT gear machine, where crack modelling is important, the six-node triangular element or the mesh delete meshing associativity from the main menu bar will be used. The FE model is used to calculate *K* near the crack tip in order to correlate the local displacement with the asymptotic values. Thus, *K* can be expressed as a function of displacement. The stress intensity factor *K* was used as an input predictor variable to the data-driven method. The compiled 10-minute–interval raw data for each attribute were collected, and diagnostics, detection, extraction of the main statistical features, and probabilistic distribution were then evaluated.

3.2.3 Data diagnostics and imputation

Once the raw data were collected, diagnostics were performed. This involved detecting, isolating, and identifying faults, and extracting statistical features. When raw data were extracted from the SCADA and OSIsoft PI servers, errors such as no data being found were observed at certain intervals for all predictor variables, which was negligible in comparison to the overall observed datapoints. A sliding-window algorithm using Jupyter notebook (formerly IPython) algorithms was used to extract statistical features such as mean, standard deviation, histogram (using n = 5 bins), and probability distribution time-series data points for each attribute (see Appendix C for details).

The operation and maintenance-failure tracking database depicts valuable raw data (see table 2), which were blended with the SCADA condition-monitoring data shown in table 3 to generate the predictive model. The overwhelming majority of the failure modes involve damage to the ring gear and tooth cracks. Figures 16 and 17 show some hypothetical locations where cracks in the tooth were observed. The next section

depicts the creation of diffirent views of the data, so that machine-learning algorithms can be used to better expose the structure of the regression problems.

New or Rebuilt	Install Date	Failure Date	Hours	KWhrs	Failure Loc.	Failure Mode
New	10/9/17	9/27/19	17256	18956791	Planetary gear	Ring gear damage
New	10/5/17	2/27/19	12035	10822299	IMS/HSS tooth	Tooth crack
New	9/18/17	6/3/19	14976	11034561	IMS/LSS	Tooth damage
New	9/24/17	1/26/19	10453	10112405	Planetary gear	Ring gear damage
New	9/3/17	11/30/18	13223	13617917	Planetary gear	Ring gear damage

Table 2: Sample data, O&M failure tracker database.

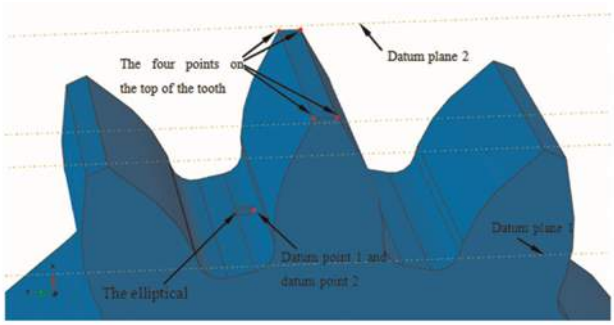


Figure 16: Hypothetical location of the crack in the tooth root.

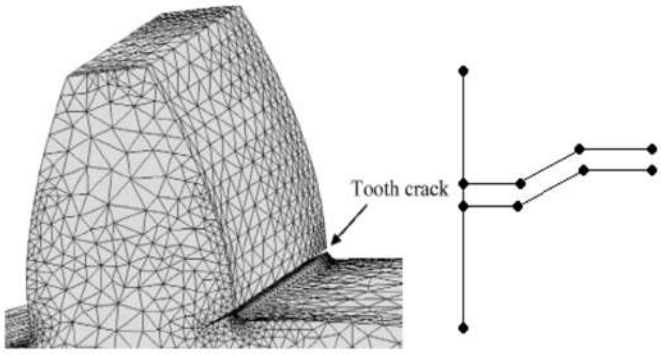


Figure 17: Element representation of the observed cracks.

3.2.4 Prepare views of the dataset

Once the full raw time series dataset is compiled from the health condition–monitoring sensors and results of the finite element analysis, it must be reframed as an unsupervised learning dataset before being executing in WEKA machine-learning algorithms. For the purposes of this research, we will use normalized and standardized datasets, followed by sliding-window Jupyter (Python) codes to extract statistical features such as mean and standard deviation.

A normalized dataset will create all the input attributes in the range 0 to 1, which benefits the regression algorithm, which could have been influenced by the scale of the attributes. In all cases, all the remaining attributes apart from the class attribute have numeric value normalized in the given dataset. If some attribute has a gaussian-like distribution, the data can also be rescaled to take this into consideration by using a standardizing filter.

3.2.5 Proposed data analysis methods

Studies have used many methods of analysis to predict and quantify these types of predictive model. For the wind industry, depending on the size of the datasets and the nature of the expected results, various machine-learning tools have been implemented. For the purposes of this research, however, appropriate regression algorithms were suggested using a machine-learning tool in WEKA. Depending on the observation of residual distribution, multi-collinearity, normality checks of the residuals, and other unforeseen factors, model reduction and simplicity are parts of the goal process.

Data-mining algorithms and classifiers such as k-nearest neighbors, support vector machines (SVM), decision tree, and multi-layer perceptron were verified to which

had the best prediction accuracy and performance in comparison to the linear regression algorithm.

3.2.5.1 Regression algorithms

In this section, various machine-learning regression algorithms that were used are outlined. In general, regression lets us estimate the mathematical relationship between a dependent variable (usually called *Y*) and an independent variable (*X*). Table 3 shows the available variables, often called *response* and *predictor* variables respectively. The analysis was completed using statistical software tools and regression algorithms in WEKA because of their robustness and for having multiple functionality alternatives that are presentable for the given dataset.

Response variable (*Y*)	**Predictor variables (*X*)**
Gearbox time to failure (TTF) (*Y*)	Environmental factors • Wind speed (X_1) • Ambient temperature (X_2) • Ambient wind direction (X_3) Rotor system factors • Nacelle position (X_4) • Rotor system pitch angle (X_5) • Rotor system RPM (torque) (X_6) • Blade A pitch angle (X_7) • Blade B pitch angle (X_8) • Blade C pitch angle (X_9) Gearbox factors • Nacelle temperature (X_{10}) • Gear bearing temperature (X_{11}) • Gear oil temperature (X_{12}) • Tower acceleration *x* direction (X_{13}) • Tower acceleration *y* direction (X_{14}) • Stress intensity factor (*K*) (X_{15}) – FE analysis result • Power generation (X_{16})

Table 3: Response and predictor variables.

WEKA has a large number of regression algortithms. Of the top five supported regression algorithms for predictive modeling, two were implemented to compare the robustness of the model, perform linear regression, and support vector machine or support vector regression.

Linear regression: This is a very simple regression algorithm that is fast to train and can perform better when the output variable has a linear combination of the input data. A common problem in multiple regression analysis is multicollinearity, which occurs when the *X*s are highly correlated. This is examined carefully using variance inflation factors (*VIFs*), which quantify the inflation of the variance.[55]

When using other statistical tools, the rule of thumb is that *VIFs* of greater than 5 warrant further investigation, but *VIFs* greater than 10 are signs of serious multicollinearity requiring correction, in which case the partial least square, a stepwise process, is used as an alternative method of analysis.[55] In WEKA, however, the algorithm is meant to perform less if the training data have input attributes that are highly correlated. It can detect and remove the highly correlated input variables automatically by setting *eliminateColinearAttributes* to true under the GUI objective editor. The machine-learning tool also generates a summary of output values that takes the accuracy, normality, and residual checks into consideration.[62]

Support vector regression (SVR) or machine (SVM): The main difference is in finding the line that best separates the training data into classes and that best minimizes errors. The SMOreg algorithm under the function group in WEKA projects the data into higher-dimensional space to draw lines and make better predictions. On some occasions, different kernels can be used to monitor the flexibility in the projection, but for this

model the polykernel was used.[62]

3.2.5.2 Regression model validation procedure

The two-step validation process involves calculating the regression model for a sample data set (training set) slanted to a higher percentage (70%) and then validating this model with a 30% different data set (test set). A tenfold cross-validation test mode will also be used to validate the model in comparision to the percentage split of the data.

To evaluate the accuracy and effectiveness of a classification regression model in WEKA, the correlation coefficient needs to be evaluated. Numbers close to 1 are considered very good, and numbers close to 0 are considered poor. Likewise, most predictive features are evaluated by examination of the greatest voting number, either positively or negatively. Once a robust and effective predictive model is developed using the equation from the model and the dataset collected by changing actionable predictor variables, such as rotor pitching system and gear oil, improvement in the response variable (time to failure of the gearbox) is evaluated using paired t-test.

3.2.5.3 Paired t-test

Actionable predictor variables such as those among the rotor system and gearbox factors were tested for paired differences of means to determine whether changes to them have any value for improving the TTF of the gearbox. Precautions were taken to ensure the data fit the model assumptions. Paired two-sample t-tests were used to compare the two TTFs using Minitab.

A paired t-test was used to evaluate whether the research hypothesis should be rejected. The core purpose was to show whether the integrated prognostics CBM approach will improve the gearbox's TTF. The null hypothesis (H_o) was that there is no

difference between the comparative variable measurements. The alternative hypothesis (H_a) was that there is significant difference between them. The significance level α, the probability of rejecting H_o when it is true, is set at .05 on the basis of a 95% confidence interval.

The probability of the resulting error being found outside the designated confidence interval was measured using the P-value. If $P < \alpha$, H_o should be rejected, and if $P > \alpha$, it should not be. Rejecting H_o implies that there is a difference or improvement in the TTF of the gearbox given the changes made to the actionable variables. This method is as good as its data, so to increase accuracy the FE analysis method, which is characterized by the physics of nature and material science, was used to develop the integrated prognostics method described.

In the next chapter 4, analysis results, ouput of the methodology discussed under this chapter is illustrated. Various interpretation and conclusion of the results obtained and hypothesized information is explained.

Chapter 4—Results

4.1 Introduction

The integrated prognostics approach described in Chapter 3 predicts TTF on the basis of probabilities of various operational values from the environmental, rotor system, and gearbox factors. It generally applies a machine-learning regression algorithm to the historical attributes of a failed turbine to make it possible to predict future conditions of the wind turbine gearbox. The first hypothesis is about the effect of environmental, rotor system, and gearbox factors on the TTF of the turbine gearbox; the second is about whether condition-based maintainance (CBM) actions on the integrated rotor and gearbox factors improve the TTF of the turbine gearbox.

4.2 Input Variable Trend Analysis

Trends of all the individual time-series input variables and attributes were studied, as categoried in table 3. Years of historical data were used to study the trends and distribution of the environmental, rotor system, and gearbox factors.

4.2.1 Environmental factors

The environmental factor considered in this analysis, as illustrated in Chapter 3 are, *wind speed* (X_1), *ambient temperature* (X_2) and *ambient wind direction* (X_3). Figures 18, 19, and 20 show the portion of the raw data trend for these three variables that exerted loads on the rotor blade prior to the gearbox failing in two years after it was installed. For

the full raw data trend, see Appendix A (Fig. A_o).

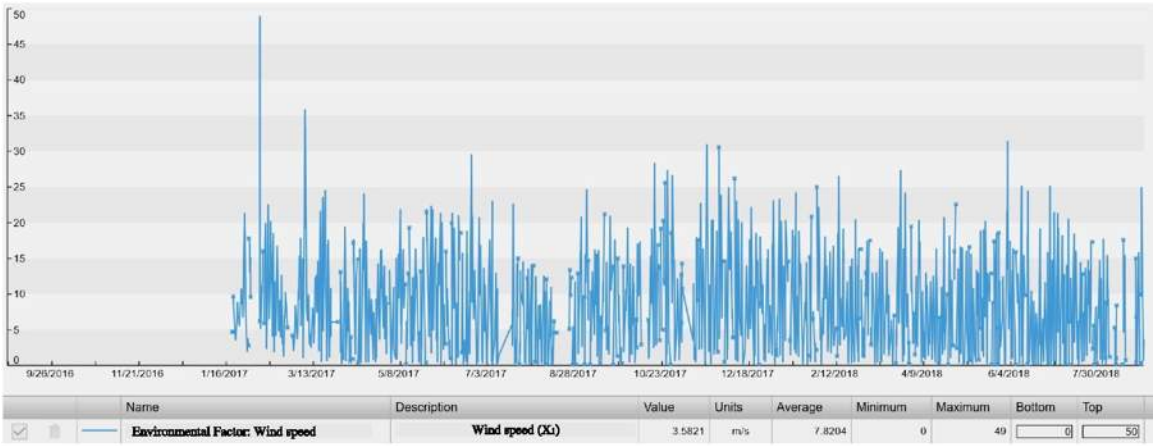


Figure 18: Raw data trend plot for wind speed (X_1).

The wind speed data show that all the datapoints were below the 50 m/s mark, which is far below the survival speed that most commercial wind turbines are designed to resist. Most of the datapoint were clustered in the 0–25 m/s range, which extended a little beyond the operational cut-off data (see Fig. 18). The average cut-in and cut-out wind speeds for a 2.0/2.2 MW–rated turbine are 3.0 m/s and 20 m/s, respectively. This means that whenever the wind speed is below 3 m/s or above 20 m/s, the turbine shuts itself down to avoid damage or reduced power-production efficiency.

Likewise, the standard operating temperature range is –20°C to 45°C for the obvious design reasons, unless the turbine is designed according to IEC 61400 standards to accommodate extreme cold or heat. Figure 18 shows that the ambient temperature oscillated between –32°C and +32°C, which make sense given the wind farm's location in the Midwest. If the temperature is too low, the rotor blades might freeze, which can cause the turbine to stop working; too high a temperature lowers the air density, which affects the energy output.

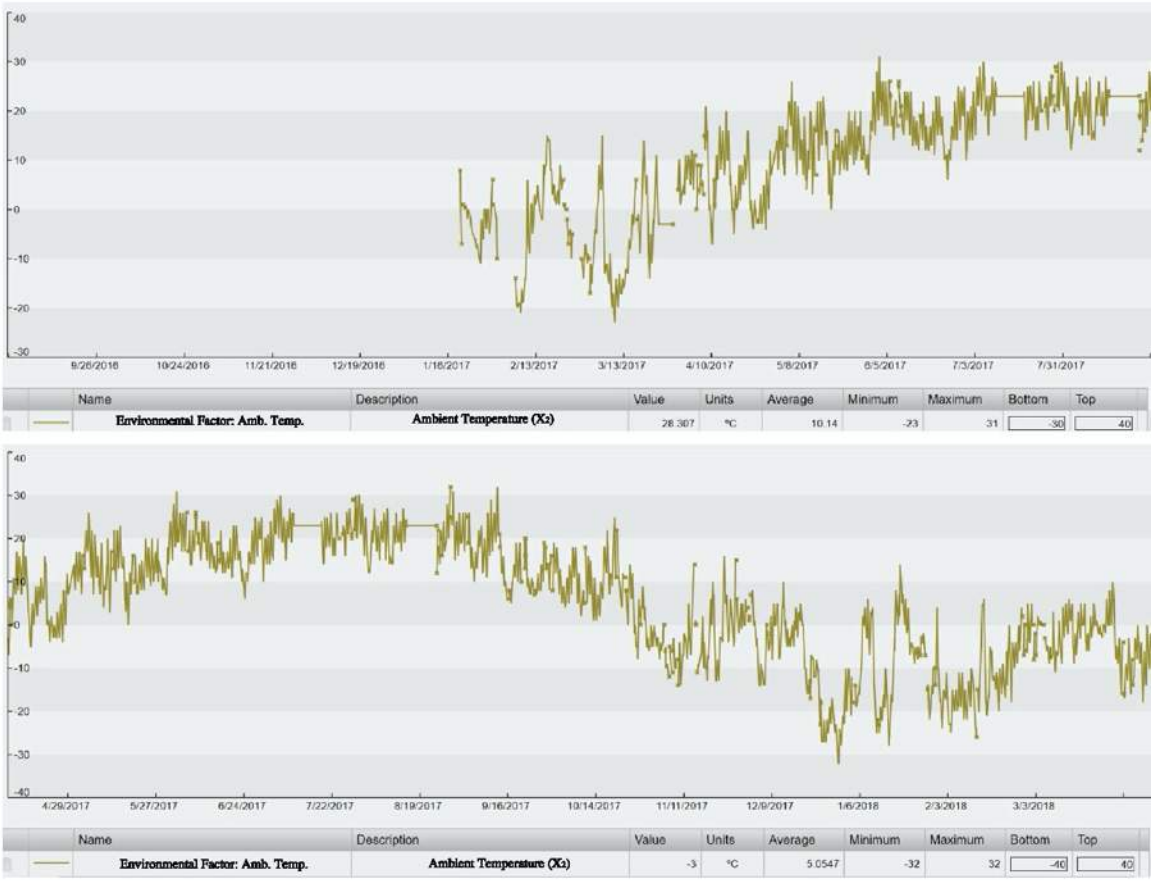


Figure 19: Raw data trend plot for ambient temperature (X_2).

Wind direction is the most stochastic variable that has a localized effect on the turbine rotor and the acceleration. The historical data showed that wind direction changes randomly from 0–359°. Some of the changes in direction are very abrupt and seem like the formation of a microburst around the turbine. The average yearly value was 260.45°, but the direction changes by 4°–10° every ten minutes (see Fig. 20 below).

The environmental factors are in general non-actionable variables, and the majority of the observed data fall within the design limits of the gearbox, except those points that are outside the range of operational limits or thresholds.

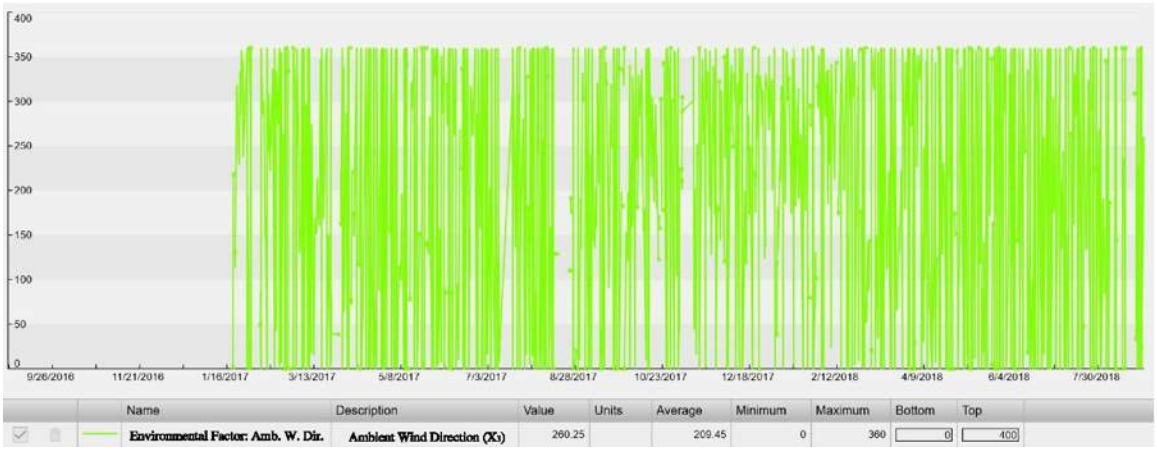


Figure 20: Raw data trend plot for wind direction (X_3).

4.2.2 Rotor System Factors

The other input variables considered under the rotor system factor are *nacelle position* (X_4), *rotor system pitch angle* (X_5), *rotor system RPM* (torque) (X_6), *blade A pitch angle* (X_7), *blade B pitch angle* (X_8), and *blade C pitch angle* (X_9). Table 4 summarizes the historical values observed. Figures 21–26 show the trend of the observed data.

Input variable	Units	Average	Maximum	Minimum
X_4	mm/ss	209.3–223.8	0.2	359.7
X_5	deg (°)	1.25–7.75	–4.6	88.2
X_6	rpm	9.23–13.39	–2.8	17.4
X_7	deg (°)	1.22–5.45	–4.7	88.2
X_8	deg (°)	1.37–5.85	–4.5	88.1
X_9	deg (°)	1.92–5.15	–4.6	88.6

Table 4: Rotor system factor summary.

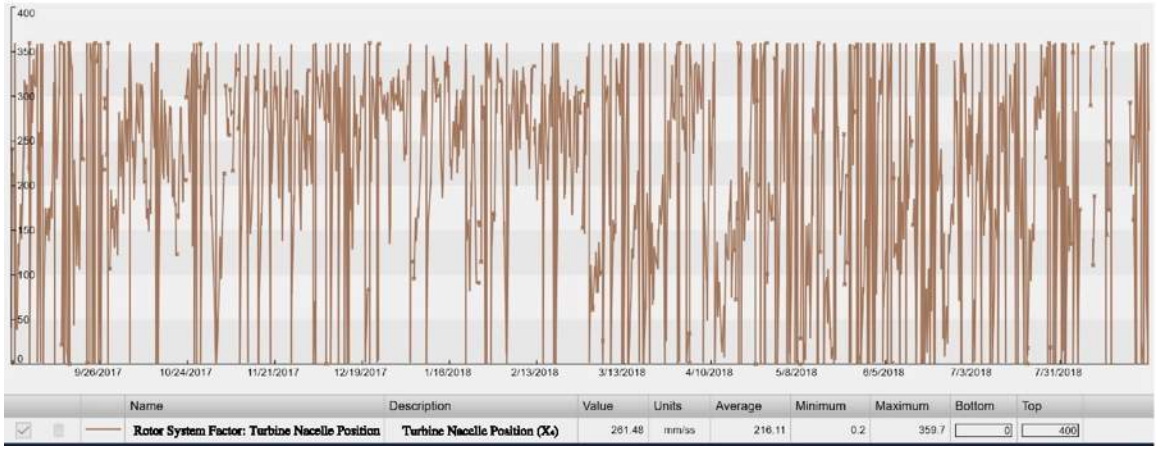


Figure 21: Raw data trend plot for turbine nacelle position (X_4).

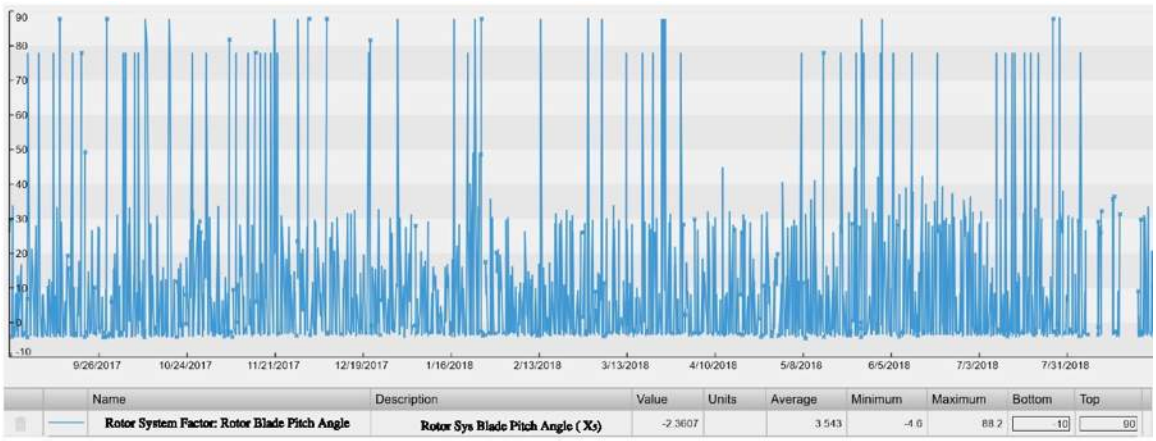


Figure 22: Raw data trend plot for rotor system blade pitch angle (X_5).

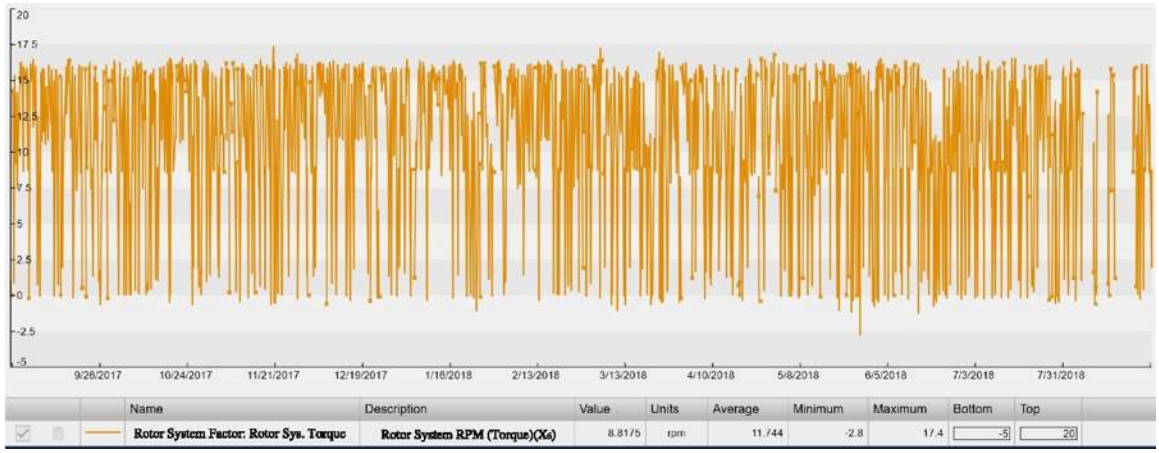


Figure 23: Raw data trend plot for rotor system rpm (torque) (X_6).

Blade pitch angles reaches its optimum value where the turbine power generation capacity is also at a maximum. The attribute interaction diagram shows that the thrust was higher when the wind speed was higher and decreased when the pitch angle increased. It is clear that the turbine's blade changing its pitch boosts wind-power efficiency. The pitch angle data below show the frequent shut-off and turn-on history the turbine experienced.

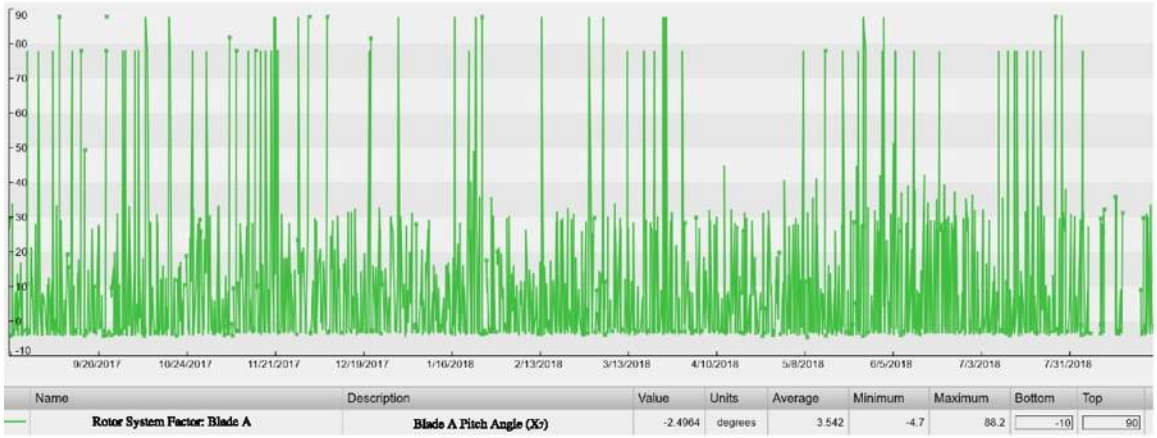


Figure 24: Raw data trend plot for blade A pitch angle (X_7).

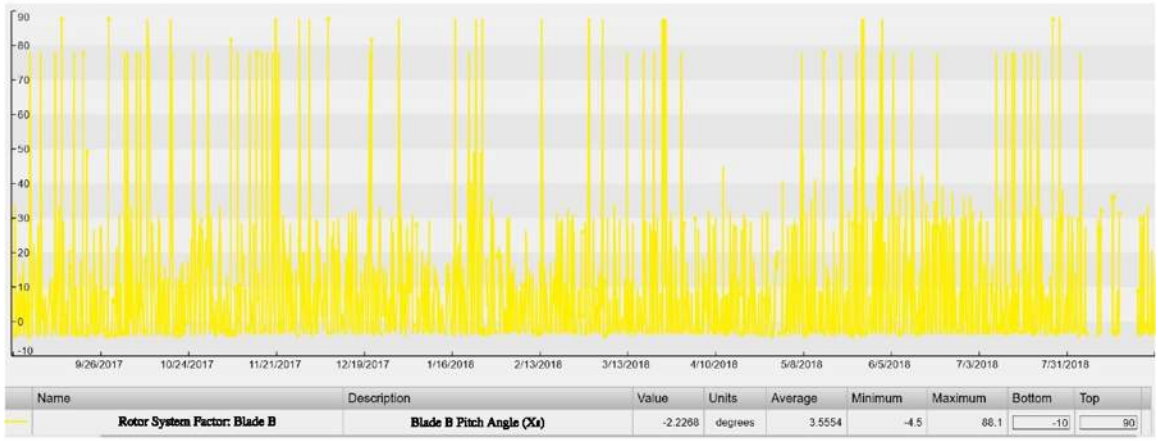


Figure 25: Raw data trend plot for blade B pitch angle (X_8).

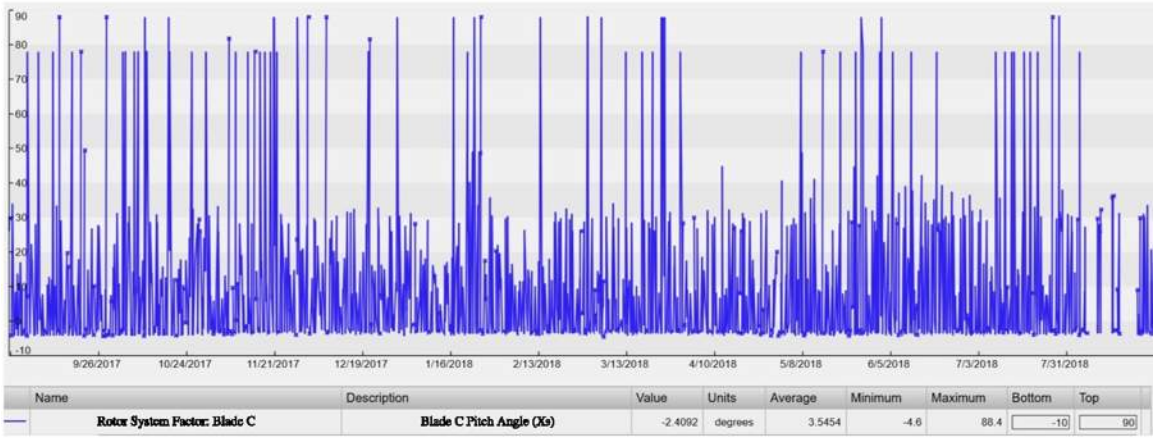


Figure 26: Raw data trend plot for blade C pitch angle (X_9).

4.2.3 Gearbox factors

The third group of input variables are called the gearbox factor, mainly because they are directly related to the gearbox and turbine. These include nacelle temperature (X_{10}), gear bearing temperature (X_{11}), gear oil temperature (X_{12}), tower acceleration in the x direction (X_{13}), tower acceleration in the y direction (X_{14}), and turbine generation (X_{16}). Figures 27–32 show the historical realtime data trends.

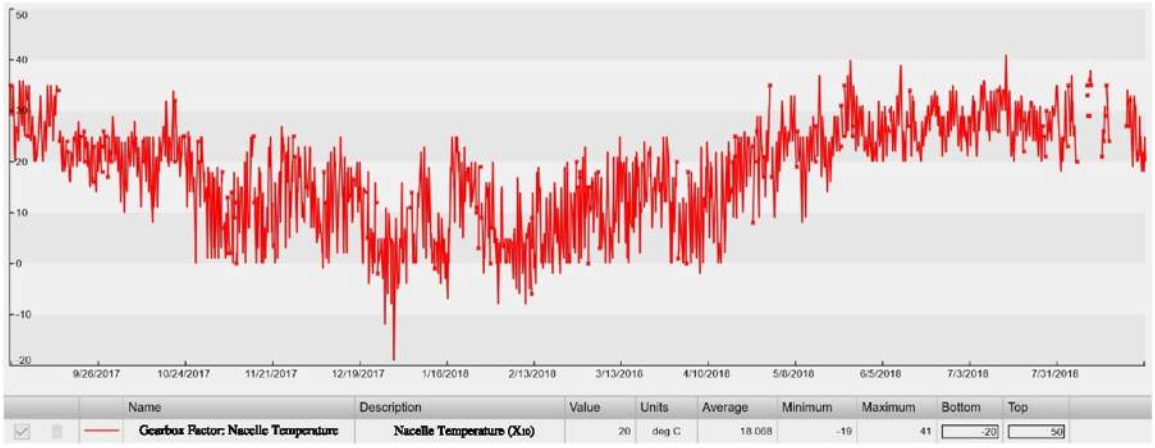


Figure 27: Raw data trend plot for nacelle temperature (X_{10}).

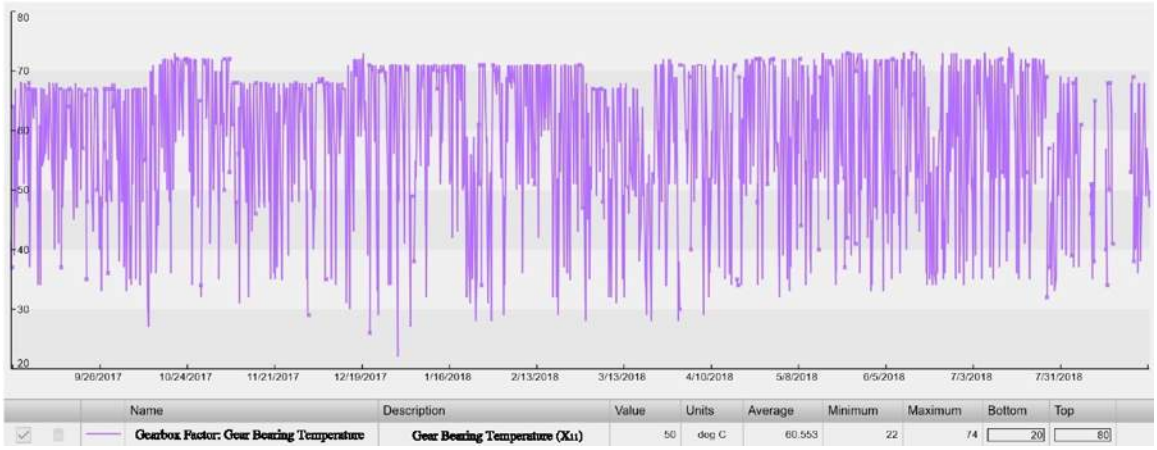


Figure 28: Raw data trend plot for gear bearing temperature (X_{11}).

Unlike the environmental factors, the operational weather temperature and the temperatures of the nacelle, gear bearing, and oil are considered actionable factors and must be monitored to assess the health of the turbine's components, including the gearbox condition.

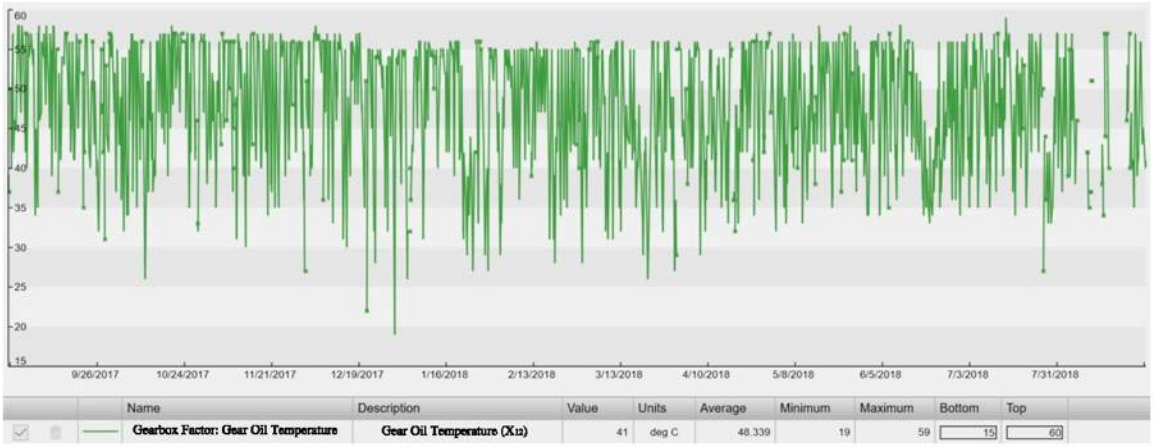


Figure 29: Raw data trend plot for gear oil temperature (X_{12}).

Higher tower acceleration is obviously the result of turbulence from sudden changes in wind direction in a short span of time. Localized microbursts are the most common phenomena that could cause the wind-direction changes in an infinitesimal time delta.

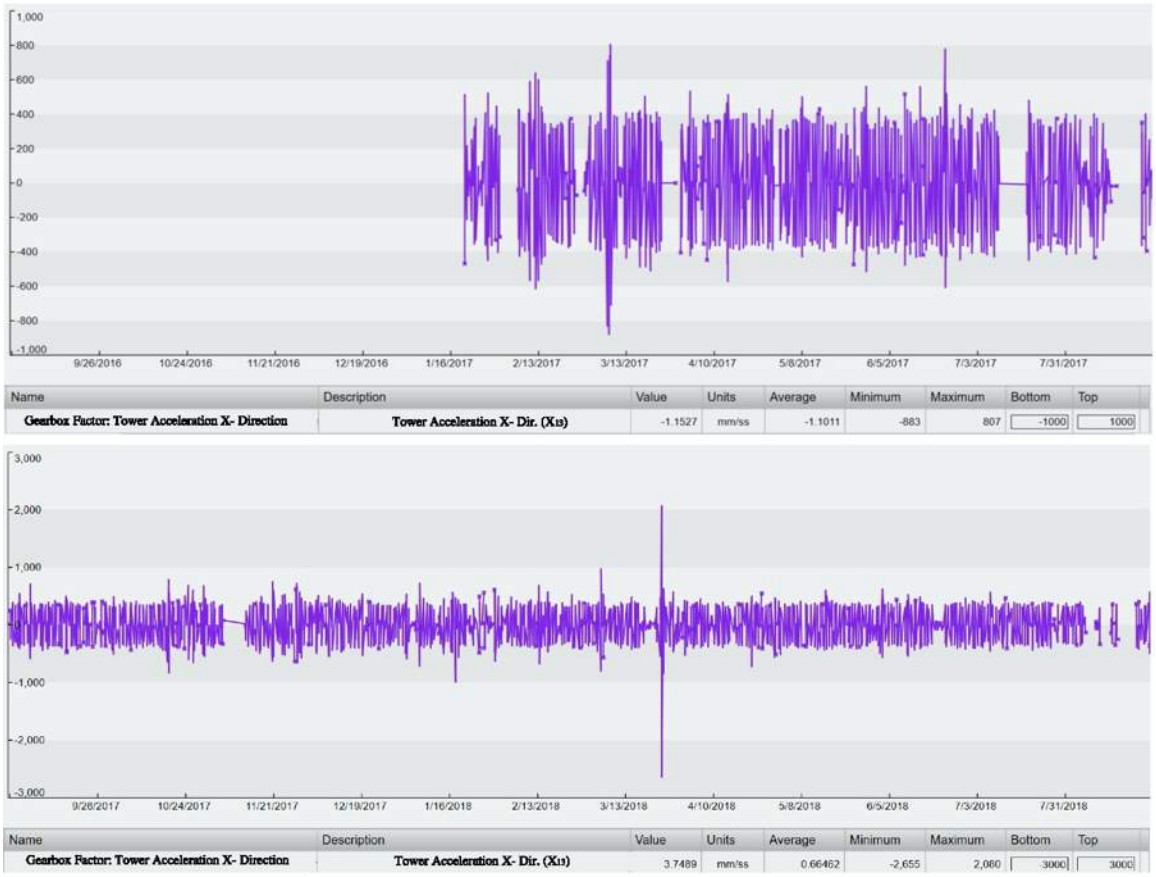


Figure 30: Raw data trend plot for tower acceleration in the *x* direction (X_{13}).

Input variable	Units	Average	Maximum	Minimum
X_{10}	°C	18.9–20.7	41	–19
X_{11}	°C	59.1–61.8	74	22
X_{12}	°C	33.5–60.7	59	19
X_{13}	mm/ss	0.41–1.11	2060	–2655
X_{14}	mm/ss	8.3–10.9	1955	–1057
X_{16}	kw	923–1.15	2178.3	60.1

Table 5: Summary of Gearbox factors.

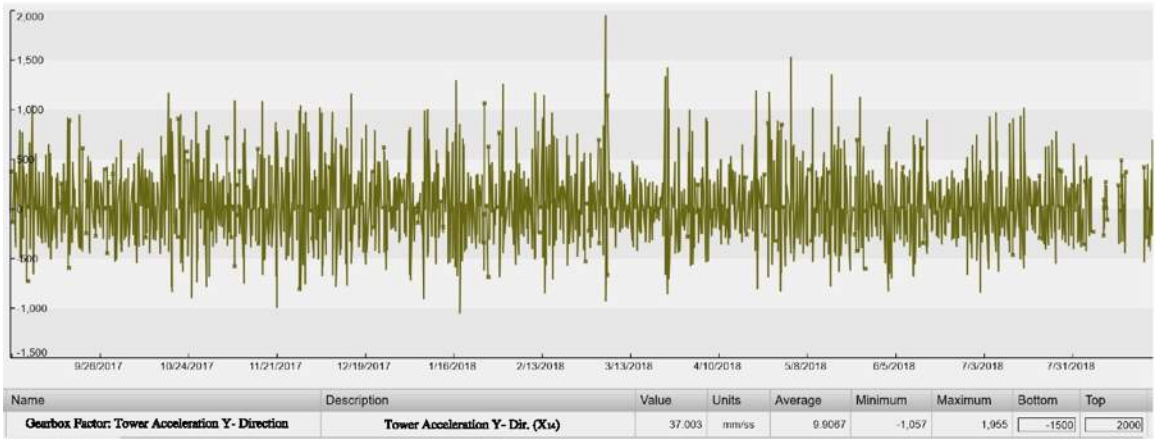


Figure 31: Raw data trend plot for tower acceleration in the *y* direction (X_{14}).

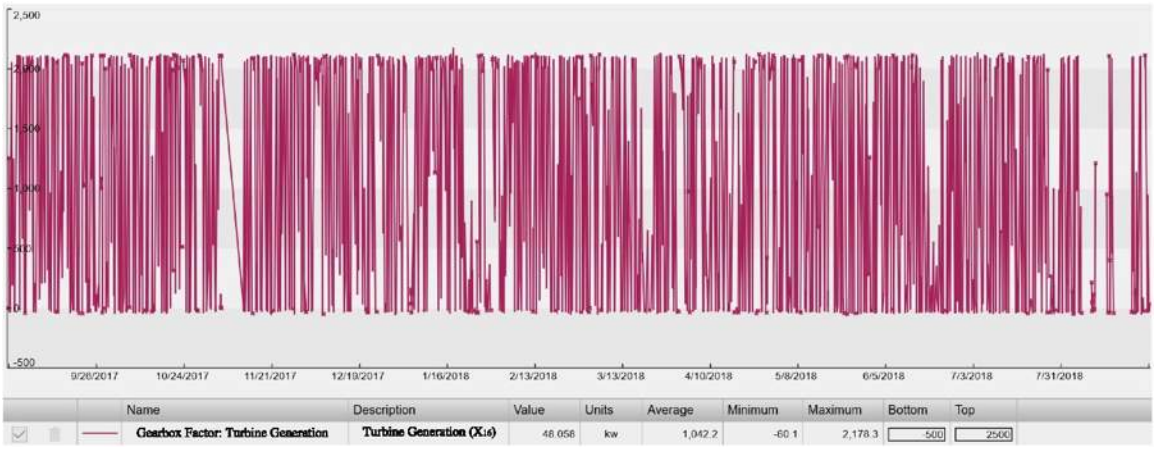


Figure 32: Raw data trend plot for turbine generation (X_{16}).

4.3 Stress Intensity and Finite Element Analysis Results

Spur gears have been modeled in finite element program using ABAQUS under static conditions with assumptions about initial crack time, and analyzed using instantaneously varying loads (see Fig. 33). The results have shown that the maximum stress is concentrated around the pitch circle of the gear tooth. The stress intensity factor for various historical wind speeds and varying torque at the pitch circle of the gear has

been evaluated and used as a predictor in the model. Figure 34 depicts the SIF trend used as an input variable in the predictive model, which is the result of varying rotor load and torque.

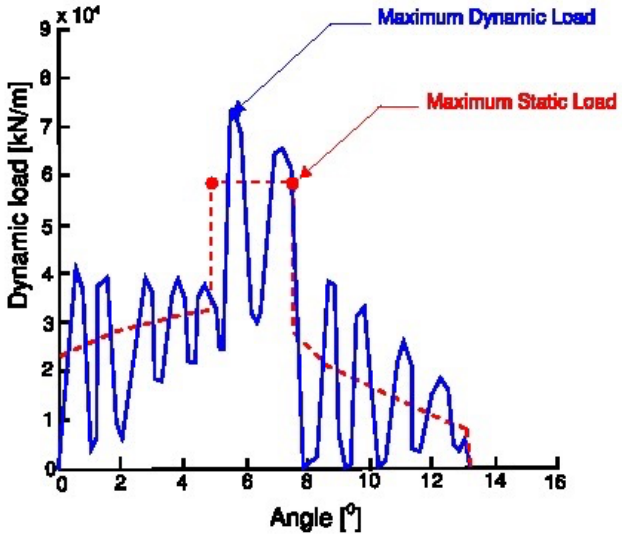


Figure 33: Static and dynamic load distribution.

The dynamic load distribution at every rotation angle with an assumed crack mesh is high, between 5° and 8°. Clearly, the exerted maximum dynamic load at that point is higher than the static load. Many other studies concur with this finding, that the location of the maximum dynamic load is where the crack propagates.[46] As figure 33 shows, the maximum dynamic load is zero during the crack tooth mesh at 8° in the x-axis.

Furthermore, the mean varying torque from the data points is applied to plot the curve of mode I and II SIFs vs. crack propagation. As mentioned in section 2.3, the mode

I is the opening mode and has a tensile stress that aligns with the plane of the crack.[46] Mode II is sliding, shear stress that normally acts parallel to the plane of the crack and perpendicular to the crack front. Therefore, mode I's SIF, which is the tensile stress, is more dominant than mode II's early in the lifecycle. Hence, K_I is used to compute the crack growth rate. Using the Paris' law and equation 13, we generate the crack degradation path and length. For the purposes of computation, an initial crack length of a_o = 0.2 mm was assumed. We then follow the fatigue crack propagation of the gear, considering uncertainities in loading and material properties following [61] see Appendix B, fig. B_3).

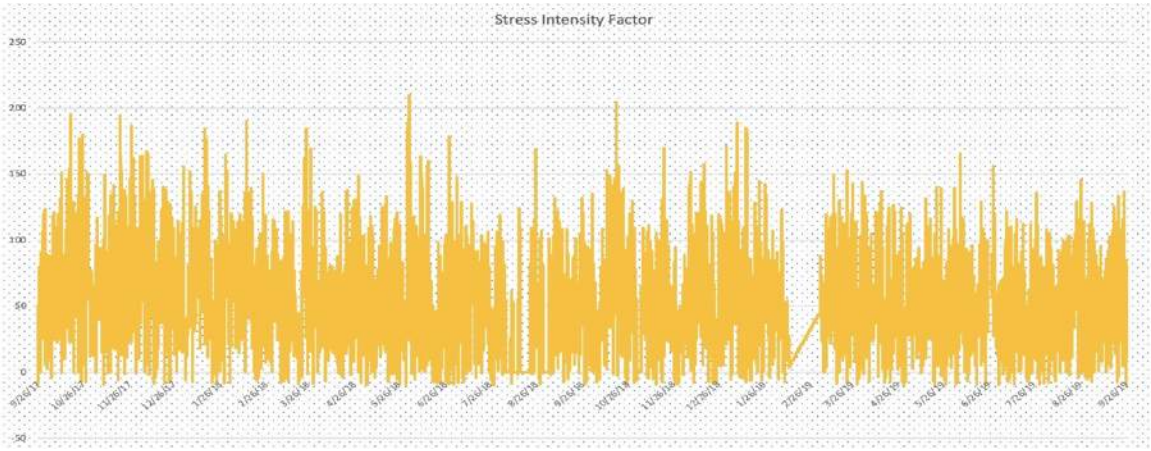


Figure 34: Raw data trend plot for stress intensity factor (X_{15}).

Input variables	Units	Average	Maximum	Minimum
X_{15}	KN/($m^{3/2}$)	41.79–65.93	89.872	8.938

Table 6: Summary of stress intensity.

The attribute interaction matrix between wind speed and SIF shows a high

correlation, and thus the stress intensity increases with the wind speed, which makes perfect sense, as torque is directly proportional to stress on the gear.

4.4 Regression Machine Learning Steps and Results

After the list of all the data attributes is compiled in the form of *Y*s and *X*s from all three categorical factors and FEA for SIF, feature-extraction, data-processing, and sliding-windows algorithms were applied to formulate a time-series problem as a supervised regression or a classification problem (see Appendix C for details), which reduced noises in the data to minimize error. The processed dataset were converted to a .cvs or .arff native file for WEKA to read it properly.

4.4.1 Analyzed Dataset

It is always important to review the distribution of each attribute and its interactions before modeling, to shed light on how to pick specific regression algorithm techniques that could be essential.

Summary of the dataset and statistics

- After feature engineered data, 58,194 instances were obtained. If 10-fold cross validation is used to evaluate the algorithms, then each fold will include 5,819 instances, which is great. However, because the number of instances was adequate, 70% of the training set and 30% of the test set were used for comparison.
- There are 17 attributes, 16 inputs, and 1 output variable (see table 6).

- By visualizing all attribute distributions, which shows a range of distributions, we can start using simple linear regression and comparing the results with nonlinear methods, such as support vector regression algorithms (see Appendix D, fig. D_0).

Input model

Response Variable	Predictor	Description	Category
TTF (Y)	X_1	Wind speed (m/s)[1]	
	X_2	Ambient temperature (°C)	Environmental factors
	X_3	Ambient wind direction (°)	
	X_4	Nacelle position (mm/ss)	
	X_5	Rotor pitch angle (°)[1]	
	X_6	Rotor system RPM	
	X_7	Blade A pitch angle (°)	Rotor system factors
	X_8	Blade B pitch angle (°)[1]	
	X_9	Blade C pitch angle (°)	
	X_{10}	Nacelle temperature (°C)	
	X_{11}	Gear bearing temperature (°C)[1]	
	X_{12}	Gear oil temperature (°C)	
	X_{13}	Tower acceleration, *x*-dir (mm/ss)	Gearbox factors
	X_{14}	Tower acceleration, *y*-dir (mm/ss)	
	X_{15}	Stress intensity factor (KN/m$^{3/2}$)	
	X_{16}	Power generation (kw)	

Table 7: Input variables of unreduced model.

4.4.2 Views of the Dataset

The view created *normalizes* attributes into the 0–1 range except for the class attribute (Y), which will benefit the regression algorithm, which could have been influenced by the scale of the attributes. This was done by choosing "unsupervised. attribute. Normalize" under the "Filter" pane. This reviews the minimum and maximum values of each attribute to confirm that they are 0 and 1 (see fig. D_1 in Appendix D).

Note that some of the attribution distributions are Gaussian-like, which means

[1] Eliminated variables in the reduced model.

they can be rescaled to take the distribution into consideration using the standardizing filter. Standardizing the dataset creates a copy in which each individual attribute has a mean of 0 and a standard deviation of 1 (see fig. D_2 in Appendix D).

4.4.3 Evaluation of regression algorithm results

An experiment was designed to evaluate a suite of standard regression algorithms on the views of the problems. Among the algorithms one can use in WEKA, linear regression and support vector machine (SMOreg) were compared for accuracy. The results of both are shown in table 8, with a split of 70% training, 30% testing.

X	Correlation Coefficient	r^2	*df*	*p*
Environmental factor	0.8165	0.7027	58179	<.001
Rotor system factor	0.8165	0.7027	58179	<.001
Gearbox factor	0.8165	0.7027	58179	<.001

Table 8: Summary of linear regression algorithms.

The same models were run using the support vector machine algorithms, and though some performed better than linear regression analysis, the overall correlation coefficient depicted lower values than for linear regression. The test mode was split with 70% training and 30% testing. The kernel was PolyKernel – E 1.0 – C 250007.

X	Correlation Coefficient
Environmental factor	0.7817
Rotor system factor	0.7817
Gearbox factor	0.7817

Table 9: Support Vector regression algorithm: Summary

Most of the models showed that 70% of the variance in the data was explained by the selected model.

Model X	Significance ($P < .05$)	$r^2 > 90$	$80 < r^2 < 90$	$75 < r^2 < 60$
X_1	12 models	2 models	13 models	31 models
X_2	39 models			
X_3	35 models			
X_4	9 models			
X_5	11 models			
X_6	36 models			
X_7	41 models			
X_8	6 models			
X_9	40 models			
X_{10}	40 models			
X_{11}	3 models			
X_{12}	42 models			
X_{13}	38 models			
X_{14}	32 models			
X_{15}	37 models			
X_{16}	35 models			

Table 10: Input variable significance distribution

The multiple linear regression algorithm performed better at generating the best model. The next section describes the results and the interpretations of the attributes. Table 9 depicts the distribution of the attributes and their significance in 45 models run for various Midwest wind farm turbines in which the gearbox failed.

4.5 Model Analysis Results and Interpretation

The predictive model depicts the analysis of the response factor which is TTF interms of the predictors which is the combination of the environmental, rotor system and gearbox factors. The hypothesis that was aimed to evaluate under this section is:

H_1: The time to failure (TTF) of the gearbox turbine component is affected by

various environmental conditions, rotor systems, and gearbox factors.

4.5.1 Reduced model output

X	Coefficient/Weight	t-Stat	P-value	Category
X_{10}	–0.8324	–3.1001	.000968	Gearbox factor
X_2	1.0031	3.9139	.000045	Environmental factor
X_3	–0.7053	–8.2356	<.00001	Environmental factor
X_7	–1.5918	–17.843	<.00001	Rotor system factor
X_9	1.3918	7.9713	<.00001	Rotor system factor
X_{12}	1.9172	14.3809	<.00001	Gearbox factor
X_6	–0.9505	–4.0759	.000023	Rotor system factor
X_{13}	–0.1408	–1.892	.029248	Gearbox factor
X_{14}	–0.4769	–6.0034	<.00001	Gearbox factor
X_{16}	0.1861	0.9146	.0180203	Gearbox factor
X_{15}	0.2941	1.3505	.008843	Gearbox factor

Table 11: Variable significance.

The linear regression model equation is given by the following equation:

$$\text{Y (TTF)} = -0.8324X_{10} + 1.0031X_2 - 0.7053X_3 - 1.5918X_7 + 1.3918X_9 + 1.9172X_{12} - 0.9505X_6 - 0.1408X_{13} - 0.4769X_{14} + 0.1861X_{16} + 0.2941X_{15} - 0.0387 \quad (14)$$

Correlation coefficient	0.8165	Degrees of freedom = 58179
Mean absolute error	0.133	R^2 value = 0.7027
Root mean squared error	0.1673	Adjusted R^2 = 0.69414
Relative absolute error	13.0365%	F-statistic = 82.2886
Root relative squared error	17.736%	
Total instances	58193	
Ignored class unknown instances	11	

Table 12: WEKA linear regression analysis result.

4.5.2 Model interpretation

In tables 7 and 11, it is clear that 2 of 3 environmental factors, 3 of 6 rotor factors, and all 6 gearbox factors are significant and affect the TTF of the wind turbine gearbox. The overall correlation coefficient is 0.81, which is good. Furthermore, the adjusted R^2 shows that 70.3% of the variance in the data is explained by this model. The greater the magnitude of the weight or the coefficient in the regression equation, the better the predictive power of that variable or attribute. And the closer the correlation coefficient is to 1, the better the predictive power of the model. The most positive and negative weights are associated with the features most predictive of the TTF of the wind turbine gearbox.

Gear oil temperature (X_{12}). According to the weight distribution, X_{12} has the most predictive power, which aligns with the influence that oil temperature has on gear failures. Research also indicates that the early wear, mircopitting, and scuffing of gears are influenced by the temperature of the oil in the lubrication system. Thin lubricant in the gear mesh may arise from low viscosity, which is typically a result of high-temperature oil.

Blade A & C pitch angles (X_7, X_9). Both the hydraulic and the electrical pitch system regulate power generation by changing the blade angles. The blade pitch angle is the attribute with the second-most predictive power, which makes sense because it affects the dynamic and static loads by changing its angle, especially when the turbine starts and stops. Like the weight, the wind speed data from the two years before gearbox failure indicates that the turbine experienced multiple starts and shut-offs. This could trigger rotor misalignment.

4.6 Test validation and results

It is important to maintain and monitor gear oil, but the rotor pitch system also clearly shows better predictive power for improving the TTF of the gearbox and avoiding sudden premature failure, which can cost a lot more than traditional time-based maintenance. The three-step process for validating the hypothesis 2 is described below. To re-iterate,

H_2: If rotor systems and gearbox-integrated CBM actions are implemented, they will reduce the premature failure of wind turbine gearboxes.

Step 1: *Optimization of pitch control system for the wind turbine* (*rotor system factor*). It is obvious that delays caused by hydraulic pressure and electrical pitch systems can destabilize or degrade the pitch-control system and hurt performance, which could lead to gearbox failure and even the total collapse of the wind turbine system. Therefore, realtime data were taken on input variables such as X_7 and X_9 of a few turbines for which the pitch control had been optimized using the novel pitch-control approach after gearbox failure.

The novel pitch-control approach was developed through opitimization and integration, delay perturbation, and signal-compensation techniques. Ultimately, it manipulates the angle at which the blade pitches and reduces the load to ensure the safety and reliability of the wind turbine system.

Step 2: Gearbox oil analysis and monitoring (gearbox factor). Routine oil analysis is a widely used proactive maintenance strategy for wind turbines. Using a test slate to evaluate the condition of the in-service lubricant was found to be effective for

assessing the condition of the gearbox and other mechanical components.[54] In this step, the main purpose is to lower the gear oil temperature, which reduces friction, keeps the gearbox clean, and protects the surface of the gear from damage such as mircopitting, sculfing, and corrosion.[54] Typically, the wind industry and the service and maintenance agreements that most OEMs have in place require changing the oil filter in the gearbox after 500 hours of operation. Every year, an oil sample should be taken to determine its viscosity, acid number, water content, additives, and contamination (particle count). Most OEMs recommend a complete oil change every two years of operation.[59]

The same wind turbine that underwent the novel pitch-control adjustment also had its oil additives and filter checked. As a result, input variables such as X_{12} were obtained to integrate the CBM approach with that of the rotor system factor. The description of step 3 depicts the process of validation by integrating steps 1 and 2 into the regression model.

Step 3: Improvement, evaluation, and validation. Realtime data from the turbines after the optimization of the pitch-control system and gearbox oil analysis were compiled. The adjusted X_7, X_9, and X_{12} variables were mainly used as predictors to solve for the new response variable (TTF) using the regression equation 14 described in section 4.5.1. Then a paired t-test was used to compare the new Y_2 variable with variables from before the gearbox failure.

Test results and interpretation

Here we describe the paired t-test analysis and its results to determine whether there is a statistically significant difference between the means of the two dependent variables Y_1 and Y_2 (TTF). As discussed in section 3.3.5, two related groups were set up

in Minitab as Y_1 *(Pre_TTF)*, which is the TTF captured before the gearbox failure, and Y_2 *(Post_TTF)*, the TTF generated using the regression equation 14 and data from after the improvement to the pitch-control system and gear oil (see the sample data in Appendix E, fig. E_0). Below is the Minitab output for the paired t-test, which provides useful descriptive statistics for two related TTF response variables.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Y_1 (Pre_TTF)	32267	577.3	6463.5	36.0
Y_2 (Post_TTF)	32267	7009.8	1552.5	8.6

Estimation for Paired Difference

Mean	StDev	SE Mean	95% CI for μ_difference
-6432.5	6350.2	35.4	(-6501.8, -6363.3)

μ_difference: mean of (Y_1 (Pre_TTF) - Y_2 (Post_TTF))

Test

Null hypothesis	H_0: μ_difference = 0
Alternative hypothesis	H_1: μ_difference ≠ 0

T-Value	P-Value
-181.96	0.000

Table 13: Paired two-sample t-test results.

The "Mean" column under Descriptive Statistics shows that the TTF after the adjustment of the pitch-system and the reduction in gear oil temperature was improved, with a mean difference of 6432.5 hours between the two TTFs (as shown in the estimation for the paired difference table). When the two TTFs are compared in the

difference estimation table, the standard deviation is 6350.2 hours, with a standard error from the mean of 35.4 hours, and the 95% CI for the mean difference rows shows a range from –6501.8 to –6363.3.

The t-value of –181.96 and the statistical significance (2-tailed *P*-value) of the paired t-test is 0.000, so clearly the *P*-value is less than .05. It can be concluded that there is a statistically significant difference between the two TTFs. In other words, the difference between the mean TTFs before and after the adjustment is not 0, which means that the rotor system and integrated CBM actions reduced the premature failure of the gearbox component.

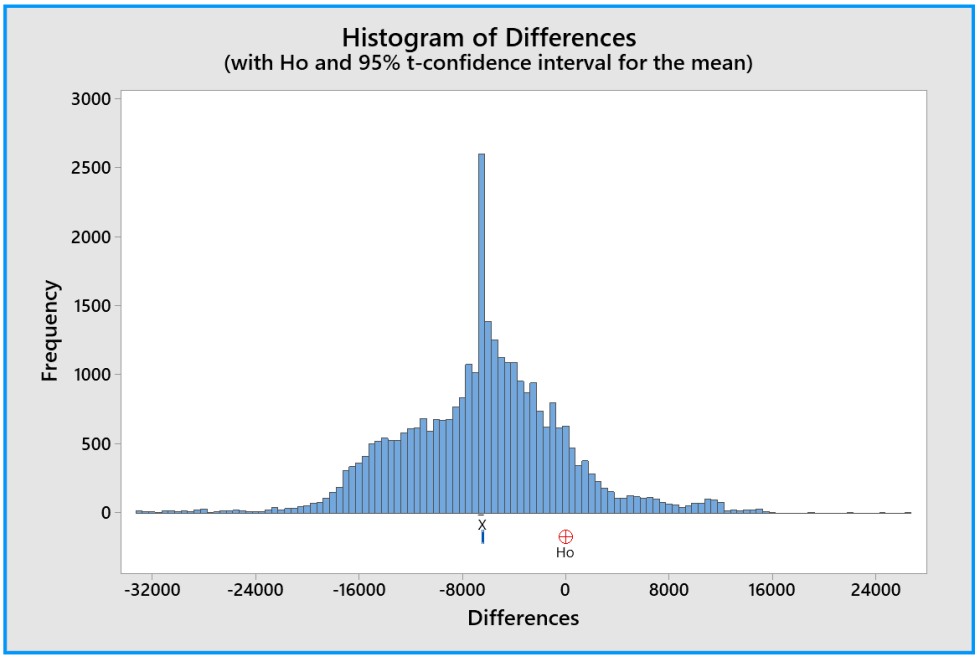


Figure 35: Histogram of difference

Though the Anderson-Darling *P*-value shows a failing test, the two sample data look fairly normal when evaluating the distribution (see appendix E). It is critical to

review the histogram and the normal proability plot to see if a small set of outliers has caused the non-normality, as is the case in this dataset. Furthermore, Ryan-Joiner normality test indicated the strength of the correlation, in addition to data transformation checks using Box-Cox and Johnson. The correlation coefficient for both dataset is near 1, which is an indication that the population is likely to be normal (refer to Fig. 35 above). Based on the analysis above, the maintenance actions proposed is discussed in the section below.

4.7 Proposed Approaches and Maintenance Actions

Integrated prognostics approaches focus mainly on premature failures of turbine gearboxes, and as such condition-based maintenance actions are highly data driven and involve rotor-alignment checks. High fluctuations in wind speed, temperature, and wind direction, which could trigger stops and starts in the pitch-control system, must be monitored thoroughly alongside the gear oil temperature. For instance, consider a wind farm with (1) a turbine with an already failed gearbox *i* and a working rotor system *j* that require adjustment, and (2) a turbine with a running rotor system *k* and gearbox *n*.

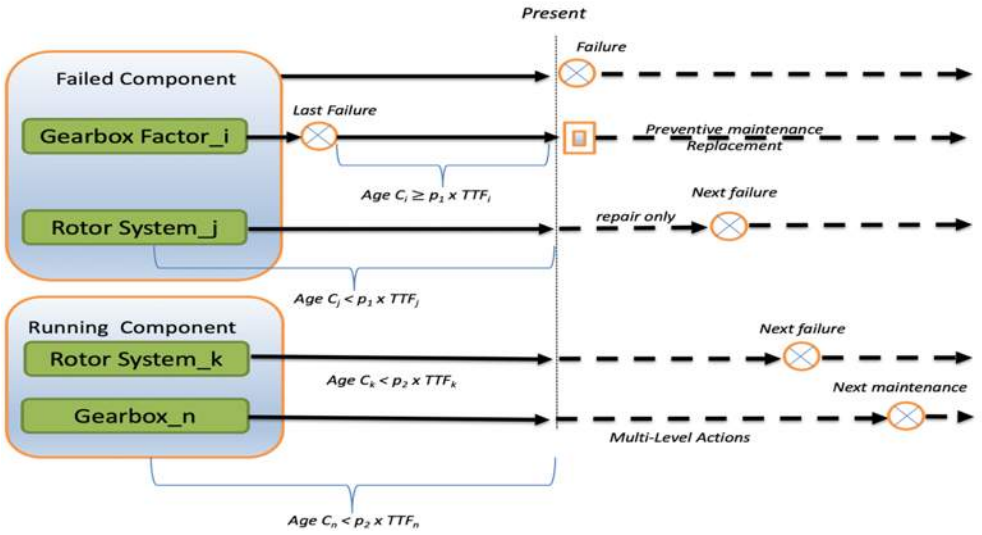


Figure 36: Proposed maintenance concept with multi-level actions.

The proposed maintenance approach, as illustrated in figure 35, is a hybrid of an age-based policy used by [59] for military aero-engines and Mohamed-Salah et al.'s [60] opportunistic maintenance approach taken between the preventive maintenance time and the instant of failure. Operators should complete an extra layer of checks during maintenance, in addition to traditional oil and vibration monitoring. A cost simulation with maximum iterations on various wind turbines must be performed before a CBM decision is made. The simulation of cost evaluation is not a part of this research but is essential for decision making to maintain other closely related turbines in a more cost-concious approach. A single maintenance action could become expensive if not coupled with other levels of actionable maintenance requiring components at the same turbine or neighboring turbine locations.

Chapter 5—Discussion, Conclusions, and Recommendations

5.1 Discussion

Given the demand for and growth in wind energy, it is obvious that the need for increasing efficiency could lead to the development of a wind turbine with a larger rotor system and hub height over 150 m. Thus, unexpected early failure of the gearbox component could present a challenge which would need to be addressed. The operation and maintenance of wind farms is always an area of interest for researchers seeking to optimize and improve the reliability of wind-turbine component systems in general.

Integrated prognostics is an approach that merges data-driven methods with physics-based methodology, taking into account rotor-systems factors such as blade-pitching angles and the amount of exerted torque on the gearbox, along with existing environmental and gearbox factors.

In this chapter, we conclude our discussion of this approach, and generalize its applicability to other WT components and to the body of knowledge in related industries. We also suggest several future directions for study, as well as potential improvements to the efficiency of the model discussed in this praxis, and to the industry available oil and vibration-based gearbox maintenance actions.

5.2 Conclusions

Operation and maintenance management in the deployment of wind energy have significant impacts on the overall levelized cost of energy. Maintenance optimization is essential and has become the most pertinent research area to reducing the number of wind-turbine component failures, including, but not limited to, failures of the gearbox

component. The main objective of this praxis is to develop an integrated maintenance approach to meet challenges in gearbox time-to-failure that the industry currently faces during the early stage of the turbine life cycle.

The current industry trend in monitoring the health of gearbox components is through oil- and vibration-based monitoring systems, and by implementing conventional corrective, time-based, and condition-based preventive maintenance strategies. This research introduces the approach of integrating prognostic CBM during the early stage of the turbine life cycle.

Based on the results of the analysis explained in Chapter 4, it is evident that environmental, rotor-system, and gearbox factors have strong predictive power in gearbox failures, blade-pitch angles and gear-oil temperature in particular. Also, it is observed that in addition to a regular maintenance schedule, optimizing rotor systems using novel pitch control along with gear-oil temperature adjustment can reduce the chance of premature failure of the wind turbine's gearbox by manipulating the angle at which the blade pitches in delaying the extra load exerted on the rotor system.

5.3 Contributions to Body of Knowledge

This praxis has contributed to the body of knowledge in the area of three related components below:

1. Insufficient research has been conducted on the topic of utilizing an integrated prognostics approach merging material science, data science, and industry-available approaches. The same analogy could also be applied to additional components of the wind-turbine system, such as

blades, generators, main bearings, and others, resulting in the establishment of a more optimized maintenance system.

2. The approach discussed fills the data gap from which the wind industry suffers by integrating the actual physical and material science of failing components. Despite the age of the industry, this analysis is very cohesive with scientific and historical data. The model is most understandable if the phenomenon is explained as a challenge.
3. The theory deliniated is logically consistent and explains the relationship between factors used and actual material science. It proposes a framework for optimized maintenance actions based on predictive models combining various variables and parameters.

5.4 Recommendations for Future Research

Recommendations for future expansion and improvement on this research's limitations are given below.

- The cost feasibility of the suggested integrated prognostics maintenance action was not considered, as cost-effectiveness of such maintenance strategies should be evaluated for the entire wind farm and included in matrix simulations for best decision-making.
- In the finite-element analysis of the spur gear itself, the varying low- and high-speed shaft of the drivetrain can be considered along with the varying load from the rotor system, in order to address the impact of these variations to the turbine which may affect stress intensity and gear-crack

propagation.

- In the physical model of the gear, it would be possible to model the entire drivetrain system rather than only the stand-alone spur gear, which would account for more accurate load transfer and better contact force on the gear-tooth profile.
- More historical health-monitoring data from wind turbines that utilize novel pitch-control systems for large turbines can be utilized to test the proposed method's validity, and improve its efficiency. More gearbox types need to be studied under varying environmental conditions in order to reach further hypothesized conclusions.

References

[1] Ozturk, S., Fthenakis, V. & Faulstich, S. (2018). Failure modes, effects and criticality analysis (FMECA) for wind turbines considering climatic regions and comparing geared and direct drive wind turbines. *Energies Journal*, doi: 10.3390/en1109 2317 (Journal Impact Factor: 2.707).

[2] Settanni, E. (2016): Applying forgotten lessons in field reliability data analysis to performance-based support contracts. *Journal of Engineering Management*, 28(1) (Journal Impact Factor: 0.548)

[3] Garabedian, N.T., Gould, B.J., Doll, G.L.. & Burris, D.L. (2018). The cause of premature wind turbine bearing failures: Overloading or underloading. *Journal of Tribology – Transaction of the ASME*, doi: 10.1080/10402004.2018.1433345 (Journal Impact Factor: 1.648).

[4] Jiang, A., Wang Y., & Cheng, Y. (2018). A condition-based opportunistic maintenance policy integrated with energy efficiency for two-component parallel systems. *Journal of industrial Engineering and Management*.

[5] F. Ding, Z. Tian, F. Zhao, and H. Xu. (2018). An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions. *Renew. Energy* vol. 129, no. Part A, pp. 260-270.

[6] Heli and Huaiju Liu, Peitang Wei and Haifeng He (2018). *Evaluation of Contact Fatigue Life of a Wind Turbine Gear Pair Considering Residual Stress*. Journal of Tribology – Transaction of the ASME, doi: 10.1115/1.4039164 (Journal Impact Factor: 1.648).

[7] International Renewable Energy Agency (IRENA) (2012), *Renewable Energy*

Technologies Cost analysis series. Vol.1 Power sector Issue 5/5.

[8] Rommerl, D. P., Di Maio, D., & Tinga, T. (2018). Calculating wind turbine component loads for improved life prediction. *Journal of Renewable Energy*. doi:10.1016/j.renene.2019.07.131 (Journal Impact Factor: 1.511).

[9] Stehly, T., Beiter, P., & Heimiller, D. (2017). *Cost of wind energy review*. National Renewable Energy Lab., NREL/TP-6A20-72167.

[10] Sheng, S. (2013). *Report on wind turbine reliability—a survey of various databases.* National Renewable Energy Laboratory, NREL/PR-5000-59111.

[11] Merizalde, Y., Hermandez-Callejo, L., Duque-Perez, O., & Alonso-Gomez, V. (2019). Maintenance models applied to wind turbines. A comprehensive overview. *Energies Journal, 12*(2), 25. doi:10.3390/en12020225 (Journal Impact Factor: 2.707).

[12] Tavner, P. J., Xiang, J., & Spinato, F. (2007). Reliability analysis for wind turbines. *Wind Energy, 10*(1). doi: 10.1002/we.204 (Journal Impact Factor: 3.125).

[13] Poore, R., & Letternmaier, T. (2003). *Alternative design study report: WindPACT advanced wind turbine drivetrain designs*. National Renewable Energy Laboratory (NREL).

[14] Department of Energy. (2010). *Advanced wind turbine drivetrain concepts.* DOE/GO-102010-3198.

[15] Schubel, P. J., & Crossley, R. J. (2012). Wind turbine blade design. *Energies Journal, 12*(5), 3425–3449. doi:10.3390/en5093425 (Journal Impact Factor: 2.707).

[16] Lloyd, G. (2010). *Guideline for the certification of wind turbines*. Germanischer

Lloyd Std. Renewables certification, Hamburg, Germany.

[17] Philippidis, T. P., Lekon, D. J., & Baharoudis, C. (2006). *Assessment of failure probability under uni-axial and multi-axial static and fatigue load*. University of Patras, TU Delft, Optimate Blades TG 2.

[18] Manwell, J., McGowan, J., & Rogers, A. (2009). *Wind energy explained: Theory, design and application* (2nd ed.). John Wiley & Sons.

[19] Hodges, D. H., Volovoi, V., & Li, L. (2007). Cross-sectional design of composite rotor blades considering manufacturing constraints. *Proceedings of the 63rd Annual Forum of the American Helicopter Society.* doi:10.4050/JAHS.53.240

[20] Ganfuli, R. (2004). Survey of recent developments in rotorcraft optimization. *Journal of Aircraft, 41*(3), 493–410.

[21] Nijssen, R. (2006). *Fatigue life prediction and strength degradation of wind turbine rotor blade composites* [Unpublished doctoral dissertation]. Delft University of Technology.

[22] Lucente, M. (2008). *Condition monitoring system in wind turbine gearbox* (XR-EE-ETK 2008:4). [Master's thesis, KTH Royal Institute of Technology]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.528.4712&rep=rep1&type=pdf

[23] Burton, T. (2011). *Wind energy handbook*. John Wiley & Sons, Ltd.

[24] Adam, T., Gathercole, N., Reiter, H., & Harris, B. (1994). Life prediction for fatigue of T800/5245 carbon fibre composites: II-variable amplitude loading. *International Journal of Fatigue, 16,* 533–457.

[25] Zhao, M., & Ji, J. (2016). Dynamic analysis of wind turbine gearbox components.

Energies Journal, 16(9), 110. doi:10.3390/en902010 (Journal Impact Factor: 2.707).

[26] Sethuraman, L., Guo, Y., & Sheng, S. (2015). *Main bearing dynamics in three-point suspension drivetrains for wind turbines*. National Renewable Energy Laboratory. NREL/PR-5000-64311.

[27] Wang, R., Han, T., & Wang, W. (2018). Fracture analysis and improvement of the main shaft of wind turbine based on finite element method. *Advances in Mechanical Engineering, 10*(4). doi.org/10.1177/1687814018769003 (Journal Impact Factor: 1.024).

[28] Irwin, G. R. (1957). Analysis of stress and strains near the end of a crack traversing a plate. *Journal of Applied Mechanics, 24,* 361–364.

[29] Westergaard, H. M. (1939). Bearing pressures and cracks. *Journal of Applied Mechanics, 6*, A49–53.

[30] Danish Wind Industry Association. (2000). *Gearbox wind turbines.* http://www.windpower.org/powtrain.htm

[31] Wan, S., Cheng, L., & Sheng, X. (2015). Effects of yaw error on wind turbine running characteristics based on the equivalent wind speed model. *Energies*, *8*(7), 6286–6301. doi:10.3390/en8076286 (Journal Impact Factor: 2.707).

[32] Wang, N., Wright, A. D., & Johnson, K. E. (2016). *Independent blade pitch controller design for a three-bladed turbine using disturbance accommodating control*. National Renewable Energy Laboratory. NREL/CP-5000-66011.

[33] Namik, H., & Stol, K. (2013). Individual blade pitch control of a spar-buoy floating wind turbine. *IEEE Transactions on Control System Technology, 22*(1), 214–223.

[34] Vattenfall, J., & Autoinvent, N. B. (2016). *Wind turbine gearboxes: Maintenance effect on present and future gearboxes for wind turbines.* Energiforskrapport.

[35] Musial, W., Butterfield, S., & McNiff, B. (2007). *Improving wind turbine gearbox reliability.* National Renewable Energy Laboratory. NREL/PR-5000-41548.

[36] Guo, Y., & Sheng, S. (2018). *Probability of failure for gearbox high-speed-stage bearings in wind turbines*. National Renewable Energy Laboratory. NREL/PR-5000-70972.

[37] Zaher, A., McArthur, S., Infield, D., & Patel, Y. (2009). Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy, 12*(6), 574–593.

[38] Dalmau, J., Gimenez, G., & De Castro, R. (2016). ISO 9001 aspects related to performance and their level of implementation. *Journal of Industrial Engineering and Management, 9*(5). doi:10.3926/jiem.2072 (Journal impact factor: 1.025).

[39] Saidi, L., Ali, J., Benbouzid, M., & Bechhofer, E. (2018). An integrated wind turbine failures prognostic approach implementing Kalman smoother with confidence bounds. *Applied Acoustics*, *138,* 199–208. doi:10.1016/j.apacoust.2018.04.005

[40] Stehly, T., Beiter, P., & Heimiller, D. (2017). *2017 cost of wind energy review*. National Renewable Energy Laboratory. NREL/TP-6A20-72167

[41] Qin, R., Grasman, S., Long, S., Lin, Y., & Thomas, M. (2012). A framework of cost-effectiveness analysis for alternative energy strategies. *Engineering Management Journal, 24*(4), 18–35. (Journal impact factor: 2.707).

[42] Bi, R., Qian, K., Zhou, C., Hepburn, D., & Rong, J. (2014). A survey of failures in wind turbine generator systems with focus on a wind farm in China. *International*

Journal of Smart Grid and Clean Energy. doi:10.12720/sgce.3.4.366-373

[43] Mazzuchi, T. A., Van Noortwijk, J. M., & Kallen, M. J. (2002). Maintenance optimization. doi:10.1002/9780470061572

[44] Mazzuchi, T. A. (1989). The proportional hazards model in reliability. *Proceedings of the Annual Reliability and Maintainability Symposium.*

[45] U. S. Department of Energy. (2015). *Wind vision*: *A new era for wind power in the United States* (Technical report, DOE/GO-102015-4557). http://energy.gov/sites/prod/files/2015/03/f20/wv_full_report.pdf.

[46] Sheng, S. (2014). *Gearbox typical failure modes, detection and mitigation methods.* National Renewable Energy Laboratory, NREL/PR-5000-60982.

[47] Gray, C. S., & Watson, S. J. (2018). *Physics of failure approach to wind turbine condition based maintenance*. Centre for Renewable Energy System Technology, Electronic and Electrical Engineering Department.

[48] Eriki, A., Ravichandra, R., & Mustaffa, M. E. (2012). Spur gear crack propagation path analysis using finite element method. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2012* (Vol II), Hong Kong.

[49] Charri, F., Fakhfakh, T., & Haddar, M. (2009). Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness. *European Journal of Mechanics a/Solids, 28*, 461–468.

[50] Mazzuchi, T. A., & van Dorp, J. R. (2011). A Bayesian expert judgement model to determine lifetime distributions for maintenance optimization. *Structure and Infrastructure Engineering, 8*(4), 307–315. doi: 10.1080/15732479.2011.563084

[51] Altamura, A., & Straub, D. (2014). Reliability assessment of high cycle fatigue under variable amplitude loading: Review and solutions. *Engineering Fracture Mechanics, 121,* 40–66.

[52] Paris, P. C., & Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Basic Engineering, 85*(4), 528–533.

[53] Keller, J., Sheng, S., & Cotrell, J. (2016). *Wind turbine drivetrain reliability collaborative*. National Renewable Energy Laboratory.

[54] Coronado, D., & Fischer, K. (2015). *Condition monitoring of wind turbines*: *State of art, user experience and recommendations.* VGB Research Foundation, Fraunhofer IWES.

[55] Penn State, Eberly College of Science: Applied Regression Analysis—Detecting multicollinearity using variance inflation factors. https://online.stat.psu.edu/stat462/node/180/

[56] A. Hoskuldson (1998): " PLS Regression Methods," *Journal of Chemometrics*, 2, 211-228.

[57] Jin-Rae Cho, Ki-Yong Jeong, Min-Ho Park, Dae-Suk Shin, O-Kaung Lim, and No-Gil Park (2013): Finite element structural analysis of wind turbine gearbox considering tooth contact of internal gear system. *Journal of mechanical science and technology*. doi:10.1007/s12206-013-0521-0

[58] Tian, Z.;Zuo, M.J.; Wu,S. (2012): *Crack propagation assessment for apur gears using model-based analysis and simulation*. J.Intell. Mnauf. 239-253.[CrossRef]

[59] J. Croker and U.D. Kumar (2000): *Age-related maintenance versus reliability*

centered maintenance, a case study on aero-engines, Reliab. Eng. Syst. Saf., Vol. 67, no. 2, pp. 113-118.

[60] K. Daoud, Mohamed -Salah O. and G. Ali (1999): *A simulation model for opportunistic maintenance strategies*, In Emergeing Technology and Factory automation, vol. 1, pp. 703-708.

[61] Endeshaw, H., Stephen E. Osire, Alemayehu, F., Dias, J. (2017): Evaluation of fatigue crack propagation of gears considering uncertainities in loading and material properties, *journal of sustainability*.doi:10.3390/su9122200.

[62] Brownlee, Jason (2019): Machine Learning Mastery with Weka. WEKA, University of Waikato.

Appendix A

Overview of Input Variable and Trend

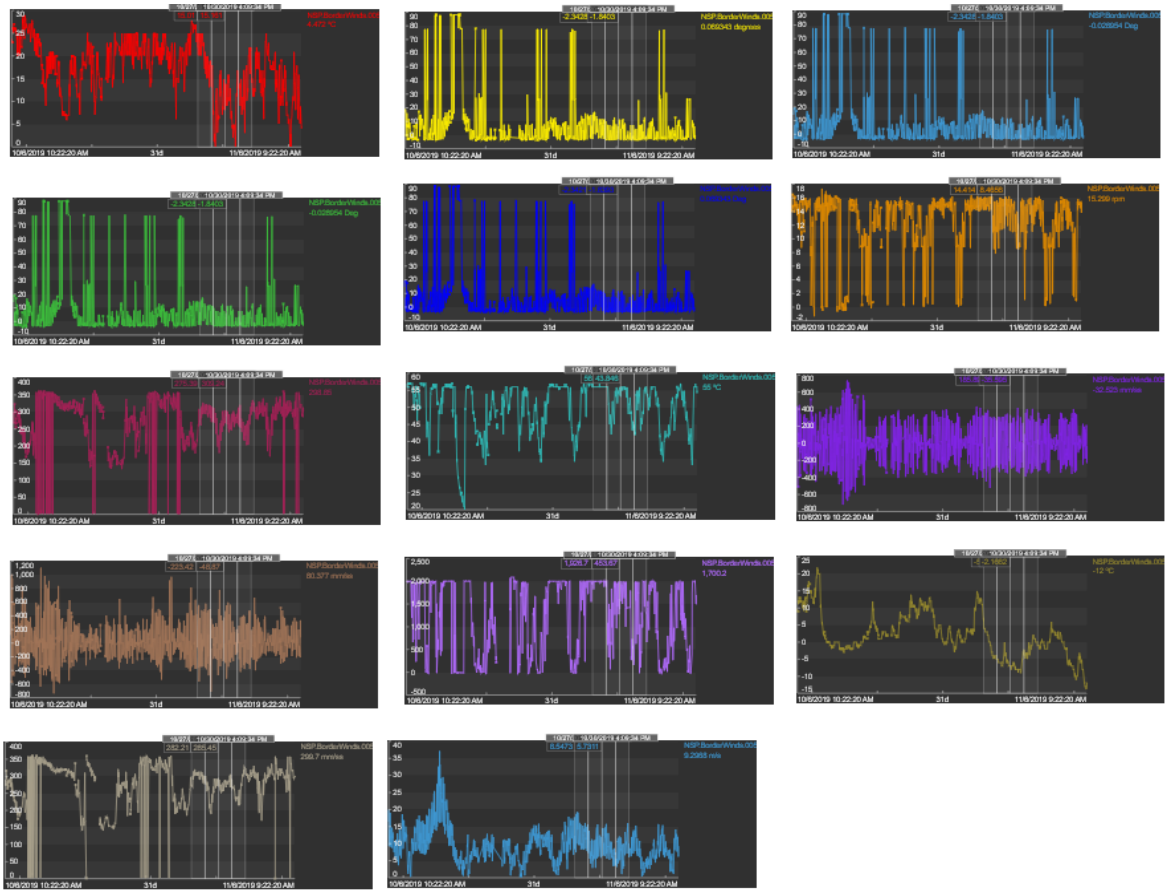
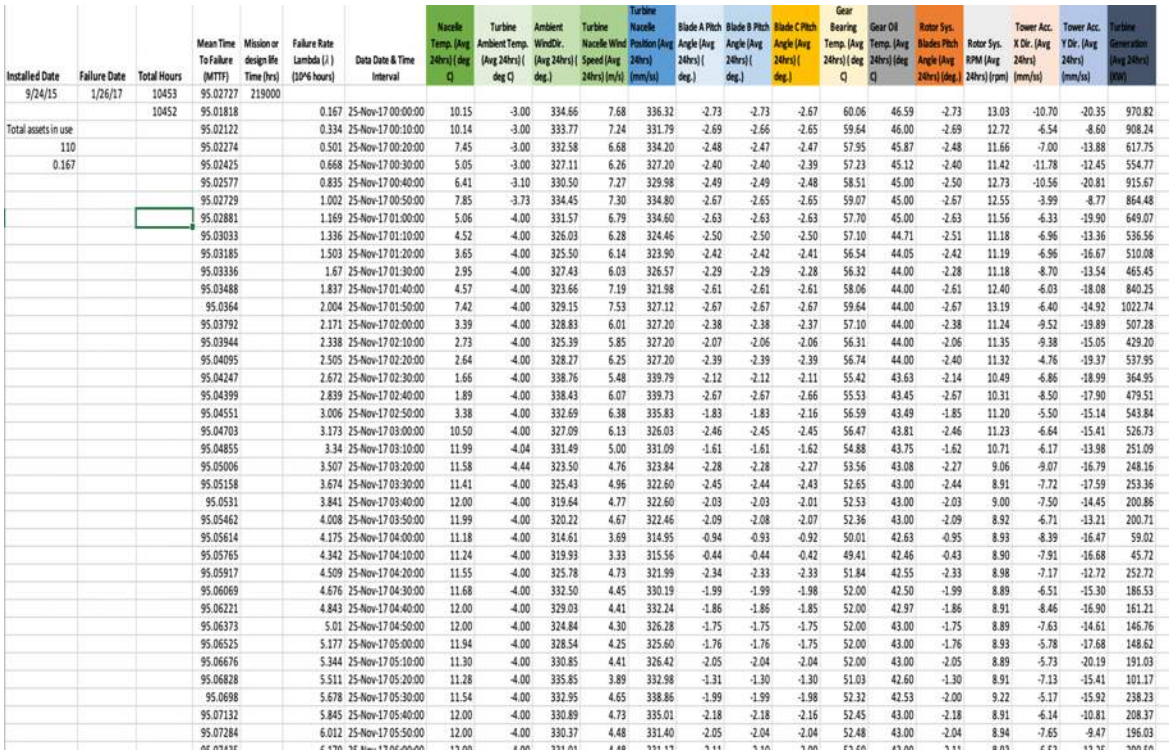


Figure A_0: Historical Data trend – PI Vision



Installed Date	Failure Date	Total Hours	Mean Time To Failure (MTTF)	Mission or design life Time (hrs)	Failure Rate Lambda (λ) (10^6 hours)	Data Date & Time Interval	Nacelle Temp. (Avg 24hrs) (deg C)	Turbine Ambient Temp. (Avg 24hrs) (deg C)	Ambient WindDir. (Avg 24hrs) (deg.)	Turbine Nacelle Wind Speed (Avg 24hrs) (m/s)	Turbine Nacelle Position (Avg 24hrs) (mm/ss)	Blade A Pitch Angle (Avg 24hrs) (deg.)	Blade B Pitch Angle (Avg 24hrs) (deg.)	Blade C Pitch Angle (Avg 24hrs) (deg.)	Gear Bearing Temp. (Avg 24hrs) (deg C)	Gear Oil Temp. (Avg 24hrs) (deg C)	Rotor Sys. Blades Pitch Angle (Avg 24hrs) (deg.)	Rotor Sys. RPM (Avg 24hrs) (rpm)	Tower Acc. X Dir. (Avg 24hrs) (mm/ss)	Tower Acc. Y Dir. (Avg 24hrs) (mm/ss)	Turbine Generation (Avg 24hrs) (KW)
9/24/15	1/26/17	10453	95.02727	219000																	
		10452	95.01818		0.167	25-Nov-17 00:00:00	10.15	-3.00	334.66	7.68	336.32	-2.73	-2.73	-2.67	60.06	46.59	-2.73	13.03	-10.70	-20.35	970.82
Total assets in use			95.02122		0.334	25-Nov-17 00:10:00	10.14	-3.00	333.77	7.24	331.79	-2.69	-2.66	-2.65	59.64	46.00	-2.69	12.72	-6.54	-8.60	908.24
110			95.02274		0.501	25-Nov-17 00:20:00	7.45	-3.00	332.58	6.68	334.20	-2.48	-2.47	-2.47	57.95	45.87	-2.48	11.66	-7.00	-13.88	617.75
0.167			95.02425		0.668	25-Nov-17 00:30:00	5.05	-3.00	327.11	6.26	327.20	-2.40	-2.40	-2.39	57.23	45.12	-2.40	11.42	-11.78	-12.45	554.77
			95.02577		0.835	25-Nov-17 00:40:00	6.41	-3.10	330.50	7.27	329.98	-2.49	-2.49	-2.48	58.51	45.00	-2.50	12.73	-10.56	-20.81	915.67
			95.02729		1.002	25-Nov-17 00:50:00	7.85	-3.73	334.45	7.30	334.80	-2.67	-2.65	-2.65	59.07	45.00	-2.67	12.55	-3.99	-8.77	864.48
			95.02881		1.169	25-Nov-17 01:00:00	5.06	-4.00	331.57	6.79	334.60	-2.63	-2.63	-2.63	57.70	45.00	-2.63	11.56	-6.33	-19.90	649.07
			95.03033		1.336	25-Nov-17 01:10:00	4.52	-4.00	326.03	6.28	324.46	-2.50	-2.50	-2.50	57.10	44.71	-2.51	11.18	-6.96	-13.36	536.56
			95.03185		1.503	25-Nov-17 01:20:00	3.65	-4.00	325.50	6.14	323.90	-2.42	-2.42	-2.41	56.54	44.05	-2.42	11.19	-6.96	-16.67	510.08
			95.03336		1.67	25-Nov-17 01:30:00	2.95	-4.00	327.43	6.03	326.57	-2.29	-2.29	-2.28	56.32	44.00	-2.28	11.18	-8.70	-13.54	465.45
			95.03488		1.837	25-Nov-17 01:40:00	4.57	-4.00	323.66	7.19	321.98	-2.61	-2.61	-2.61	58.06	44.00	-2.61	12.40	-6.03	-18.08	840.25
			95.0364		2.004	25-Nov-17 01:50:00	7.42	-4.00	329.15	7.53	327.12	-2.67	-2.67	-2.67	59.64	44.00	-2.67	13.19	-6.40	-14.92	1022.74
			95.03792		2.171	25-Nov-17 02:00:00	3.39	-4.00	328.83	6.01	327.20	-2.38	-2.38	-2.37	57.10	44.00	-2.38	11.24	-9.52	-19.89	507.28
			95.03944		2.338	25-Nov-17 02:10:00	2.73	-4.00	325.39	5.85	327.20	-2.07	-2.06	-2.06	56.31	44.00	-2.06	11.35	-9.38	-15.05	429.20
			95.04095		2.505	25-Nov-17 02:20:00	2.64	-4.00	328.27	6.25	327.20	-2.39	-2.39	-2.39	56.74	44.00	-2.40	11.32	-4.76	-19.37	537.95
			95.04247		2.672	25-Nov-17 02:30:00	1.66	-4.00	338.76	5.48	339.79	-2.12	-2.12	-2.11	55.42	43.63	-2.14	10.49	-6.86	-18.99	364.95
			95.04399		2.839	25-Nov-17 02:40:00	1.89	-4.00	338.43	6.07	339.73	-2.67	-2.67	-2.66	55.53	43.45	-2.67	10.31	-8.50	-17.90	479.51
			95.04551		3.006	25-Nov-17 02:50:00	3.38	-4.00	332.69	6.38	335.83	-1.83	-1.83	-2.16	56.59	43.49	-1.85	11.20	-5.50	-15.14	543.84
			95.04703		3.173	25-Nov-17 03:00:00	10.50	-4.00	327.09	6.13	326.03	-2.46	-2.45	-2.45	56.47	43.81	-2.46	11.23	-6.64	-15.41	526.73
			95.04855		3.34	25-Nov-17 03:10:00	11.99	-4.04	331.49	5.00	331.09	-1.61	-1.61	-1.62	54.88	43.75	-1.62	10.71	-6.17	-13.98	251.09
			95.05006		3.507	25-Nov-17 03:20:00	11.58	-4.44	323.50	4.76	323.84	-2.28	-2.28	-2.27	53.56	43.08	-2.27	9.06	-9.07	-16.79	248.16
			95.05158		3.674	25-Nov-17 03:30:00	11.41	-4.00	325.43	4.96	322.60	-2.45	-2.44	-2.43	52.65	43.00	-2.44	8.91	-7.72	-17.59	253.36
			95.0531		3.841	25-Nov-17 03:40:00	12.00	-4.00	319.64	4.77	322.60	-2.03	-2.03	-2.01	52.53	43.00	-2.03	9.00	-7.50	-14.45	200.86
			95.05462		4.008	25-Nov-17 03:50:00	11.99	-4.00	320.22	4.67	322.46	-2.09	-2.08	-2.07	52.36	43.00	-2.09	8.92	-6.71	-13.21	200.71
			95.05614		4.175	25-Nov-17 04:00:00	11.18	-4.00	314.61	3.69	314.95	-0.94	-0.93	-0.92	50.01	42.63	-0.95	8.93	-8.39	-16.47	59.02
			95.05765		4.342	25-Nov-17 04:10:00	11.24	-4.00	319.93	3.33	315.56	-0.44	-0.44	-0.42	49.41	42.46	-0.43	8.90	-7.91	-16.68	45.72
			95.05917		4.509	25-Nov-17 04:20:00	11.55	-4.00	325.78	4.73	321.99	-2.34	-2.33	-2.33	51.84	42.55	-2.33	8.98	-7.17	-12.72	252.72
			95.06069		4.676	25-Nov-17 04:30:00	11.68	-4.00	332.50	4.45	330.19	-1.99	-1.99	-1.98	52.00	42.50	-1.99	8.89	-6.51	-15.30	186.53
			95.06221		4.843	25-Nov-17 04:40:00	12.00	-4.00	329.03	4.41	332.24	-1.86	-1.86	-1.85	52.00	42.97	-1.86	8.91	-8.46	-16.90	161.21
			95.06373		5.01	25-Nov-17 04:50:00	12.00	-4.00	324.84	4.30	326.28	-1.75	-1.75	-1.75	52.00	43.00	-1.75	8.89	-7.63	-14.61	146.76
			95.06525		5.177	25-Nov-17 05:00:00	11.94	-4.00	328.54	4.25	325.60	-1.76	-1.76	-1.75	52.00	43.00	-1.76	8.93	-5.78	-17.68	148.62
			95.06676		5.344	25-Nov-17 05:10:00	11.30	-4.00	330.85	4.41	326.42	-2.05	-2.04	-2.04	52.00	43.00	-2.05	8.89	-5.73	-20.19	191.03
			95.06828		5.511	25-Nov-17 05:20:00	11.28	-4.00	335.85	3.89	332.98	-1.31	-1.30	-1.30	51.03	42.60	-1.30	8.91	-7.13	-15.41	101.17
			95.0698		5.678	25-Nov-17 05:30:00	11.54	-4.00	332.95	4.65	338.86	-1.99	-1.99	-1.98	52.32	42.53	-2.00	9.22	-5.17	-15.92	238.23
			95.07132		5.845	25-Nov-17 05:40:00	12.00	-4.00	330.89	4.73	335.01	-2.18	-2.18	-2.16	52.45	43.00	-2.18	8.91	-6.14	-10.81	208.37
			95.07284		6.012	25-Nov-17 05:50:00	12.00	-4.00	330.37	4.48	331.40	-2.05	-2.04	-2.04	52.48	43.00	-2.04	8.94	-7.65	-9.47	196.03

Figure A1: Excel exported 10 min interval datapoints (Screen shot)

Appendix B

FEA Stress Intensity Calculations

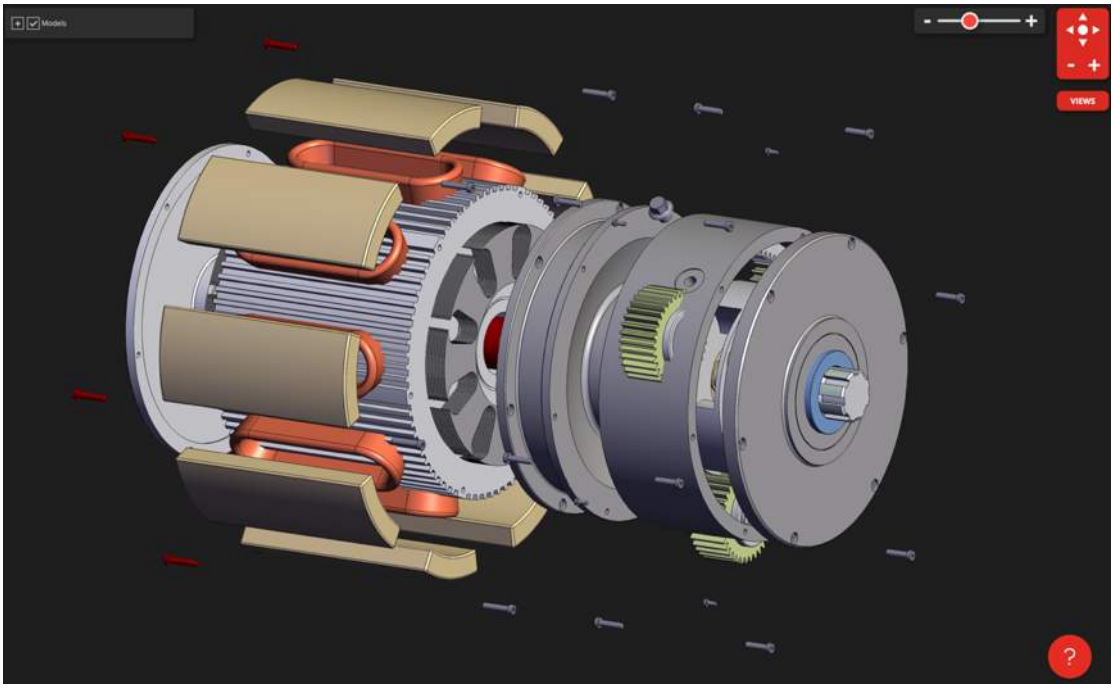


Figure B_0: Gearbox Assembly – Solidworks (CADgrab)

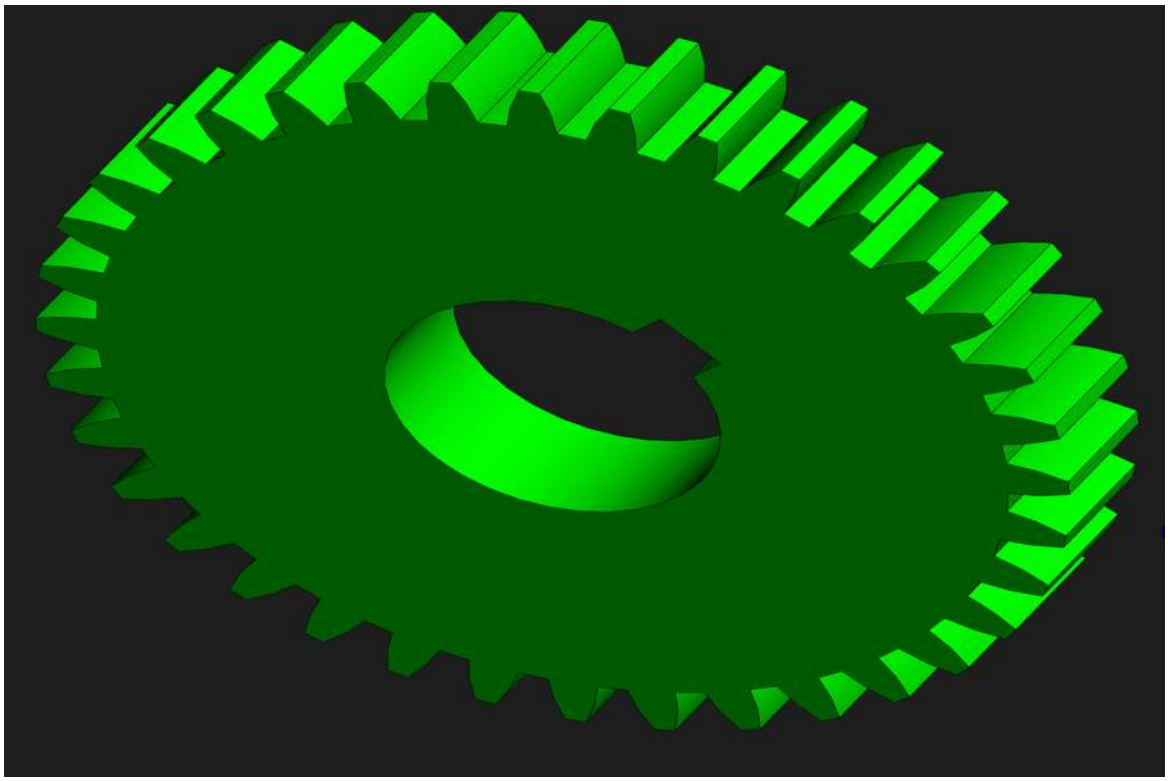


Figure B_1: Gear model (CADgrab)

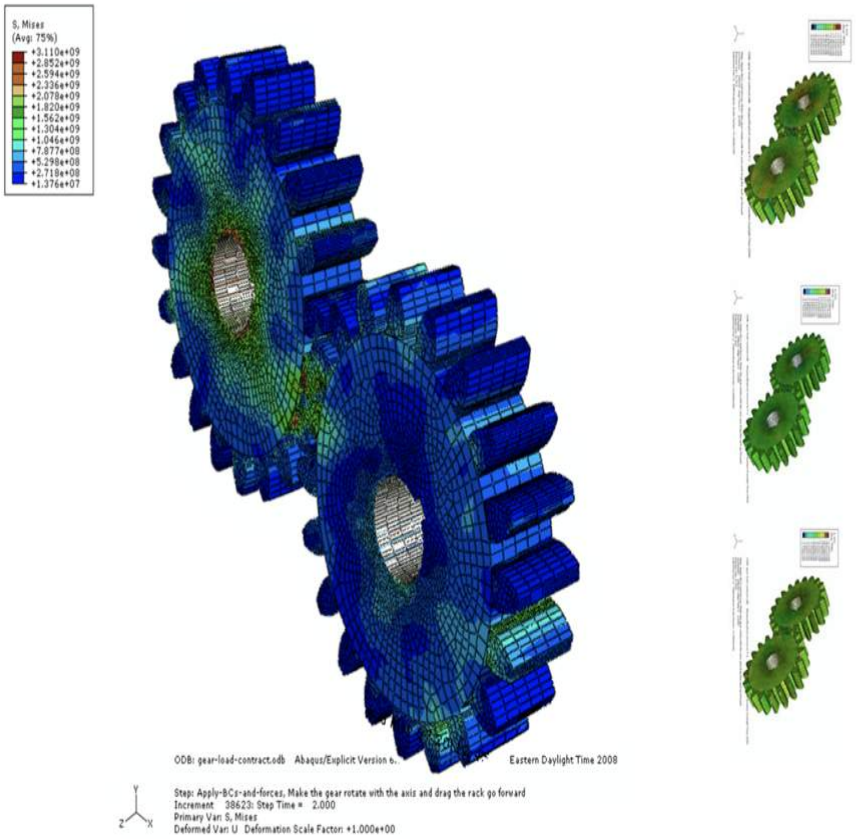


Figure B_2: Gear Stress distribution – FEA Abaqus Result

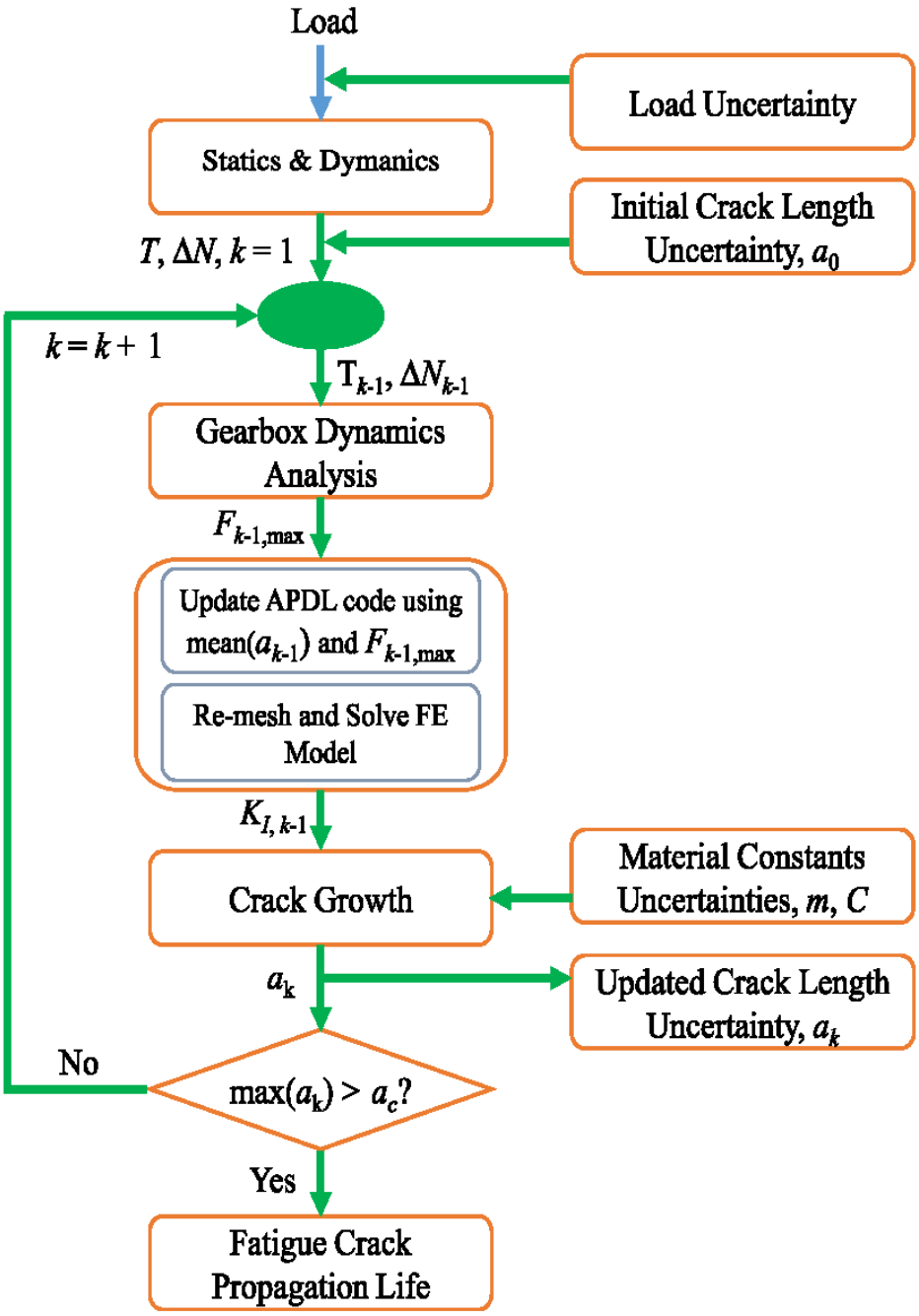


Figure B3: Framework of Fatigue crack propagation of gear [credit: to [61]]

Appendix C

Feature extraction, Data processing and Sliding window algorithms

Extract Features

```
class FeatureExtractor:

    def __init__(self, input_column, dims):
        self._input_column = input_column
        self._dims = dims

    @property
    def input_column(self):
        return self._input_column

    @property
    def output_dims(self):
        return self._dims

    @property
    def output_columns(self):
        raise NotImplementedError()

    def extract(self, input_data):
        raise NotImplementedError()

    def full_pass(self, full_data):
        """Makes a full pass over data.

        Some feature extractions mechanisms require a full pass
        over the data before features can be extracted.
        Args:
            full_data: Pandas DataFrame of the full data.
        """
        pass


class MeanFeatureExtractor(FeatureExtractor):

    def __init__(self, input_column):
        super(MeanFeatureExtractor, self).__init__(input_column, dims=1)

    @property
    def output_columns(self):
        return [self.input_column +
                ' ' +
                self.__class__.__name__.replace('FeatureExtractor', '')]

    def extract(self, input_data):
        # Replace missing value with NaN
        input_data = pd.to_numeric(input_data, errors='coerce')
        input_data = np.nan_to_num(input_data)
        return [np.mean(input_data)]
```

Figure C_0: Feature Extraction Algorithm – Jupyter (Pyhton)

```python
def process_data(data,
                 window_size_minutes,
                 skip_days,
                 feature_extractors,
                 timestamp_column,
                 target_column):
    """Process the given data and extract features by running a sliding window.

    Args:
        data: Pandas DataFrame containing the data
        window_size_days: The size of the sliding window in days.
        skip_days: the number of days to move the sliding window by.
        feature_extractors: a list feature extractors.
        timestamp_column: the name of the timestamp column.
    """
    for feat in feature_extractors:
        if feat.input_column not in data.columns:
            raise ValueError('Column name "%s" does not exist in the table.' % feat.input_column)
    # Initialize feature extractors
    for feat in feature_extractors:
        feat.full_pass(data)

    # Initilize output data
    output_data = []

    # Move the sliding window over the data and extract features.
    for start, end in _sliding_window(
        data[timestamp_column], window_size_minutes, skip_days):
        output_row = []
        # Extract feature and add them to output row.
        for feat in feature_extractors:
            output_row.extend(feat.extract(data[feat.input_column][start:end]))
        # Calculate time to failure
        output_row.append(_calculate_time_to_failure(data, start, end, target_column))
        # Add extracted feature to output_data
        output_data.append(output_row)

    # Create output dataframe
    output_columns = []
    for feat in feature_extractors:
        output_columns.extend(feat.output_columns)
    output_columns.append('Processed ' + target_column)
    return pd.DataFrame(output_data, columns=output_columns)
```

Figure C_1: Data Processing Algorithm – Jupyter (Pyhton)

```
def _sliding_window(timestamp_column_data,
                    window_size_minutes,
                    skip_days):
    start = 0
    end = -1
    n_rows = timestamp_column_data.shape[0]
    for i in range(n_rows):
        delta = timestamp_column_data[i] - timestamp_column_data[start]
        if delta.days >= window_size_minutes:
            end = i
            break
    if end < 0:
        end = n_rows

    while True:
        yield start, end

        # Exit loop if we have reached the end of the input data.
        if end >= n_rows:
            break

        for i in range(start, n_rows):
            delta = timestamp_column_data[i] - timestamp_column_data[start]
            if delta.days >= skip_days or i == n_rows - 1:
                end = min(end + (i - start), n_rows)
                start = i
                break

def _calculate_time_to_failure(data, start, end, target_column):
    if end == 0:
        raise ValueError('end index can not be zero.')
    target_column_data = data[target_column]
    max_target_value = np.amax(target_column_data)
    time_to_failure = max_target_value - target_column_data[end-1]
    return time_to_failure
```

Figure C_2: Sliding Window Algorithm – Jupyter (Pyhton)

```
extractors = [
    MeanFeatureExtractor('X10 ( deg C)'),
    MeanFeatureExtractor('X2 ( deg C)'),
    MeanFeatureExtractor('X3 ( deg.)'),
    MeanFeatureExtractor('X1 (m/s)'),
    MeanFeatureExtractor('X4 (mm/ss)'),
    MeanFeatureExtractor('X7 ( deg.)'),
    MeanFeatureExtractor('X8 ( deg.)'),
    MeanFeatureExtractor('X9 ( deg.)'),
    MeanFeatureExtractor('X11 ( deg C)'),
    MeanFeatureExtractor('X12 (deg C)'),
    MeanFeatureExtractor('X5 (deg.)'),
    MeanFeatureExtractor('X6 (rpm)'),
    MeanFeatureExtractor('X13 (mm/ss)'),
    MeanFeatureExtractor('X14 (mm/ss)'),
    MeanFeatureExtractor('X16 (KW)'),
    MeanFeatureExtractor('X15 (KN/(m)^(3/2))'),
]
```

Figure C_3: Mean Feature Extractor Algorithm – Jupyter (Pyhton)

Appendix D

WEKA Linear Regression algorithms

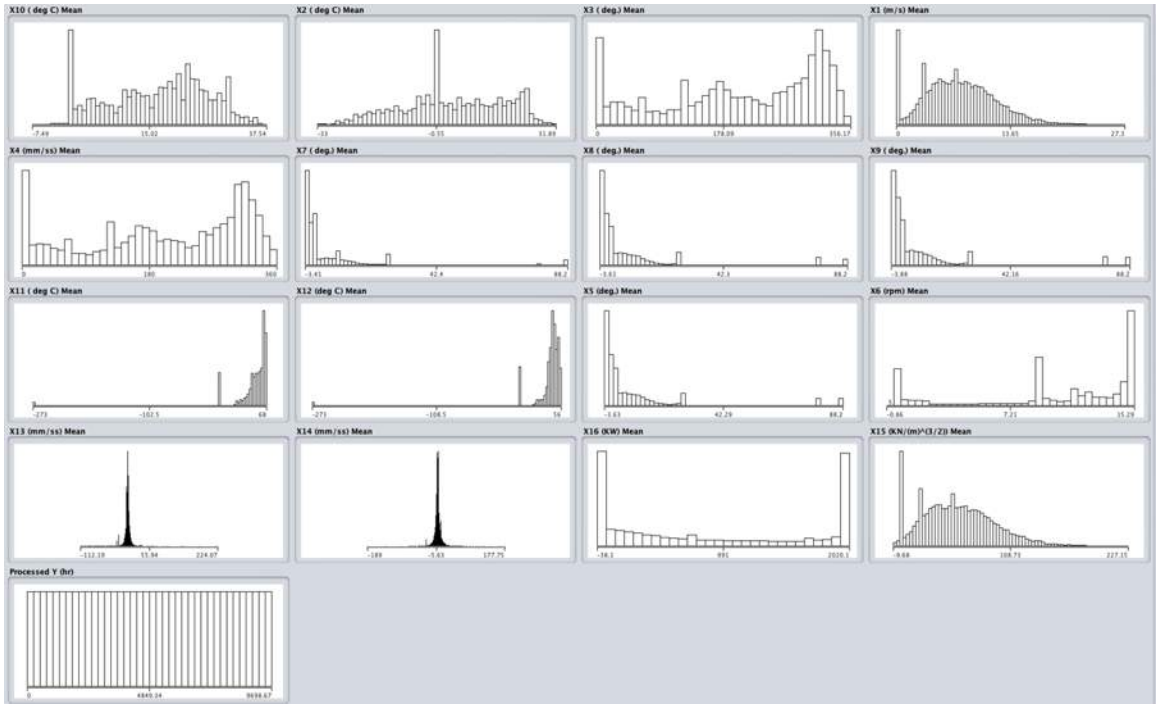


Figure D_0: Dataset: 17 Attribute distributions

Preparation of Dataset View

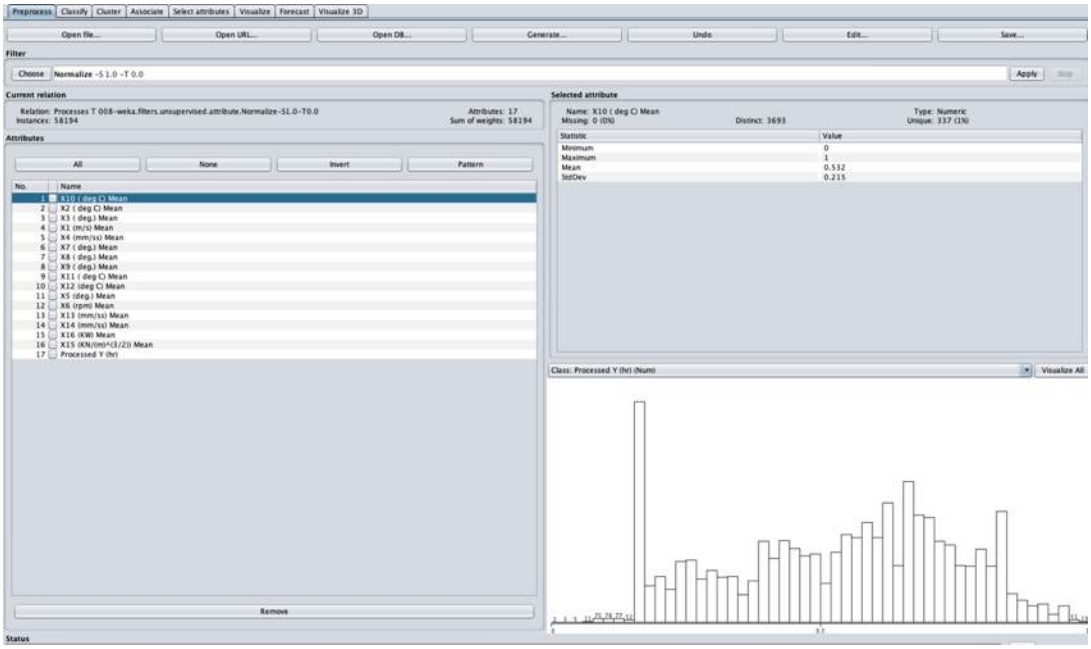


Figure D_1: Normalize Attribute (X_{10})

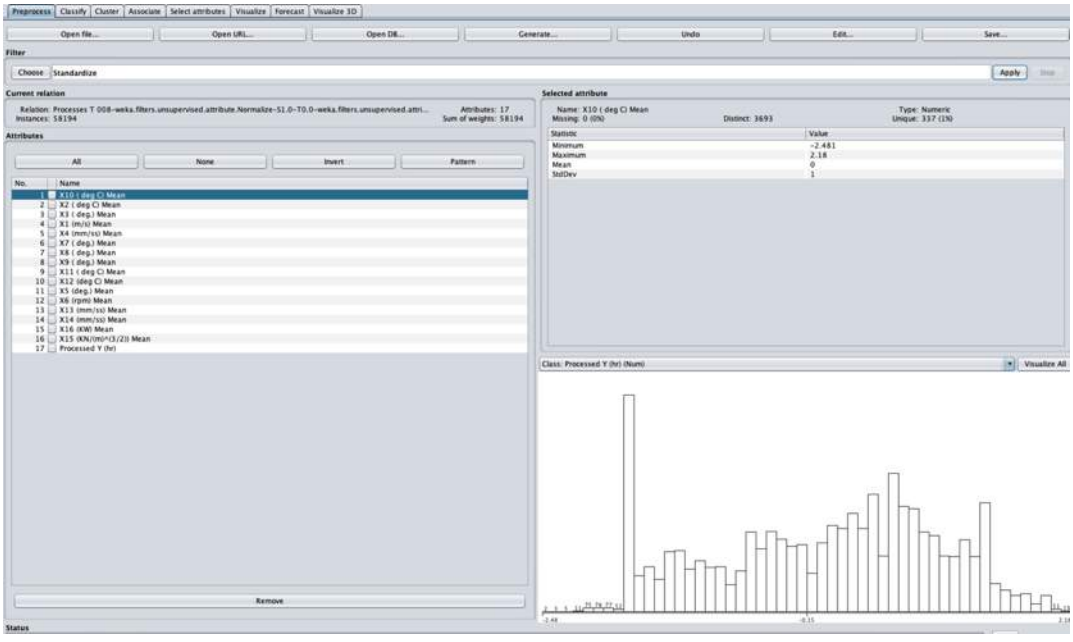


Figure D_2: Standardize Attribute (X_{10})

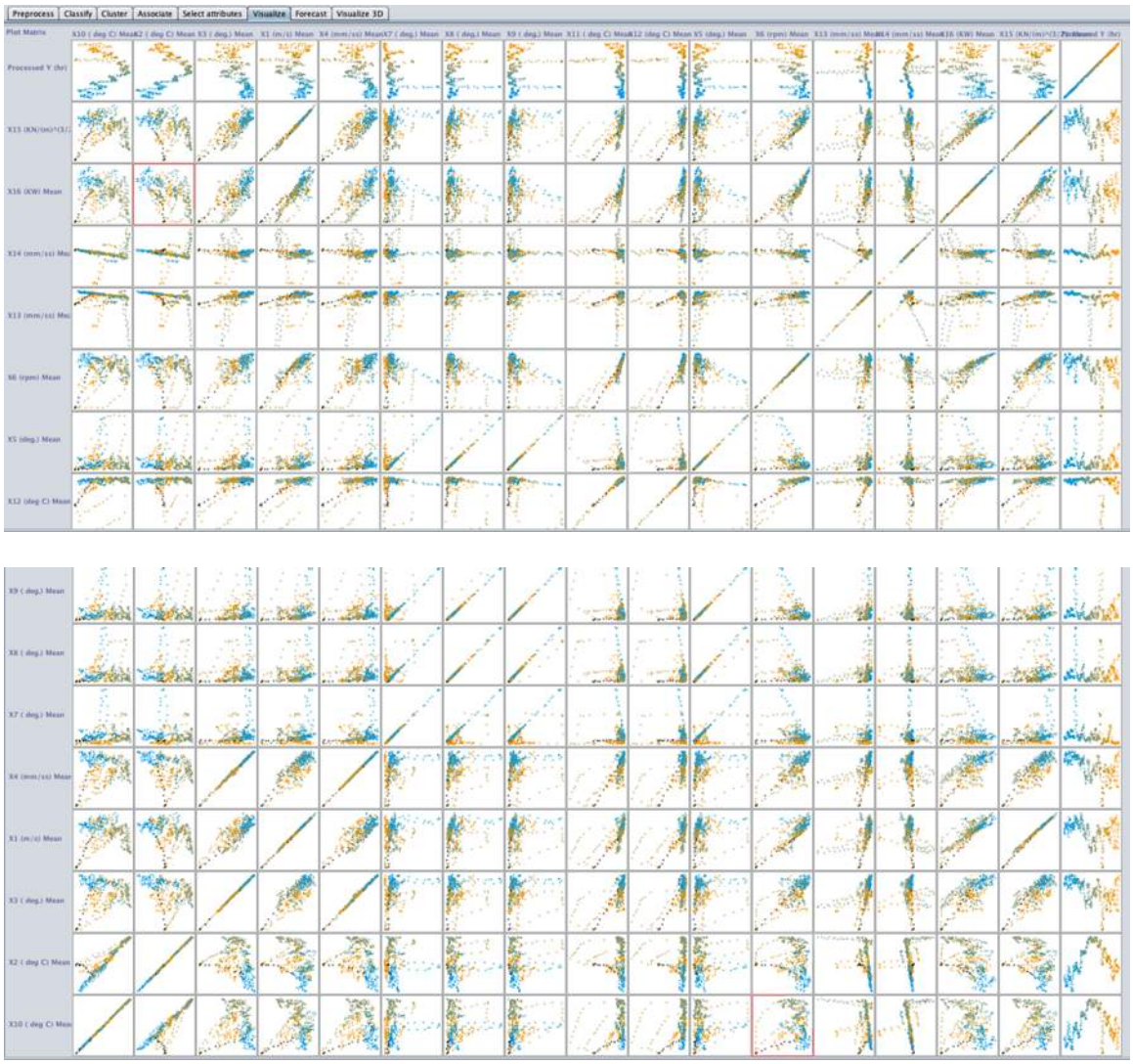


Figure D3: Attribute interaction matrix

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -additional-stats -do-not-check-capabilities -num-decimal-places 4
Relation:     Processes BWF 008-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-unset-class-temporarily-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.unsupervised.attribute.Norma
Instances:    406
Attributes:   17
              X10 ( deg C) Mean
              X2 ( deg C) Mean
              X3 ( deg.) Mean
              X1 (m/s) Mean
              X4 (mm/ss) Mean
              X7 ( deg.) Mean
              X8 ( deg.) Mean
              X9 ( deg.) Mean
              X11 ( deg C) Mean
              X12 (deg C) Mean
              X5 (deg.) Mean
              X6 (rpm) Mean
              X13 (mm/ss) Mean
              X14 (mm/ss) Mean
              X16 (KW) Mean
              X15 (KN/(m)^(3/2)) Mean
              Processed Y (hr)
Test mode:    split 70.0% train, remainder test

=== Classifier model (full training set) ===


Linear Regression Model

Processed Y (hr) =

     -0.8324 * X10 ( deg C) Mean +
      1.0031 * X2 ( deg C) Mean +
     -0.7053 * X3 ( deg.) Mean +
     -1.5918 * X7 ( deg.) Mean +
      1.3918 * X9 ( deg.) Mean +
      1.9172 * X12 (deg C) Mean +
     -0.9505 * X6 (rpm) Mean +
     -0.1408 * X13 (mm/ss) Mean +
     -0.4769 * X14 (mm/ss) Mean +
      0.1861 * X16 (KW) Mean +
      0.2941 * X15 (KN/(m)^(3/2)) Mean +
     -0.0387
```

Figure D4: Regression equation

```
Linear Regression Model

Processed Y (hr) =

     -0.8324 * X10 ( deg C) Mean +
      1.0031 * X2 ( deg C) Mean +
     -0.7053 * X3 ( deg.) Mean +
     -1.5918 * X7 ( deg.) Mean +
      1.3918 * X9 ( deg.) Mean +
      1.9172 * X12 (deg C) Mean +
     -0.9505 * X6 (rpm) Mean +
     -0.1408 * X13 (mm/ss) Mean +
     -0.4769 * X14 (mm/ss) Mean +
      0.1861 * X16 (KW) Mean +
      0.2941 * X15 (KN/(m)^(3/2)) Mean +
     -0.0387

Regression Analysis:

Variable                    Coefficient     SE of Coef        t-Stat
X10 ( deg C) Mean             -0.8324         0.2685         -3.1001
X2 ( deg C) Mean               1.0031         0.2563          3.9139
X3 ( deg.) Mean               -0.7053         0.0856         -8.2356
X7 ( deg.) Mean               -1.5918         0.0892        -17.8428
X9 ( deg.) Mean                1.3918         0.1746          7.9713
X12 (deg C) Mean               1.9172         0.1333         14.3809
X6 (rpm) Mean                 -0.9505         0.2332         -4.0759
X13 (mm/ss) Mean              -0.1408         0.0744         -1.892
X14 (mm/ss) Mean              -0.4769         0.0794         -6.0034
X16 (KW) Mean                  0.1861         0.2035          0.9146
X15 (KN/(m)^(3/2)) Mean        0.2941         0.2178          1.3505
const                         -0.0387         0.169          -0.2288

Degrees of freedom = 58179
R^2 value = 0.7027
Adjusted R^2 = 0.69414
F-statistic = 82.2886

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.8165
Mean absolute error                      0.133
Root mean squared error                  0.1673
Relative absolute error                 13.0365 %
Root relative squared error             17.736  %
Total Number of Instances           58193
Ignored Class Unknown Instances                11
```

Figure D_5: Regression analysis result

Appendix E

Minitab Output Results

The following sample data Y_1 (Post_TTF) and Y_2 (Pre_TTF) are the datas obtained from the regression question of historical data pre-failure of the gearbox and post failure of the gearbox component. Post gearbox failure, novel pitching system adjustment has been made to the rotor system.

	C1	C2
	Y_2 (Pre_TTF)	Y_1 (Post_TTF)
1	3708.8	9698.67
2	3489.0	9698.50
3	1961.6	9698.34
4	2358.5	9698.17
5	3819.4	9698.00
6	2120.4	9697.84
7	4413.2	9697.67
8	3785.2	9697.50
9	4286.7	9697.34
10	4297.8	9697.17
11	4462.2	9697.00
12	4017.1	9696.84
13	4684.6	9696.67
14	6411.4	9696.50
15	8964.9	9696.34
16	11110.1	9696.17

Figure E_0: Paired t-test Input Variable (partial view)

The Figure E1 and E2 below depicts the histogram of the differences between the two sample data and individual value plot.

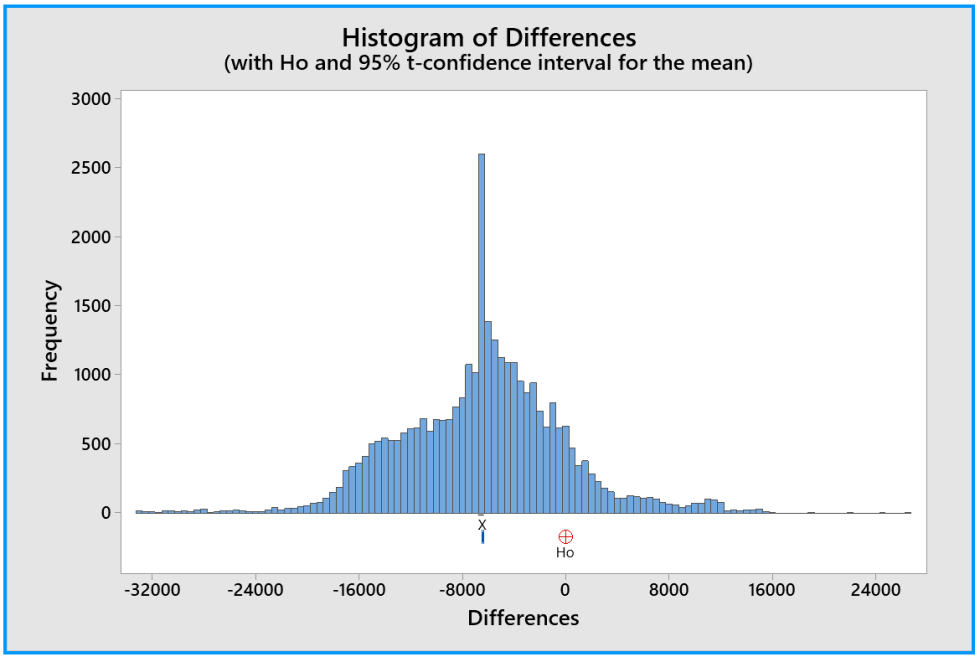


Figure E1: Histogram of Differences

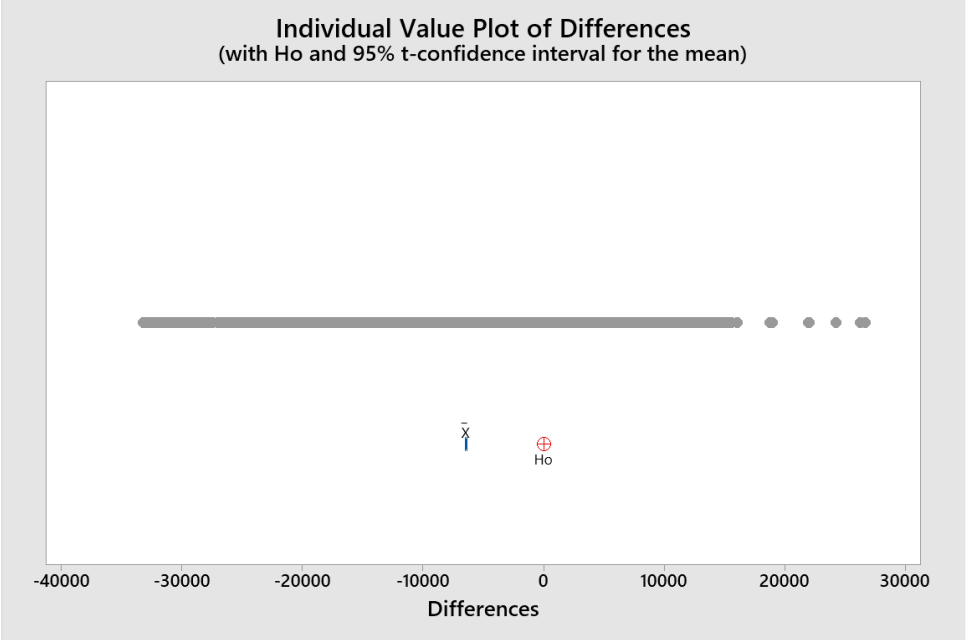


Figure E2: Value Plot of Differences

Clearly, the distribution shows that its normaly distributed tailing out fairly on both sides.

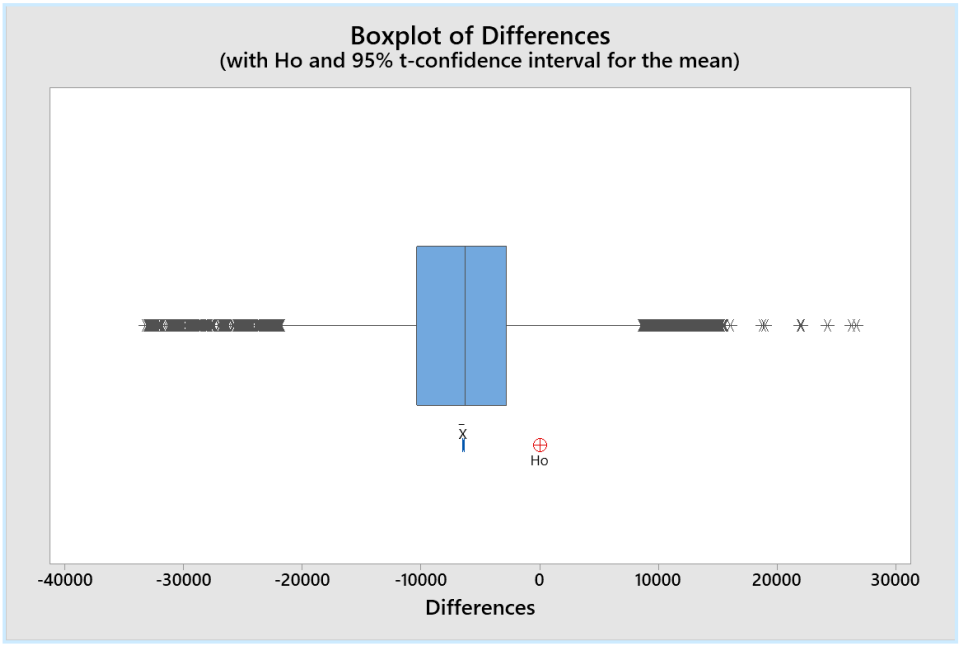


Figure E3: Boxplot of Differences

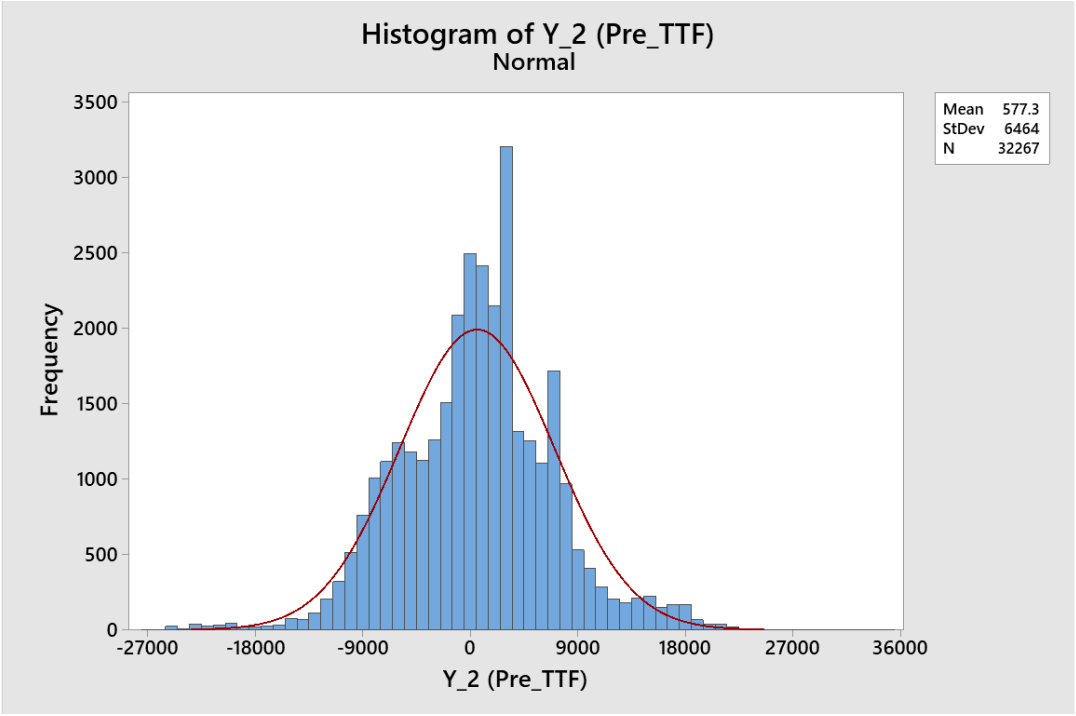


Figure E4: Histgorgram -Fairly normality distributed

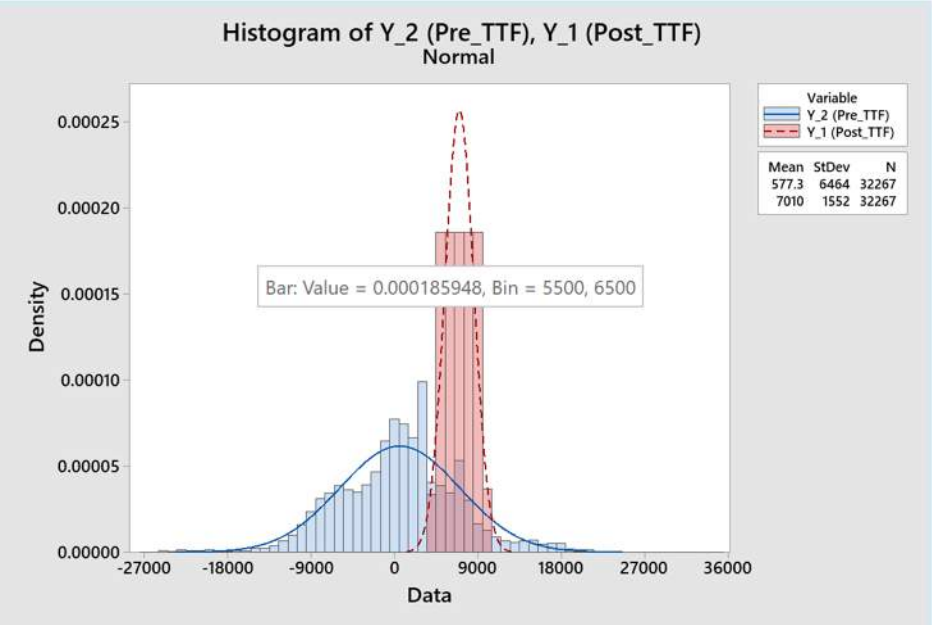
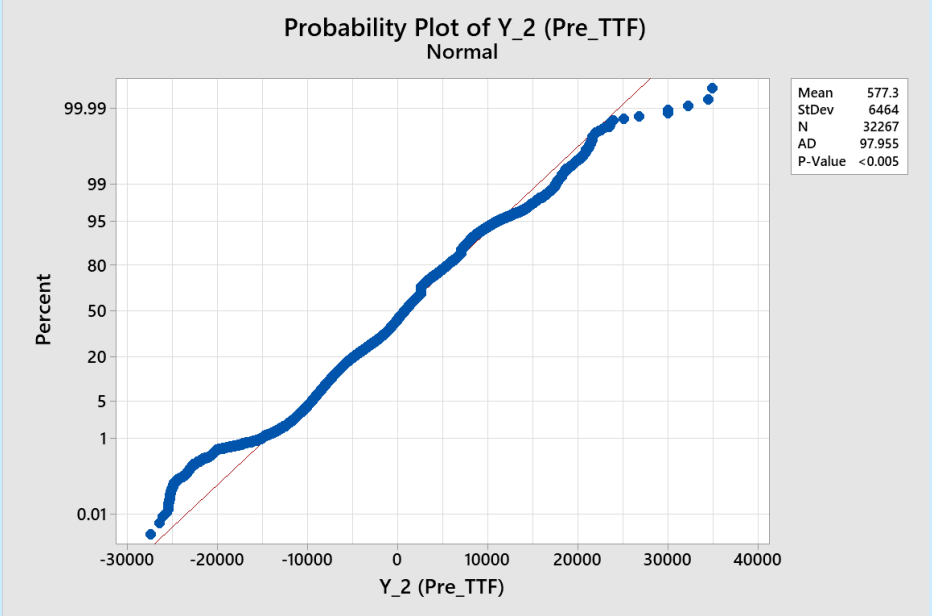



Figure E5: Distribution of the two dataset.