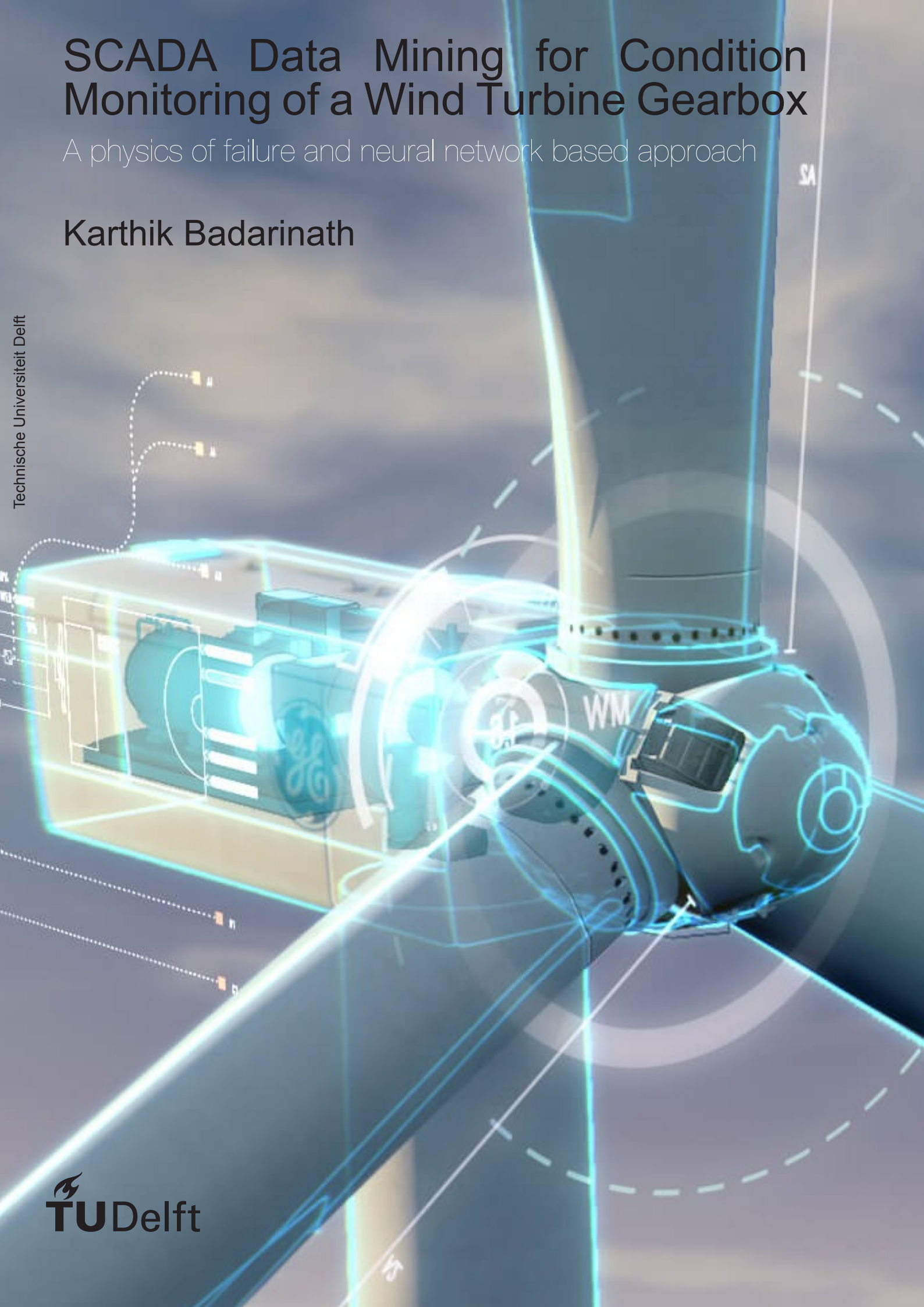


SCADA Data Mining for Condition Monitoring of a Wind Turbine Gearbox

A physics of failure and neural network based approach

Karthik Badarinath

Technische Universiteit Delft



SCADA Data Mining for Condition Monitoring of a Wind Turbine Gearbox

A physics of failure and neural network based approach

by

Karthik Badarinath

to obtain the degree of Master of Science
in Sustainable Energy Technology
at the Delft University of Technology,
to be defended publicly on Thursday September 12, 2019 at 2:00 PM.

Student number:	4727673	
Project duration:	December 4, 2018 – August 30, 2019	
Thesis committee:	Prof. dr. S.J. Watson	TU Delft, Internal supervisor
	Dr. R. A. J. Ostayen	TU Delft, Committee
	Dr. ir. X. Jiang	TU Delft, Committee
	MSc. B. Van Den Kieboom	Eneco, External supervisor

Cover Image courtesy - GE Research Digital Twin

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



Preface

Wind power is the fastest-growing energy source and the offshore wind industry is expected to grow by over 80 GW through 2024. There is a need to reduce all associated costs to be competitive in a market that might be fully subsidy-free soon. Adapting a predictive or condition-based maintenance strategy can help in reducing O&M costs, as it accounts for more than 30% of the lifetime costs in offshore turbines.

This master thesis focuses on developing a novel gearbox condition monitoring technique using raw Supervisory Control And Data Acquisition (SCADA) data from an offshore wind farm. The developed approach attempts to leverage the use of physics of failure technique and artificial neural network models for continuous monitoring of gearbox condition. Utilizing SCADA data for monitoring and fault detection can help in eliminating retrofitting additional expensive sensors such as an online oil monitoring system, vibration monitoring systems and so on for the same purpose.

Firstly, a physics of failure methodology is implemented to monitor turbines on a farm level. Using Failure Mode, Effects Analysis (FMEA), the operating regimes under which gears and bearings can potentially get damaged are ascertained. A correlation between the damage operating conditions and the root cause of failure is analyzed using boxplots. In particular, a correlation between operating conditions and the unexplained *White Etching Cracks or axial cracking* failure that occur frequently in the high-speed shaft bearings is studied. It is observed that turbines that experienced atleast one of the below damage operating conditions compared to the rest of the fleet had a failure in high speed bearing due to axial cracking -

1. Higher number of braking events at high wind speeds and rated power.
2. Higher number of emergency brake events.
3. Idling for longer periods which results in bearings operating at boundary/mixed lubrication regimes.
4. Higher number of rotor start-stop cycles.

Additionally, this methodology is also used to complement the Artificial Neural Network model developed for anomaly detection, as it was observed that approximately 70% of failed gearboxes were operating at higher risk of damage, while the rest might be due to a random failure.

Finally, an Artificial Neural Network (ANN) based on a Recurrent Neural Network (RNN) is developed for fault detection. The model is trained on data when the turbine was operating normally, to predict gear bearing and gearbox lubrication oil temperatures. A *stacked Gated Recurrent Unit (GRU)*, a type of RNN model with two hidden layers resulted in predictions with the lowest mean squared error of 0.1464°C for gear bearing temperature model and a mean squared error of 0.2375°C for gearbox oil temperature along with good generalization. The difference/residual between the measured and ANN predicted temperatures are used to flag anomalies. The ANN-based condition monitoring method is validated through case studies using real data from wind turbines. The results from the case studies indicate that the RNN based fault detection method can detect a failure in the wind turbine gearbox components as early as 3-6 months before it fails.

Keywords: Wind turbine gearbox, Physics of failure, White etching cracks, Condition monitoring, Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), Supervisory Control And Data Acquisition (SCADA)

Acknowledgement

It always seems impossible until it's done.

- Nelson Mandela

The last two years of my Master's life has been like a roller-coaster ride. Completion of this thesis marks the culmination of my Master's program. This journey would not have been possible without the help and assistance of several people.

First and foremost, I would like to thank my supervisor *Prof. Simon Watson* for his constant guidance and assistance throughout this thesis. A very big thanks to all my supervisors at Eneco - *Wouter Haans, Bas van den Kieboom and Bart van der Wielen* for all the belief they had in me since the start of this research work and for proofreading my thesis. They gave me an opportunity to work with Eneco though my knowledge on machine learning was really poor when I started out. I would also like to thank my Eneco colleagues - *Rody, Ufuk and Marina*.

I am grateful for all the data and support provided by Eneco, without which this work would not have been possible. Also special thanks to *Prof. Andrew Ng* for his MOOCs on machine and deep learning, and also *Jason Brownlee* and his blogs on machinelearningmastery.com which helped me massively in understanding and implementing my ML models. I would also like to express my gratitude to *Prof. Ron van Oostayen* for his valuable opinion on my work and steering me in the right direction.

I would like to thank all my friends, *Arya, Anurag, Anoosh, Arvind, Akhilesh, Nikhilesh, Mihir, Sidharth, Samruddhi, Varun, Sharab, D-Mama aka Sushanth, Kishan* who made my stay in Delft memorable and enjoyable. Surviving without you guys in Delft would have been impossible. Special thanks to my teammates from Energy for Refugees *Anurag, Maria, Manolis, Gamze, Shivam and Roos*. I would also like to thank *Kushal aka Rafa* for coming out to play tennis every weekend which helped me to bust all the stress from academics. Thanks, *Bose* for encouraging me to take up this thesis topic.

Finally but most importantly, I will forever be indebted to my parents for always supporting my dreams and for their constant encouragement. I also want to thank my brother for always being there and supporting me.

Karthik Badarinath
Delft, September 2019

Contents

List of Figures	ix
List of Tables	xiii
Abbreviations	xv
Nomenclature	xviii
1 Introduction	1
1.1 Wind Energy Outlook and Need for Condition Monitoring	1
1.2 Overview of Wind Turbine SCADA	1
1.3 Failure Statistics	2
1.4 Maintenance Strategies and Signal Monitoring	2
1.5 Research Scope	4
1.6 Layout.	5
2 Background	7
2.1 Prognostics Approaches	7
2.1.1 Physics based Prognostics	7
2.1.2 Data-driven Prognostics	9
2.2 Machine Learning Basics	9
2.3 Artificial Neural Networks	10
2.3.1 Loss Function.	12
2.3.2 Training Neural Networks	12
2.4 Nonlinear Autoregressive Exogenous Input.	13
2.5 Recurrent Neural Network	13
2.5.1 Need for RNNs for Sequential Data	14
2.5.2 RNN Theory	14
2.5.3 Training RNN	14
2.6 Long Short Term Memory	15
2.7 Gated Recurrent Unit.	18
2.8 Relevant works on Fault Detection using ANN	19
2.9 Research Problem and Objectives	20
2.10 Proposed Condition Monitoring Methodology	20
3 Failure Risk Assessment of Gearbox using Physics of Failure Technique	23
3.1 Introduction	23
3.2 Gearbox Specification	24
3.3 Damage Operating Conditions for Bearing Failures	25
3.4 Case Study - Part I	27
3.4.1 Turbine 26 - Planetary stage bearing spalling.	27
3.4.2 Turbine 32 - PS planet pinion pitting.	27
3.4.3 Turbine 2 - PS gear ring fracture/cracking.	28
3.4.4 Turbine 53 - IMS bearing spalling	28
3.5 HSS Bearing Configuration and Load Distribution	30
3.6 Bearing Rating Life	34
3.6.1 Equivalent Dynamic Bearing Load (P).	35
3.6.2 Life Modification Factor a_{ISO}	35
3.6.3 L10 Life based on Field Operating Conditions	36
3.6.4 Minimum Requisite Load.	37

3.7	Contact Mechanics	38
3.8	Contact Lubrication	41
3.8.1	Elastohydrodynamic Lubrication (EHL) Theory	41
3.8.2	Contact Friction	44
3.9	White Etching Cracks (WECs)	46
3.9.1	Potential Root Causes of WEC	46
3.10	Case Study II - HSS Bearing Failure	49
3.10.1	Turbine 2 - HSS CRB axial cracking	50
3.10.2	Turbine 8 - No Failure	50
3.10.3	Turbine 14 - No Failure	51
3.11	Limitations	52
4	Temperature based Anomaly Detection using Artificial Neural Networks	53
4.1	Normal Behavior Modeling	53
4.2	Input and Output Parameter Selection	54
4.3	Data Preprocessing	56
4.3.1	General Filter	57
4.3.2	Cluster Filter	57
4.3.3	Missing Data Filter	57
4.3.4	Masking for RNN	58
4.4	Artificial Neural Network Model Architecture	59
4.4.1	NARX model	59
4.4.2	Stacked RNN model	62
4.4.3	Data scaling and hyperparameter tuning for RNN	62
5	Case Studies	69
5.1	Data Post-processing	69
5.2	Case studies of HSS bearing failures	70
5.2.1	Turbine - 48	71
5.2.2	Turbine - 23	75
5.2.3	Turbine - 7	78
5.3	Discussion	81
5.4	Damage Risk Assessment	81
5.5	Case Study for non HSS bearing failures	81
5.5.1	Turbine - 32	81
5.5.2	Turbine - 17	83
6	Conclusions and Future Work	85
6.1	Using PoF on a wind farm level to improve condition monitoring and risk assessment	85
6.2	RNN model for temperature based anomaly detection	86
6.3	Future Work	86
6.3.1	Physics of failure model	87
6.3.2	ANN based condition monitoring	87
6.3.3	Engineering recommendations	88
	Bibliography	89
A	Bearing Design and Lubricant Properties	95
B	Supporting Plots	97
B.1	Input/Output parameters for RNN models	97
B.1.1	Turbine 48	97
B.1.2	Turbine 23	99
B.1.3	Turbine 32	100
B.1.4	Turbine 7	102
B.2	ANN Model Predictions	104

List of Figures

1.1	Normalised Failure Rates and Downtimes for Geared $G \geq 1$ MW Turbines . Adapted from [3].	2
1.2	Overview on failure rate and mean down time per WT as published by different initiatives. Adapted from [4].	3
1.3	Evolution of maintenance strategies.	3
1.4	Costs associated with traditional maintenance strategies [5].	4
1.5	Typical development of a mechanical failure. Adopted from [5].	5
2.1	Categorization and definition of prognostics methods. Adapted from [6]	7
2.2	Illustration of physics-based prognostics. Adapted from [6]	8
2.3	Illustration of an artificial neuron [21].	10
2.4	List of activation functions [22].	11
2.5	Representation of an Artificial Neural Network.	11
2.6	Architectures of the NARX neural network. Adapted from [25].	13
2.7	A Standard RNN. Adapted from [26].	14
2.8	A schematic diagram of the LSTM unit with forget gates. Adapted from [31].	15
2.9	A schematic representation of LSTM network [32].	16
2.10	Forget gate operation.	16
2.11	Input/Update gate operation.. . . .	17
2.12	Cell state update.	17
2.13	Output gate operation.	18
2.14	Pictorial representation of GRU.	18
2.15	Overview of proposed condition monitoring method.	21
2.16	A - Proposed condition monitoring technique using only SCADA data.	21
2.17	B - Proposed methodology to find the correlation between White Etching cracks and damage operating conditions in the High Speed Stage of gearbox.	22
2.18	C - ANN based condition monitoring technique (Normal Behaviour modeling).	22
3.1	Wind turbine gearbox arrangement. Adapted from [46].	24
3.2	Wind turbine gearbox vulnerability map. Adapted from [47].	24
3.3	Turbine 26 - PS bearing spalling.	27
3.4	Turbine 32 - PS planet pinion pitting.	28
3.5	Turbine 2 - PS gear ring fracture/cracking.	29
3.6	Turbine 53 - IMS bearing spalling.	29
3.7	HSS free-body diagram.	31
3.8	Plot of Power versus HSS Torque for one year of SCADA data.	32
3.9	TRB face-face matched pair radial load sharing [55].	32
3.10	TRB face-face matched pair axial load sharing [55].	33
3.11	Histogram of HSS torque taking one year SCADA data.	34
3.12	Histogram of generator speed taking one year SCADA data.	34
3.13	Histogram of radial load on TRB pair taking one year SCADA data.. . . .	34
3.14	Histogram of axial load on TRB pair taking one year SCADA data.	34
3.15	High speed shaft and bearing roller loads at 100% power (750 KW) calculated using the Transmission3D model. The contact areas on the rollers are highlighted. Adapted from [53].	34
3.16	Histogram of radial load on CRB taking one year SCADA data.	36
3.17	Histogram of dynamic equivalent load on TRB pair taking one year SCADA data.	36
3.18	Life modification factor, a_{SKF} . Adapted from [58].	37
3.19	A pictorial representation of data removed from the training data set due to filtering.	39

3.20	Normalized plot of the depth distribution of the σ_x , σ_y , and σ_z main normal and of the v. Mises equivalent stress below the center line of the Hertzian contact area [62].	40
3.21	Plot of maximum contact pressure versus radial load on CRB.	41
3.22	Plot of maximum contact pressure versus maximum von mises stress on CRB.	41
3.23	Pressure distribution and film-thickness variation in an EHD contact.	41
3.24	Lubricant film thicknesses for point contact and line contact. Adapted from [63].	42
3.25	Typical Stribeck curve of a lubricant under constant load, temperature and slide-to-roll ratio and its correlation to the film thickness curve under the same operating conditions. Adapted from [67].	43
3.26	Scatter plot of λ vs Generator RPM of CRB for one year SCADA data.	44
3.27	Histogram of λ in CRB for one year of SCADA data.	45
3.28	Bearing frictional moment as a function of speed. Adapted from [68].	45
3.29	Bearing frictional moment as a function of speed using SCADA data.	46
3.30	Failure appearance: a) straight cracks; b) straight cracks and small spalls; and c) spalls. [69]	47
3.31	(a) Raceway top view of a typical WEC-associated spall; (b) circumferential LOM with discrete WEC networks displaying vertical links to the surface and a step-wise propagation; (c) axial LOM with WECs laying parallel to the surface in accordance with the respective steps in (b); (d) circumferential LOM closer to the DEA region; (e) SEM analysis revealing ultra-thin cracks, adjacent refined microstructure and darker alterations. LOM cross sections are all Nital 2% etched [70].	47
3.32	(A) Roller-raceway contact with areas of local high traction due, for example, to local mixed friction, leading to tensile stresses that can, b) lead to damage such as a small crack; c) surface crack or crack connected to the raceways allows the entry of oil. Adapted from [69].	48
3.33	Schematic representation of rapidly evolving torque reversals.	48
3.34	Dynamometer test results on HSS (A) during braking events (B) during grid loss events [53]	49
3.35	Importance of surface traction (not to scale): (a) pure rolling: critical stress threshold exceeded relatively deep—micro-cracking at inclusion may never reach surface and cause failure; (b) rolling and sliding: critical stress field shifted closer to surface—micro-cracking at inclusion propagates to surface, leading to failure [74]	49
3.36	White etching crack occurrence (simplified) in rolling element bearings. [69]	50
3.37	Turbine 2 - HSS bearing axial cracking	51
3.38	Turbine 8 - No failure	51
3.39	Turbine 14 - No failure	52
4.1	A schematic representation of an ANN based condition monitoring method. Adapted from [39]	53
4.2	Correlation Matrix of SCADA variables.	55
4.3	A pictorial representation of data removed from the training data set due to filtering - Power curve	58
4.4	Power curve	58
4.5	A pictorial representation of masking operation in LSTM models [77]	59
4.6	Implemented NARX Neural Network.	60
4.7	Bearing temperature - modeled versus measured NARX.	60
4.8	Mean squared error for NBM.	61
4.9	Training error histogram for bearing temperature.	61
4.10	Representation of stacked GRU model.	62
4.11	Representation of ReLU activation function.	64
4.12	MSE vs learning rate for final GRU network.	65
4.13	Gradient descent batch size and convergence.	65
4.14	Network configuration and number of weight parameters to be tuned - 726.	67
4.15	Measured and GRU predicted gear bearing temperature.	68
4.16	GRU network performance with learning rate 0.001 for gear bearing temperature model.	68

5.1	Gear bearing residual temperature for a healthy turbine..	71
5.2	Gearbox oil residual temperature for a healthy turbine.	72
5.3	Gear bearing residual temperature for a faulty turbine..	73
5.4	Gearbox oil residual temperature for a faulty turbine.	73
5.5	Turbine 48 - HSS bearing axial cracking	74
5.6	Gear bearing residual temperature for a healthy turbine - Year 2011..	75
5.7	Gearbox oil residual temperature for a healthy turbine - Year 2011..	75
5.8	Gear bearing residual temperature for a faulty turbine..	76
5.9	Gearbox oil residual temperature for a faulty turbine.	77
5.10	Turbine 23 - HSS bearing axial cracking	77
5.11	Gear bearing residual temperature for a healthy turbine..	78
5.12	Gearbox oil residual temperature for a healthy turbine.	78
5.13	Gear bearing residual temperature for a faulty turbine..	79
5.14	Gearbox oil residual temperature for a faulty turbine.	79
5.15	Turbine 7 - HSS bearing axial cracking	80
5.16	Gear bearing residual temperature for a faulty turbine..	82
5.17	Gearbox oil residual temperature for a faulty turbine.	82
5.18	Gear bearing residual temperature for a faulty turbine..	83
5.19	Gearbox oil residual temperature for a faulty turbine.	84
5.20	Turbine 17 - No Failure.	84
6.1	Framework for hybrid prognosis.	87
B.1	Input Parameters for Turbine 48.	97
B.2	Output Parameters for Turbine 48.	98
B.3	Vibration analysis of planetary stage gearbox before replacement of Turbine 48.	98
B.4	Input Parameters for Turbine 23.	99
B.5	Output Parameters for Turbine 23.	99
B.6	Vibration analysis of high speed bearing before replacement - Turbine 23..	100
B.7	Input Parameters for Turbine 32.	100
B.8	Output Parameters for Turbine 32.	101
B.9	Vibration analysis of high speed bearing before replacement - Turbine 32..	101
B.10	Input Parameters for Turbine 7.	102
B.11	Output Parameters for Turbine 7.	102
B.12	Vibration analysis of high speed bearing before replacement of Turbine 7..	103
B.13	Normal behaviour predictions of gear bearing temperature.	104
B.14	Anomalous behaviour of gear bearing temperature.	104
B.15	Normal behaviour predictions of gearbox oil temperature.	105
B.16	Anomalous behaviour of gearbox oil temperature.	105

List of Tables

1.1	Basic SCADA parameters.	2
3.1	Identification of failure root causes and critical operating conditions - Bearings.	26
3.2	Actual vs Calculated L10 life of failed bearings.	38
4.1	Input/Output parameters for ANN [Set-1]	56
4.2	Input/Output parameters for ANN [Set-2]	56
4.3	Boundaries for parameters of the gearbox model.	57
4.4	Specification of NARX model.	60
4.5	Final tuned hyperparameters for GRU network.	66
4.6	Specification of GRU network.	66
4.7	MSE metric results for some of the configurations.	67
5.1	Vibration monitoring system alarms.. . . .	70
5.2	Summary of all case studies with alarm trigger dates.	70
5.3	Signals triggered after ANN based temperature model detects anomaly.	81
A.1	Bearing design and mechanical properties.. . . .	95
A.2	Lubricant properties.	95

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
BPTT	Backpropagation Through Time
CBM	Condition-Based Monitoring
CMS	Condition Monitoring Systems
CRB	Cylindrical Roller Bearing
EHD	Elasto-Hydro Dynamic
EHL	ElastoHydrodynamic Lubrication
FFNN	Feed Forward Neural Network
GRC	Gearbox Reliability Collaborative
GRU	Gated Recurrent Unit
HSS	High Speed Stage
LSS	Low Speed Stage
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NARX	Nonlinear AutoRegressive eXogenous
NBM	Normal Behaviour Modeling
NREL	National Renewable Energy Laboratory
O&M	Operations and Maintenance
PCD	Pitch Circle Diameter
PoF	Physics of Failure
RMS	Root Mean Square
RNN	Recurrent Neural Network
SCADA	Supervisory Control And Data Acquisition
TRB	Tapered Roller Bearing
WEC	White Etching Cracks

Nomenclature

Greek Symbols

α	Pressure-viscosity coefficient	$[\text{Pa}^{-1}]$
α_p	Pressure angle of HSS pinion	$[10^\circ]$
β	Helix angle of HSS pinion	$[14^\circ]$
η_c	Lubrication contamination factor	$[-]$
κ	Viscosity ratio	$[-]$
λ	Lubrication lambda	$[-]$
μ_p	Dynamic viscosity	$[\text{cP}]$
ν	Kinematic viscosity	$[\text{cSt}]$
ν_1	Rated viscosity	$[\text{cSt}]$
ν_p	Poisson's ratio	$[-]$
ω	Generator RPM	$[\text{revolutions per minute}]$
ρ	Density	$[\text{kg}/\text{m}^3]$
σ_v	Von mises stress	$[\text{MPa}]$
$\sigma_x, \sigma_y, \sigma_z$	Principal stress in x, y and z axis respectively	$[\text{MPa}]$

Latin Symbols

a_1	Life adjustment factor for reliability	$[-]$
a_{ISO}	Life modification factor	$[-]$
B	Length of the contacting surface between bearings	$[\text{mm}]$
b	Half width of elliptical contact area	$[\text{mm}]$
C	Basic dynamic load rating	$[\text{kN}]$
d_m	Mean bearing diameter	$[\text{mm}]$
D_{pw}	Bearing pitch diameter	$[\text{mm}]$
E	Elastic modulus of contacting surface	$[\text{MPa}]$
F	Dynamic equivalent load acting between contacting surfaces	$[\text{kN}]$
F^{CB}	Radial load on cylindrical roller bearing	$[\text{kN}]$
F_m	Tangential load on HSS shaft	$[\text{kN}]$
F_s	Radial load on HSS shaft	$[\text{kN}]$
F_a	Axial load on tapered roller bearing pair	$[\text{kN}]$
$F_a A$	Axial load on upwind tapered roller bearing	$[\text{kN}]$

F_aB	Axial load on downwind tapered roller bearing	[kN]
F_{rm}^{CRB}	Minimum requisite load for pure rolling of CRB	[kN]
F_r^{TRB}	Radial load on tapered roller bearing pair	[kN]
F_rA^{TRB}	Radial load on upwind tapered roller bearing	[kN]
F_rB^{TRB}	Radial load on downwind tapered roller bearing	[kN]
h_{min}	Minimum lubricant thickness	[μm]
K_a	Axial load on HSS shaft	[kN]
K_r	Resultant load on HSS shaft	[kN]
L	Length of high speed shaft	[m]
L_B	Distance between pinion and upwind CRB	[m]
L_{10}	Bearing rating life	[millions of revolutions]
L_{UW}	Distance between TRB pair and upwind CRB	[m]
n	Bearing speed	[revolutions per minute]
P	Power	[kW]
p	Exponent of life equation	[-]
P_u	Fatigue load limit	[kN]
P_{dy}	Equivalent dynamic bearing load	[kN]
P_{max}	Maximum contact pressure along the center line	[MPa]
R_1, R_2	Radii of the contacting surfaces	[mm]
R_q	RMS surface roughness of bearing	[μm]
T	Generator torque	[kN · m]
U	Cycle fraction	[μm]
X	Radial load factor of bearing	[-]
Y	Axial load factor of bearing	[-]
Y_A, Y_B	Axial load factor of tapered roller bearing	[-]
z	Depth below the contacting surfaces	[mm]

Introduction

In this chapter, the background of the thesis is presented along with a small introduction on the growth of wind energy. A short note on the failure statistics of wind turbines, maintenance techniques along with the research scope of the thesis are discussed.

1.1. Wind Energy Outlook and Need for Condition Monitoring

Wind energy is one of the fastest-growing sources of energy. The overall capacity of installed wind energy worldwide by the end of 2018 is 597 GW [1]. It is expected to grow by over 80 GW by the end of 2023. With the development of technology, bigger wind turbines are being built and installed far offshore to tap more wind. With such exponential development, a failure in one of the wind turbine component can result in higher revenue losses and it is also impractical to have frequently scheduled maintenance as it can become tedious and expensive. Especially for offshore turbines, operations and maintenance (O&M) accounts for 25-30% of the lifetime costs [2]. To overcome the high costs of O&M it is necessary to move to more intelligent and proactive monitoring of wind turbines, often termed as condition monitoring.

A number of different expensive solutions are already available for condition monitoring of wind turbines such as - vibration monitoring systems, real-time oil quality monitoring systems and so on. These systems are normally installed as additional systems with extra costs and do not come retrofitted with a turbine. Hence, operators are in search of alternative cheap solutions to monitor the condition of the wind turbine. All major wind turbine manufacturers install a Supervisory Control And Data Acquisition (SCADA) system to monitor the performance of wind turbines which records various parameters. As SCADA data is a potentially low-cost solution for condition monitoring, this thesis discusses using these data for monitoring and improving maintenance strategies of wind turbines.

1.2. Overview of Wind Turbine SCADA

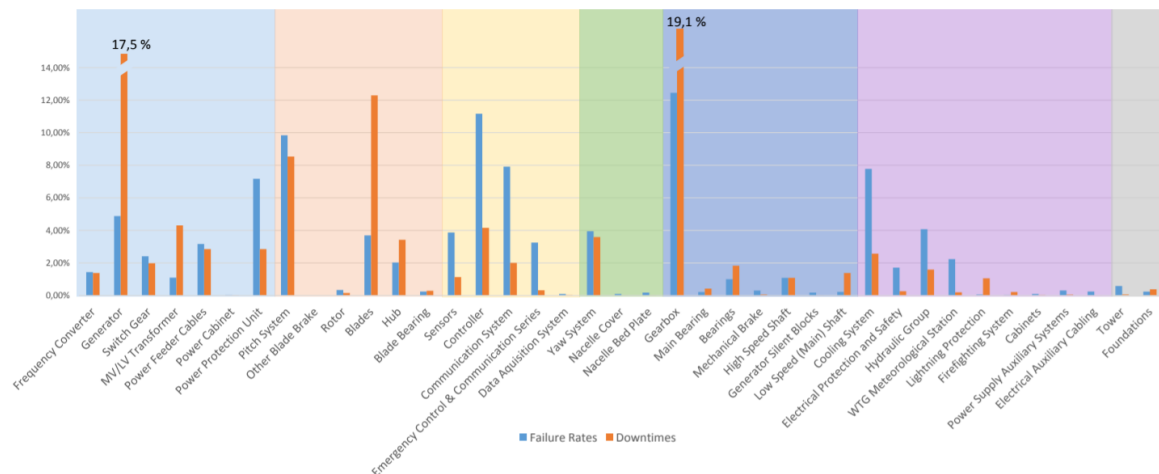
Supervisory **C**ontrol **A**nd **D**ata **A**cquisition (SCADA) is a system of software and hardware elements that allow wind farm owners to monitor wind turbine performance along with gathering and storing various data parameters. Typically, all large scale utility wind turbines come retrofitted with a SCADA system. SCADA records data usually at 10-minute resolution, though the range and type of signals recorded can vary from one type of turbine to another. A list of typically recorded SCADA parameters is tabulated in Table 1.1.

Table 1.1: Basic SCADA parameters.

Environmental	Electrical characteristics	Part Temperatures	Control Variables
Wind speed	Grid power output	Gear bearing	Rotor speed
Wind direction	Reactive power	Gearbox oil	Yaw angle
Ambient temperature	Generator voltages	Generator bearing	Pitch angle
Nacelle temperature	Generator phase current	Hub controller	Generator speed
	Voltage frequency	Top controller	Cooling pump status
	Power factor	Main bearing	Operational status code
		Rotor inverter	Number of starts / stops
		Hydraulic oil	

1.3. Failure Statistics

Understanding of common failures and typical downtime of a wind turbine due to the failure of a particular component is essential for optimizing O&M strategies. Several surveys of wind turbine failures have been conducted in the last two decades to identify failure rates and associated downtime for different sub-assemblies.

Figure 1.1: Normalised Failure Rates and Downtimes for Geared $G \geq 1\text{MW}$ Turbines . Adapted from [3].

Reder et al., 2016 [3] conducted a comprehensive survey and study on onshore wind turbine failure statistics. An evaluation of more than 4000 wind turbine from 14 different manufacturers was conducted. Most of the observed downtime and failures occurred in generators and gearboxes, as shown in Figure 1.1. Similar trends can be observed in Figure 1.2 which compares various studies done by different stakeholders. Therefore, gearbox and drivetrain problems are certainly a major challenge to be addressed to further reduce downtime which will be the focus of this thesis.

1.4. Maintenance Strategies and Signal Monitoring

The O&M costs of an offshore wind turbine account for up to 30% of the life cycle cost and hence the cost of energy. A reduction of maintenance cost and therefore a reduction of the cost of wind energy can be achieved by further improvements in wind turbine design leading to improvement of its inherent reliability, but also by systematic solutions for maintenance management.

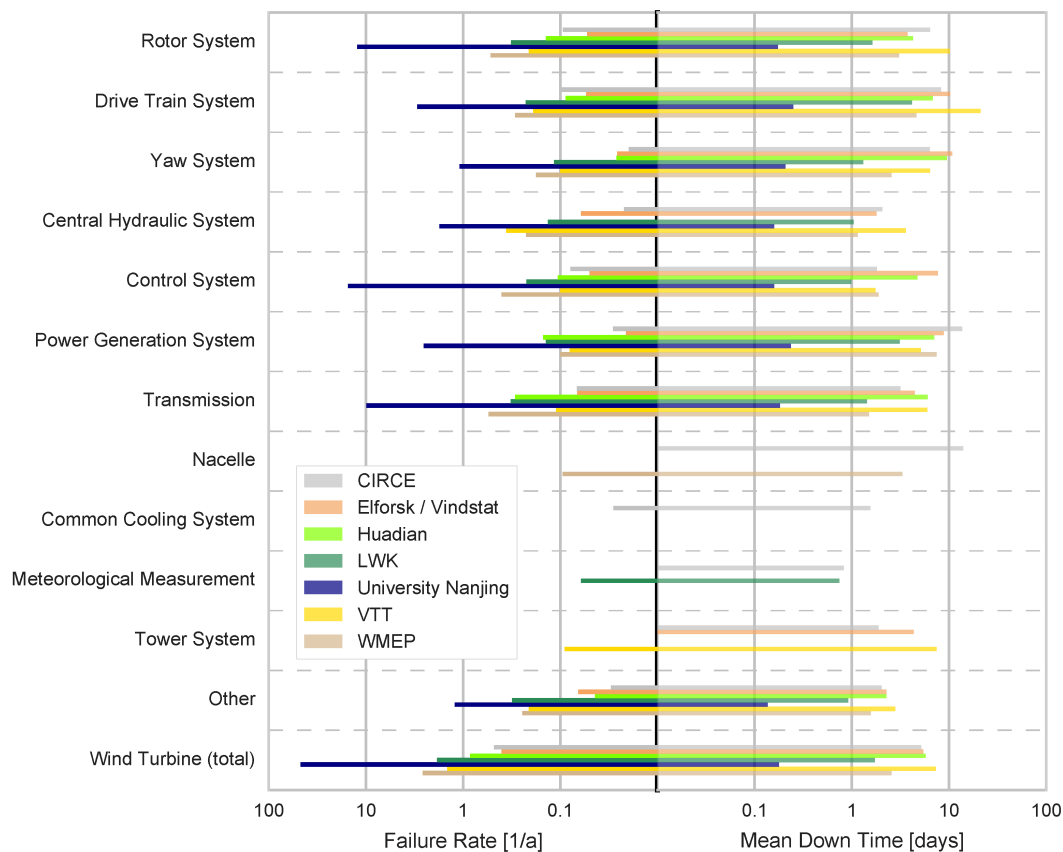


Figure 1.2: Overview on failure rate and mean down time per WT as published by different initiatives. Adapted from [4].

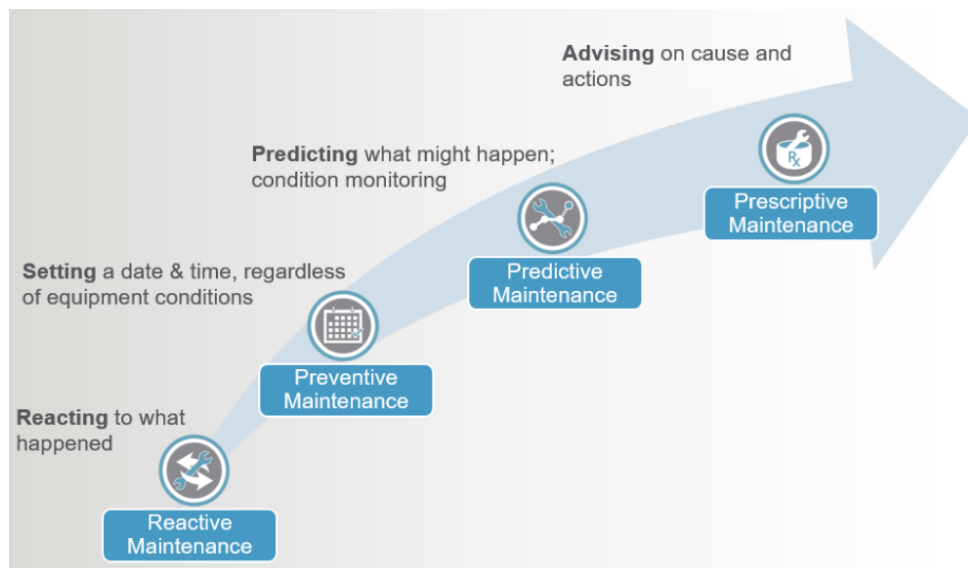


Figure 1.3: Evolution of maintenance strategies.

The basic types of maintenance strategies discussed in the asset management industry are - (a) Reactive maintenance, (b) Preventive maintenance, (c) Condition-based maintenance / Predictive maintenance and (d) Prescriptive maintenance, as shown in Figure 1.3. Currently, the industry is evolving from preventive maintenance to predictive maintenance. This thesis is focused on developing predictive maintenance algorithms.

Reactive maintenance is also known as run-to-failure strategy. In this strategy, the component is allowed to run until the end of its life and replacement is done once it completely fails. This type of maintenance leads to higher downtime and higher costs of repair.

Preventive maintenance takes place before a component fails. It is a time-based approach that is carried out at predetermined intervals to reduce failure risk or performance degradation of assets. The main aim of this strategy is to minimize unplanned downtime and reduce repair costs.

Predictive maintenance or condition-based maintenance strategy is based on evaluating the health or condition of the components before scheduling maintenance. The health of the equipment is monitored by various sensors. This way an optimum between downtime costs and repair costs can be found as shown in Figure 1.4. Examples of predictive maintenance include oil sample analysis, acoustic emission, thermal analysis, and vibration monitoring.

Prescriptive maintenance is a step-up of predictive maintenance. Artificial Intelligence-enabled Prescriptive Maintenance is unique in that instead of just predicting impending failure it strives to produce outcome-focused recommendations for operations and maintenance from analytics.

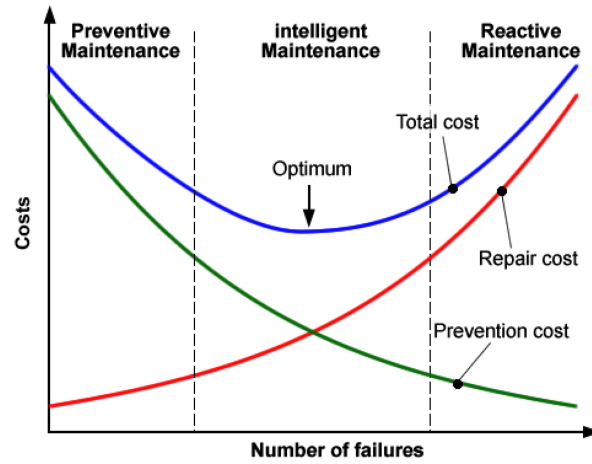


Figure 1.4: Costs associated with traditional maintenance strategies [5]

An impending failure in any typical mechanical device can be tracked using various signals as shown in Figure 1.5, depending on available sensors. A vibration analysis gives the earliest signals, however expensive vibration detection systems need to be installed. Similarly special expensive devices have to be installed for online oil quality monitoring and noise monitoring. A typical wind turbine SCADA records various temperature signals and comes with the wind turbine without any additional costs. The focus of this thesis, is therefore to develop temperature based condition monitoring strategies using SCADA data where many temperature signals are recorded.

1.5. Research Scope

The main focus of this thesis is to develop a novel technique leveraging the advantages of both the Physics of Failure and Artificial Neural Network model-based approach for effective condition monitoring of wind turbine gearbox. This research work also attempts to correlate operational history of wind turbines with *White Etching cracks* as observed in high speed shaft bearings of a wind turbine gearbox. The data for this research work was obtained from

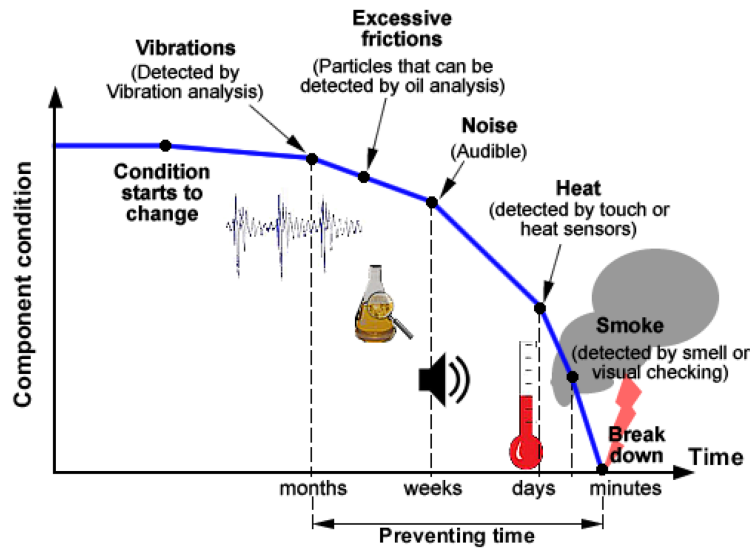


Figure 1.5: Typical development of a mechanical failure. Adopted from [5].

Prinses Amalia offshore wind farm owned by Eneco. The wind farm consists of 60 Vestas V-80 2MW turbines.

1.6. Layout

This master thesis will follow the following outline -

Chapter 2 details about the basics of physics of failure based approach for risk-based maintenance. Also a detailed explanation of artificial neural networks, in particular, the idea behind of Recurrent Neural Networks (RNN) and NARX is provided. The research objectives and methodology are presented at the end of the chapter.

Chapter 3 discusses in detail the physics of failure based approach for condition-based monitoring. Also, a correlation between the common failure mode that occurs in high-speed shaft bearings of a gearbox, namely *White Etching Cracks* and the historical damage operating conditions of the failed turbine are presented. A basic introduction of tribo-mechanical aspects of bearings is also provided to help better understand the operations of bearings.

Chapter 4 covers the details about the proposed artificial neural network model for temperature based anomaly detection of wind turbine gearbox. Details of hyperparameter tuning of the neural network are also presented.

Chapter 5 contains 5 case studies performed to validate the proposed condition monitoring model. Some limitations of the proposed model are also covered in the chapter.

Chapter 6 concludes this work and lists opportunities for future work that could not be addressed in the scope of this thesis.

2

Background

In this chapter, the necessary background information required for the understanding of the rest of the thesis report is discussed. Also, previous works done relating to condition-based monitoring using SCADA is also presented. Finally, the research problem, objectives and the methodology is detailed.

2.1. Prognostics Approaches

Prognostics is an engineering discipline focused on predicting the time-to-failure of a system or component. The core of prognostics lies in estimating the Remaining Useful Life (RUL). An accurate estimation of RUL can help in condition-based monitoring of wind turbines and can help to avoid long downtimes. The reliability of the system also increases with more accurate prognosis. Due to its significance, there has been a rapid rise in the research of Prognosis and Health Management (PHM) in recent years. Prognostics approaches can be classified into three main types - 1) physics-based prognostics, 2) data-driven prognostics and 3) hybrid prognostics. The physics based and data driven prognostics are further explained.

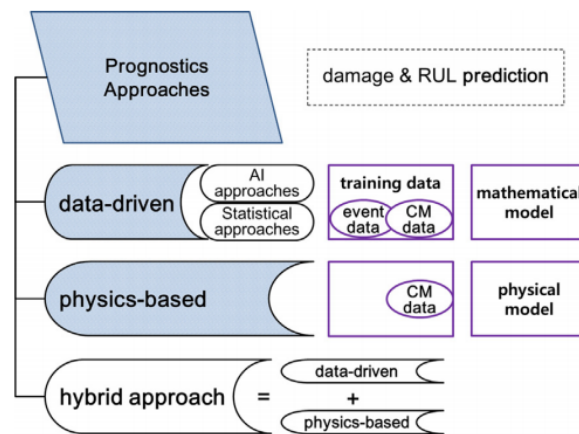


Figure 2.1: Categorization and definition of prognostics methods. Adapted from [6]

2.1.1. Physics based Prognostics

A physical understanding of the system is necessary for physics based prognostics models. To estimate the RUL in physics based prognostics (or white-box model), the field data measured is combined with a damage model based on first principles. Physical damage models

are mostly based on physics of failure (PoF) techniques. These types of models can be used in applications which require high accuracy. An illustration of physics based prognostics is shown in Figure 2.2.

A general outline of a physics of failure methodology can be described as follows [7], [8]:

1. System Analysis: The failure mode and effect analysis (FMEA) is carried out in this part. Also, the component design and material specifications are included in FMEA.
2. Load Spectra: The loads acting on the component of interest should be determined from the measured data. With the help of aeroelastic models loads in all six-degrees on freedom can be modelled accurately. Alternatively, torque calculations can also be used in damage models as it is a major damage driver.
3. Damage Model: Based on the failure mode occurring in the component, various damage models can be developed. Damage models are used to evaluate the damage accumulation of a component based on various loads acting on it. For a gearbox, typical damage models include crack propagation models, L10 life models and so on.
4. RUL Estimation: Based on the damage accumulation, various methods like Weibull reliability models can be used to estimate the remaining useful life of the component.

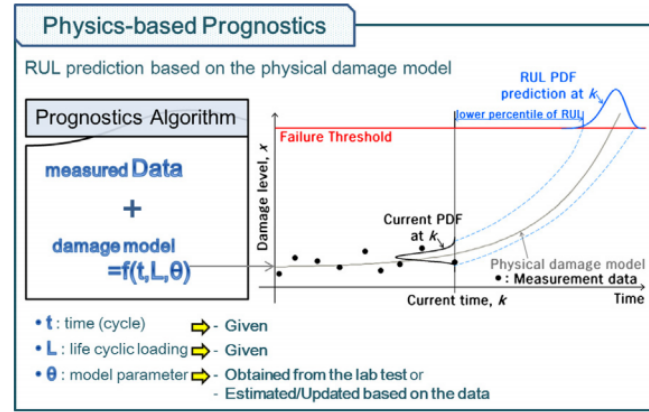


Figure 2.2: Illustration of physics-based prognostics. Adapted from [6]

Though physics based models give accurate results, there are several challenges in applying them in real-life situations. Firstly, failure modes vary from component to component, and also one component can fail due to different modes. Therefore, determining and modeling a specific failure mode can be difficult without interrupting operating machinery [9]. Most of the times, system specifications like material composition, or design of the component is not always available. Fault propagation also depends largely on future loading conditions. Hence, the model might not be accurate in dynamic operating conditions due to errors and certain assumptions made during the modeling phase.

Gray and Watson, 2010 [10] developed a physics of failure approach for the use of SCADA performance data in combination with applied knowledge of failure physics to calculate theoretical damage accumulation, and hence the risk, of failure for gearboxes. A similar approach was applied to predict electric yaw failures with some promising results [11].

Yingning Qiu et al., 2017 [12] used linear fatigue theory in combination with SCADA data to assess bending fatigue failures in planetary gears of the turbine. Stress cycles were calculated using detailed multibody system analysis, and Palmgren-Miner theory was applied to estimate the damage during the operation of the turbine.

Philippe Cambron et al., 2017 [13], proposed a physics based model to monitor the temperature of the main bearing of the turbine. The model used is based on the law of energy conservation. The difference between temperature measured by the SCADA and physics based temperature model was used to detect anomalies. A failure was detected three weeks before the current actual alarm settings used by SCADA were able to identify the issue.

Feng et al., 2013 [14], proposed a failure detection method for WT gearbox by predicting oil temperature rise using SCADA data and basic physics laws. Wilkinson et al., 2014,[15] compared physics based models based on temperature with signal trending techniques. It was concluded that physics based models give more accurate failure results and with lesser alarms.

Bretler et al., 2015 [16] proposed a physics of failure method for investigating the bending fatigue failure of a helical gear. An additional load generator module was proposed to consider external factors. The loads were calculated using finite element method, and meshing misalignment was measured using a laser. Even though in a system-level physics based model is proposed for failure prognosis, it still keeps the assumption that each component works independently. Al-Tubi et al., 2015 [17] also used micro-pitting failure models and SCADA data to predict failure in pinion gears on HSS. They concluded that torque variations have a great effect on micro-pitting in gears.

2.1.2. Data-driven Prognostics

Data-driven (DD) prognostics does not rely on the physics of the system. It uses the information of the measured data to learn the system behavior. Therefore these models are also known as black-box models. With respect to wind turbines, this measured data is obtained from supervisory control and data acquisition (SCADA) or condition monitoring systems (CMS) like accelerometers, oil monitoring systems, etc. Data-driven approaches are well suited for predicting the near future behavior of the system, especially towards the end of life as they are based on trend analysis of the data[6]. Data-driven approaches can mainly be classified into two categories- 1) Machine learning approach and 2) Statistical approach [18]. As the focus of the thesis is on using machine learning approach along with PoF models, the basics of machine learning and neural networks is discussed later in the chapter.

There are some limitations of using a data-driven approach for prognosis. The accuracy of the DD approach depends on both the quantity and the quality of the historical data. Sometimes getting access to good quality data is challenging. Also, DD approaches can lead to physically inconsistent results as there is no physics involved considered while predicting future states [19].

2.2. Machine Learning Basics

Machine Learning (ML) is a sub-field of Artificial Intelligence (AI) based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. The crux of ML lies in probabilistic and statistical theory. The 3 fundamental types of ML paradigms are:

Supervised Learning: In supervised learning data is in the form of input-output pairs. Each input x is matched with a target y and the task of ML algorithm is to learn mapping from input to output, $\mathbf{f}: \mathbf{X} \rightarrow \mathbf{Y}$. Supervised learning requires labeled data sets for effective training. In probabilistic terms it intends to infer a conditional probability function, $p(x|y)$ [20]. In the last decade most of the developments in ML has been done in supervised learning. Supervised learning uses classification and regression techniques to develop predictive models.

Unsupervised Learning: In unsupervised learning, labelled data is absent. Given only a list of inputs x , the algorithm tries to find a hidden or intrinsic structure in data. The most common type of unsupervised learning is clustering. In probabilistic terms, an unsupervised learning estimated the prior probability density function $p(x)$ [20].

Reinforcement Learning: It is an area of machine learning where learning occurs through a system of actions and rewards. The algorithm learns a policy of how to act and maximize rewards in a given environment.

A supervised machine learning model based on ANN is used to achieve the objectives of the thesis. A small introduction to ANN is given in Section 2.3.

2.3. Artificial Neural Networks

An Artificial Neural Network (ANN) is a machine learning model inspired by the function and structure of the biological brain. ANN helps in determining non-linear relations between observations or input data. The building blocks of an ANN are artificial neurons, referred to as 'nodes'.

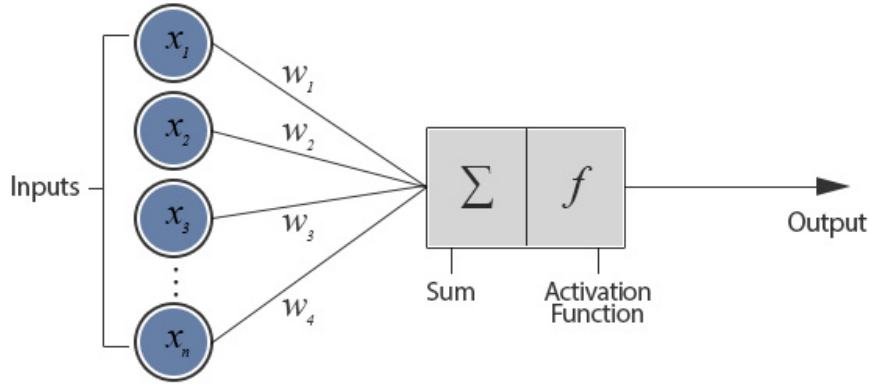


Figure 2.3: Illustration of an artificial neuron [21].

An illustration of functioning of a neuron is shown in Figure 2.3. Typically it takes in many inputs ($x_1, x_2, x_3, \dots, x_n$) which are individually weighed ($w_1, w_2, w_3, \dots, w_n$). These inputs are multiplied with their individual weights and a bias is added. A non-linear function known as *activation function* is then applied to the weighed sum and an output is obtained. Equation 2.1 is the mathematical representation of the functioning of an artificial neuron, where w , x , f , b , and y represent weight vector, input vector, activation function, bias and output vector respectively.

$$y(x) = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2.1)$$

There are various activation functions. A specific activation function can be selected based on the task that the neural network performs. An activation function decides whether to activate a neuron or not. A list of commonly used activation functions is shown in Figure 2.4.

An ANN is made up of several of these neurons as shown in Figure 2.5. It contains an input layer, a hidden layer and an output layer. There can be any number of nodes in a given layer. Also, there can be any number of hidden layers. Based on the task being performed, the number of nodes and the layers can be optimized.






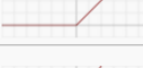


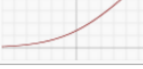
Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parametric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

Figure 2.4: List of activation functions [22].

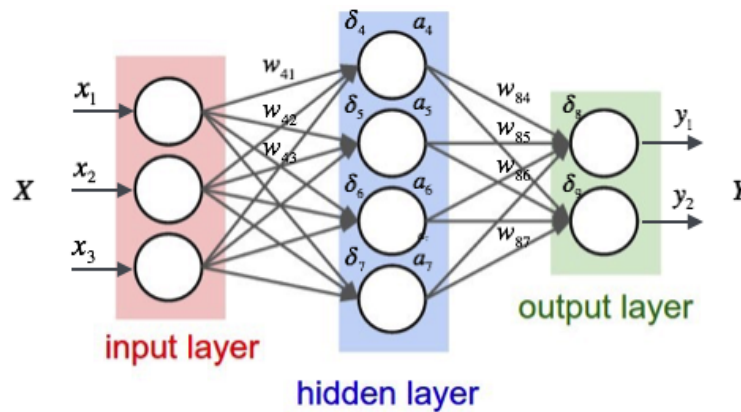


Figure 2.5: Representation of an Artificial Neural Network.

ANN can be classified into two categories- 1) Feed Forward Neural Networks (FFNN) and 2) Recurrent Neural Networks (RNN). FFNNs is the simplest form neural networks devised. Figure 2.5 represents a simple FFNN. In this type of neural network, the information always propagates in the forward direction; it never goes backward [23]. In this thesis, a Nonlinear AutoRegressive eXogenous (NARX) and two forms of RNN called Long Short Term Memory (LSTM) & Gated Recurrent Unit (GRU) are individually implemented to detect anomalies in the gearbox.

2.3.1. Loss Function

The loss function is an important part of an ANN. It is used to measure the inconsistency between the predicted value (\hat{y}) and the target value (y). The loss functions are nonnegative values. As the value of loss function decreases, the robustness of the model increases. Usually, optimization functions like gradient descent are used to minimize the empirical risk of the loss function. Loss function consists of an empirical risk function as well as a structural risk function, also known as the regularization term. This is shown in Equation 2.2.

$$\underset{f}{\operatorname{argmin}} \underbrace{\operatorname{Loss}(\hat{y}, y)}_{\text{Empirical Error}} + \underbrace{\lambda \cdot R(f)}_{\text{Structural Error}} \quad (2.2)$$

where, λ is a trade-off hyperparameter and $R(\cdot)$ measures the complexity of the model.

There are various commonly used loss functions for regression problems. These include Mean Absolute Error (MAE) (Equation 2.3), Mean Squared Error (MSE) (Equation 2.4), Mean Absolute Percentage Error (MAPE) (Equation 2.5) and so on. Here n is the number of data points in the variable, Y_i and \hat{Y}_i are the measured and the predicted values respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2.3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.4)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (2.5)$$

2.3.2. Training Neural Networks

Backpropagation is the most commonly used training algorithm for neural networks. It is used to calculate the gradient that is needed to update the weights to be used in the network.

Backpropagation:

The basic steps involved in backpropagation are-

1. Present a training input pattern and propagate it through the network to get an output.
2. Compare the predicted outputs to the expected outputs and calculate the error.
3. Calculate the derivatives of the error with respect to the network weights.
4. Adjust the weights to minimize the error.
5. Repeat.

The rule for updating weights on each iteration of gradient descent can be expressed mathematically as shown in Equation 2.6. Here α is the learning rate, w_{ij} is the weight matrix and E is the loss function that the algorithm is trying to minimize.

$$w_{ij} = w_{ij} - \alpha \cdot \frac{\partial E}{\partial w_{ij}} \quad (2.6)$$

If the dataset size is large, calculating the loss and gradient over the entire dataset may be too slow and computationally infeasible. Thus in practice, a variant of gradient descent called Stochastic Gradient Descent (SGD) is commonly used. In SGD, data is divided into subsets called batches, and the parameters are updated after calculating the loss function over one batch. Other popular variants are: RMSprop, AdaGrad, Adam. In some cases *decay* is used to decrease the learning rate gradually as parameters approach the optimum values.

2.4. Nonlinear Autoregressive Exogenous Input

Nonlinear AutoRegressive eXogenous Input (NARX) models are an extension of autoregressive linear exogenous input models which are used for time series modeling of dynamic non-linear systems. NARX models have been used for time series modeling [24].

NARX is a dynamic feedforward network with feedback connections enclosing several layers of the network. NARX is particularly used in time-series analysis as it utilizes its memory ability using it to store the past values of predicted or true time series as one of the input to predict the next value.

As seen in Figure 2.6, there are two different architectures of NARX- 1) Series-Parallel (also called open-loop) and 2) Parallel architecture (also called closed-loop). Mathematically, open-loop and close-loop NARX architectures can be represented by Equations 2.7 and 2.8 respectively [25].

$$\hat{y}(t+1) = f(y(t), y(t-1), \dots, y(t-n_y), x(t+1), x(t), x(t-1), \dots, x(t-n_x)) \quad (2.7)$$

$$\hat{y}(t+1) = f(\hat{y}(t), \hat{y}(t-1), \dots, \hat{y}(t-n_y), x(t+1), x(t), x(t-1), \dots, x(t-n_x)) \quad (2.8)$$

where, $f(\cdot)$ is the mapping function of the neural network, $\hat{y}(t+1)$ is the predicted output by NARX at time t , $\hat{y}(t), \hat{y}(t-1), \dots, \hat{y}(t-n_y)$ are the previous outputs of NARX, $y(t), y(t-1), \dots, y(t-n_y)$ are true past values in the time series, $x(t), x(t-1), \dots, x(t-n_x)$ are the inputs to NARX, n_x and n_y are the number of input and output delays respectively.

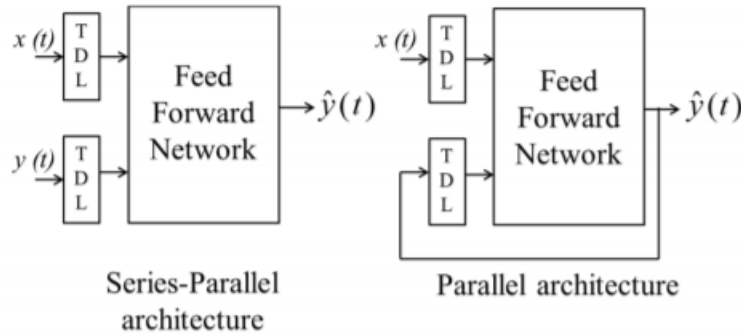


Figure 2.6: Architectures of the NARX neural network. Adapted from [25].

From the above equations and figures, it is clear that open-loop and close loop architecture differentiate themselves from each other only in terms of the input fed into the neural network. In close-loop architecture, the predicted output of the NARX neural network is used as a feedback to the feedforward network. But in cases where the true values are available, a open-loop or series-parallel architecture is used in which the true output is used instead of feeding back the estimated output.

2.5. Recurrent Neural Network

In this section, the need for Recurrent Neural Networks (RNN) to model sequential data along with the functioning and architecture of such networks is discussed.

2.5.1. Need for RNNs for Sequential Data

Traditional neural networks and other machine learning algorithms consider each sample in the training set to be independent of each other, that is they do not remember anything that has happened in the past. This can be a major limitation while working with data which is sequential in nature like time-series, video, speech, and so on. In sequential data types, there is an interdependency between individual elements across time. Similarly, predicting time-to-failure or detecting anomalies in a system is a time-series problem and RNNs are well suited to model such problems.

RNNs can work well with sequential data as they have an internal memory state which helps them to learn to selectively retain relevant information allowing them to capture dependencies across several time steps. Therefore, an RNN utilizes both previous information that it has learned and the current input to make future predictions.

2.5.2. RNN Theory

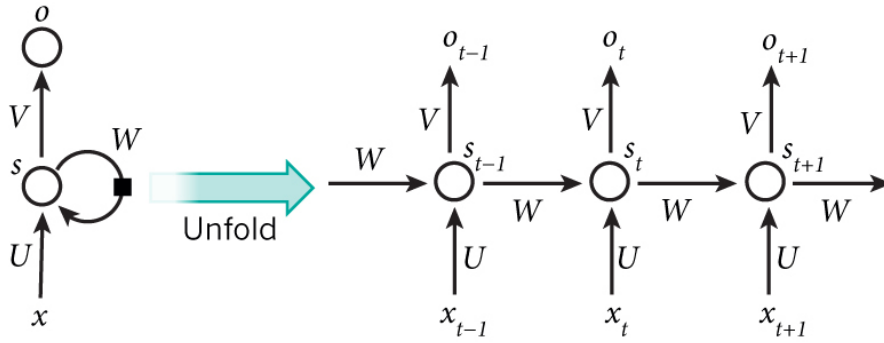


Figure 2.7: A Standard RNN. Adapted from [26].

RNNs are a powerful and robust type of neural networks which are suitable for processing sequential data. The presence of a state vector (in the hidden units) is one of the main features which distinguishes it from other types of neural network [26]. This state vector maintains the memory in the network. As seen in Figure 2.7, s is the state vector, and RNN has a feedback connection which connects the hidden neurons across time. At time t , the RNN receives the following inputs - the current sequence element x_t and hidden state from previous time step s_{t-1} . The state vector is updated to s_t and the output o_t is obtained. This ensures that the output s_t is dependent on all previous inputs x_{t-n} . U is the weight matrix between the input and hidden layers similar to a conventional NN. W is the weight matrix for the recurrent transition between one hidden state to the next which remains the same throughout the network. V is the weight matrix for hidden to output transition. Mathematically, the above simple RNN can be represented by equations 2.9 and 2.10, which refers to calculations taking place at every time-step in the network. Here, b_s and b_h are the biases added. σ and $f(\cdot)$ are the activation functions used in input and output layers respectively.

$$s_t = \sigma(Ux_t + Ws_{t-1} + b_s) \quad (2.9)$$

$$o_t = f(Vs_t + b_h) \quad (2.10)$$

2.5.3. Training RNN

Training of RNN is almost similar to training a feedforward neural network. A gradient based algorithm called *backpropagation through time* is used.

BackPropagation Through Time (BPPT): BPPT is an application of the backpropagation algorithm used for training recurrent neural networks [27]. BPPT is a gradient-based technique. Conceptually, BPPT works by unrolling the RNN in time. Each time step is treated as an individual neural network and trained as in backpropagation. Errors are then accumulated across each time-step and weights are updated. BPPT is computationally intensive and expensive. Also, it is a well-known fact that standard RNNs like vanilla RNN cannot be trained to learn long term dependencies [28].

There are two major problems with standard RNNs - 1) Exploding gradients and 2) Vanishing gradients. During training, the updating of weights is proportional to the partial derivative of the error function with respect to the current weight in each iteration of training. Therefore, while training the networks with longer dependencies, there is a high risk of the gradients becoming too large (due to large learning rates) or decay to zero. These problems are referred to as *exploding gradients* and *vanishing gradients* respectively. In such situations, the neural network might become unstable or training times can increase drastically.

The problems of exploding and vanishing gradients by certain methods like gradient clipping, changing the architecture of the neural network, or by changing the value of gradients [29]. Also, proper initialization of the W matrix can help reduce the effect on vanishing gradient. Using *ReLU* activation function instead of *tanh* or *sigmoid* is also one of the solutions. But the most effective solution to deal with these issues are to use LSTM networks or Gated Recurrent Unit (GRU), which are evolved types of RNNs. LSTMs, GRUs are very useful in modeling long term dependency problems and have become one of the most widely used types of RNN.

2.6. Long Short Term Memory

Long Short Term Memory (LSTM) is a type of unit in RNNs [30]. An RNN made up of LSTM units are called LSTM networks. LSTM networks were explicitly designed to solve the long term dependency problem.

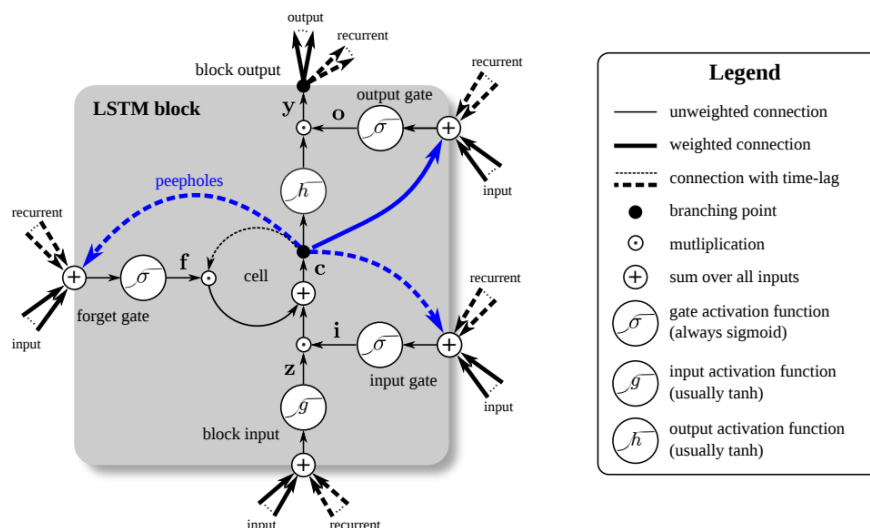


Figure 2.8: A schematic diagram of the LSTM unit with forget gates. Adapted from [31].

An LSTM cell consists of three gates namely input gate, forget gate and output gate as shown in Figure 2.8. The key to LSTMs is the cell state, the horizontal line running through

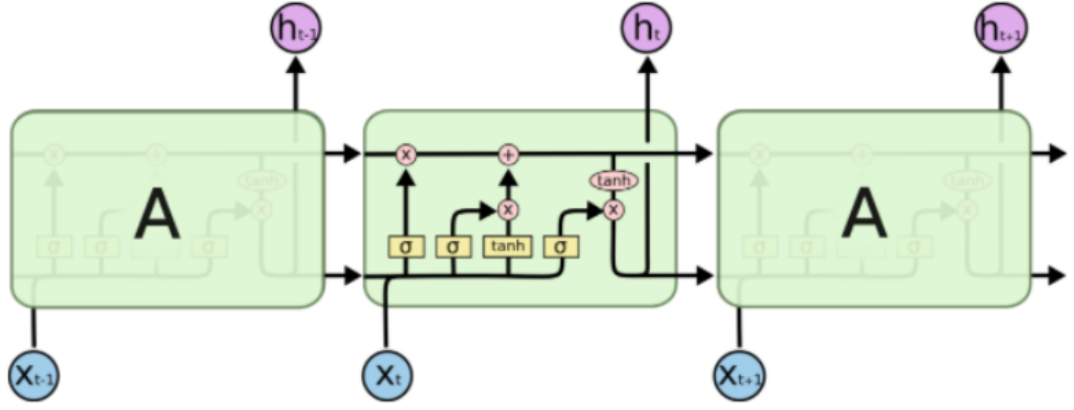


Figure 2.9: A schematic representation of LSTM network [32].

the top of the Figure 2.9. The three gate as mentioned above protects and controls the cell state.

Forget Gate Operation: The first step that is performed in an LSTM cell is to throw away unnecessary information from the cell state. This function is done by the sigmoid layer in the forget gate. It takes in previous state output h_{t-1} and input x_t from current time-step. The sigmoid functions then outputs f_t , which is a value between 0 and 1 for each number in the cell state C_{t-1} . If output value is 1, the information is passed through, or else if the output is 0 then the information is completely lost. A representation of the forget gate is shown in Figure 2.10.

The mathematical operation performed at the forget gate is given by Equation 2.11.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.11)$$

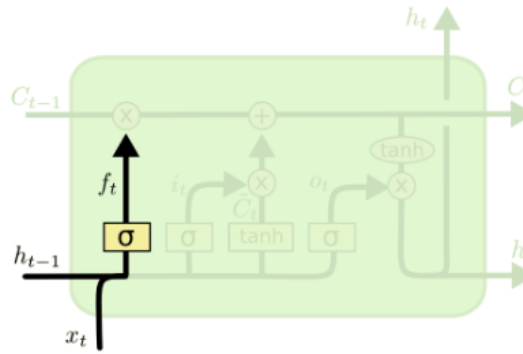


Figure 2.10: Forget gate operation.

Input/Update Gate Operation: The input/update gate is shown in Figure 2.11. In this step, the new information to be stored in the cell state is calculated. This step is split into two parts. In the first part, the sigmoid function also known as *input gate layer*, makes a decision on the values to be updates. Later, the \tanh function prepares a new vector \tilde{C}_t which is concatenated with the output of sigmoid function to update the new state to C_t . The mathematical Equations 2.12 and 2.13 represent the calculations taking place in input/update layers.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.12)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.13)$$

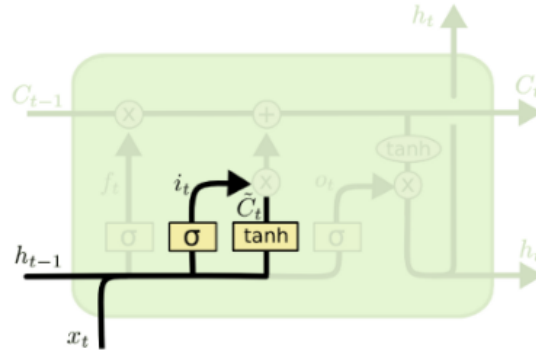


Figure 2.11: Input/Update gate operation.

In the next step, the old cell state C_{t-1} is updated to C_t . Here, the old state C_t is multiplied with f_t and then this is added to the product of i_t and \tilde{C}_t . The mathematical representation of the update step is shown in the Equation 2.14.

$$C_t = (f_t \times C_{t-1}) + (i_t \times \tilde{C}_t) \quad (2.14)$$

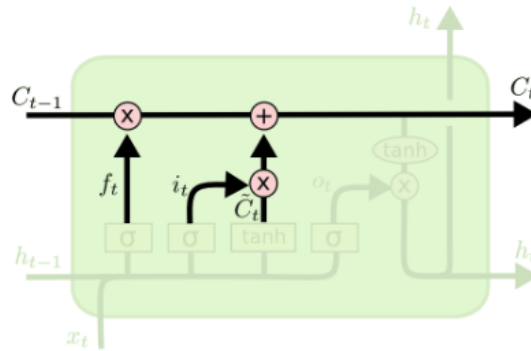


Figure 2.12: Cell state update.

Output Gate Operation: In this step, the final output of the LSTM cell is calculated. First, h_{t-1} and x_t is run through a *sigmoid* function to decide the part of the cell state that will sent to the output. The output o_t of the sigmoid function is multiplied with \tanh (to push values between -1 and 1). The mathematical representation of the output gate is shown in Equations 2.15 and 2.16.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.15)$$

$$h_t = o_t \times \tanh(C_t) \quad (2.16)$$

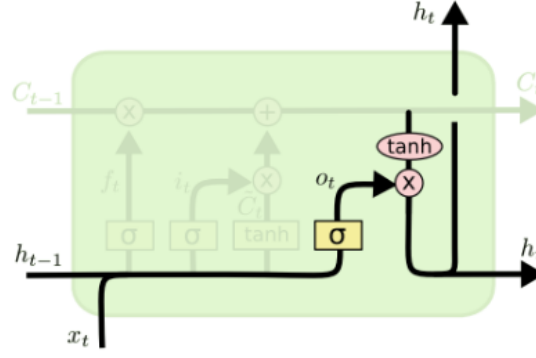


Figure 2.13: Output gate operation.

2.7. Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a type of recurrent unit similar to LSTM, with only two gates. GRUs were proposed by Chung et al., [33]. A GRU unit only has two gates namely update gate z and reset gate r . GRUs got rid of the cell state and used the hidden state to transfer information. The pictorial representation of a GRU is shown in Figure 2.14. With only two gates, the GRUs have found to require lesser computation and power, while having performance on par with LSTMs [33]. Just like the gates in LSTMs, the gates in the GRU are trained to selectively filter out any irrelevant information while keeping what's useful.

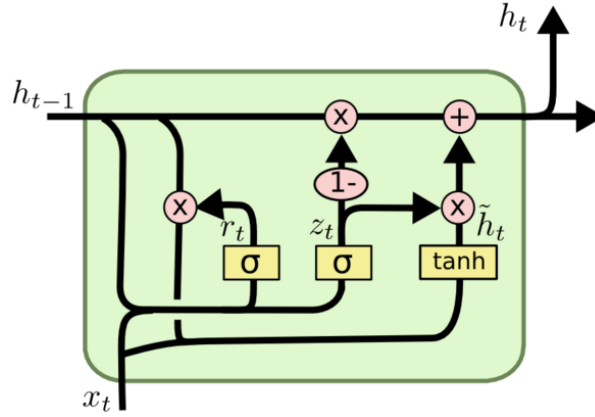


Figure 2.14: Pictorial representation of GRU.

The equations governing the gating mechanisms of GRU are 2.17 - 2.20.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (2.17)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2.18)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \times h_{t-1}, x_t]) \quad (2.19)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (2.20)$$

2.8. Relevant works on Fault Detection using ANN

Recently, there has been a lot of research interest in using SCADA data to model ANNs for condition monitoring of wind turbines. ANNs are quite robust and flexible in mapping non-linear functions. The ANN method is easily scalable for application on a large set of wind turbines. Some of the research that has been done in this field is discussed in this section.

Garcia et al., 2006 [34] used multilayer feedforward auto-regressive ANN for CM of wind turbine drivetrain. The model was trained on normal behaviour data. The ANN model output is compared with the measured value in real time, and a difference outside confidence bands, defined by the normal behavior model, is termed as an anomaly. In the case study, the model was about to predict a fault 26 hours in advance.

Zeher et al., 2009 [35] also used multilayer feedforward autoregressive ANN to model normal behaviour of the wind turbine gearbox bearings. The model was able to detect faults 6 months in advance. The anomaly detection was achieved by observing an increase in the frequency of the errors between the predicted and measured output parameter. This method of anomaly detection cannot be applied to a large set of turbines, and it is better to have a threshold value so that alarms can be set off in the CMS.

Kusiak et al., 2012 [36] proposed a feedforward ANN based fault detection model to predict generator bearing faults with different number of nodes in hidden layer. The best configuration consisted of 18 neurons, logistic hidden activation and identity output activation. The anomalies were detected based on observation of errors. This method could detect faults only 1.5 hours in advance, which does not allow for any kind of maintenance planning.

Zhang et al., 2014 [37], used ARX ANN models to predict min shaft rear bearing temperature. The model inputs were output power, nacelle temperature and turbine speed which were used as exogenous inputs. The model could predict faults three months in advance.

Bangalore et al., 2015 [38] used NARX ANN model for Normal Behaviour Modeling (NBM) of gearbox bearing temperatures. The selection of the training data was automated by using filtering and selection. The ANNs used power, gearbox oil temperature, nacelle temperature and the rotational speed as inputs as well as up to two additional temperatures of the other investigated bearings as inputs. Mahalanobis distance averaged over three days was used for detecting anomalies. A recorded gearbox failure due to spalling in one bearing was successfully detected one week before the vibration-based system identified the failure. Similar approach was adapted by Bangalore et al., 2017 [39], where NARX ANN is used to model NBM. Some additional data-preprocessing was such as cluster filter was done while preparing datasets. It was seen that better data preprocessing can build better models. A fault in the HSS bearing was detected three months in advance.

In the field of wind energy, there is not much literature available on the use of recurrent neural networks for condition monitoring using SCADA, although the use of RNNs for anomaly detection in various fields is increasing rapidly due the rise in computation power. Malhotra et al, 2015 [40], proposed stacked LSTM networks for anomaly detection of temporal time-series data. A network is trained on non-anomalous data and used as a predictor over many time steps. The resulting prediction errors are modeled as a multivariate Gaussian distribution, which is used to assess the likelihood of anomalous behavior. Nanduri et al., 2016 [41], used LSTMs and gated recurrent networks to detect anomalies from Flight Data Recorder. The RNN algorithms detected 9 out the 11 anomalies in the test dataset and performed better than feedforward networks. Jian Fu et al., 2019 [42] introduced the idea of using deep learning algorithms for temperature based fault detection in wind turbine gearboxes. A combination of convolutional neural network and LSTM is proposed and validated

against data from a real wind farm. Georg Helbing et al., 2018 [43] have also suggested exploring more recent RNN architectures like GRU and LSTM for better anomaly detection.

2.9. Research Problem and Objectives

Utilizing SCADA data for condition monitoring of wind turbines is becoming more appealing to asset owners as they come installed along with a standard wind turbine. Installing additional condition monitoring systems (CMS) systems to detect vibrations in the drivetrain, real-time oil quality monitors are expensive. As condition monitoring using SCADA data is a potentially low-cost solution requiring no additional sensors, several approaches using these data for early failure detection have been developed in recent years.

Presently, most of the condition monitoring using SCADA is based on data-driven models. Data-driven models, specifically artificial neural networks do a good job in mapping non-linear functions and learning from the data. One of the limitations of the data-driven models is that there is no physics involved in predicting the future state of the system.

Based on the literature study, two research questions are formulated :

"Is it possible to develop a robust condition monitoring technique for a wind turbine gearbox that utilizes both physics of failure model and machine learning techniques?"

"Can a correlation be deduced between White Etching Cracks and hypothetical failure theories using SCADA data and Physics of Failure technique?"

Based on the research problem, the research objectives of this thesis includes -

1. Using physics of failure techniques to carry out risk based assessment of wind turbine gearboxes on a wind farm level.
 - Using operational SCADA data to determine the tribo-mechanical behavior of the gearbox bearings and correlate them to the observed failures.
 - Correlating observed mode of failure- white etching cracks in high speed shaft bearings of gearbox to corresponding operating conditions on a wind farm level.
2. Exploring recurrent neural networks namely Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) for temperature based anomaly detection.
3. Investigating if utilizing both *Physics of Failure* and *Artificial Neural Network* models can help in robust condition monitoring of gearboxes.

2.10. Proposed Condition Monitoring Methodology

The research methodology for condition monitoring of gearboxes is split into 3 main parts. This study is based on SCADA data from an offshore wind farm off the coast of The Netherlands. The wind farm consists of 60 turbines of 2 MW rated power.

To improve the robustness of the condition monitoring system of gearboxes, a novel approach to combine physics of failure models with data driven models is proposed. This is shown in Figure 2.15. In the first part, a general methodology to assess turbine failure risk where the gearbox specifications is not available is proposed. This is shown in Figure 2.16.

White etching cracks are commonly seen in HSS bearings is a mode of failure which is not yet understood fully, as there are multiple postulates without much field data. A methodology to correlate the observed failures to the operational SCADA data on a wind farm level is shown in Figure 2.17.

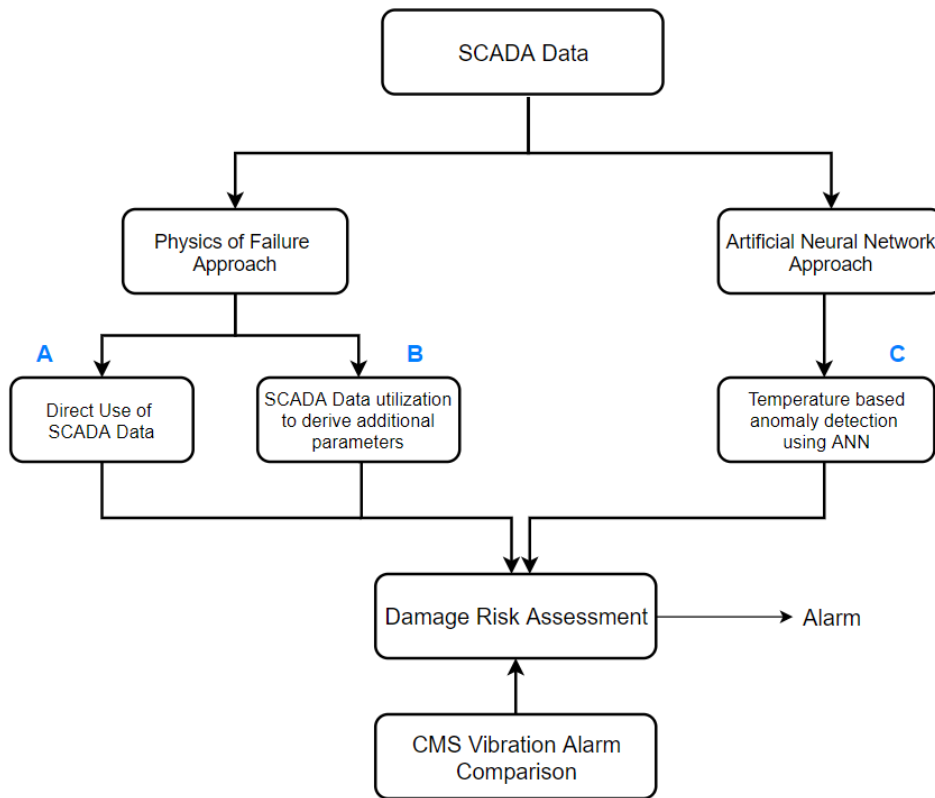
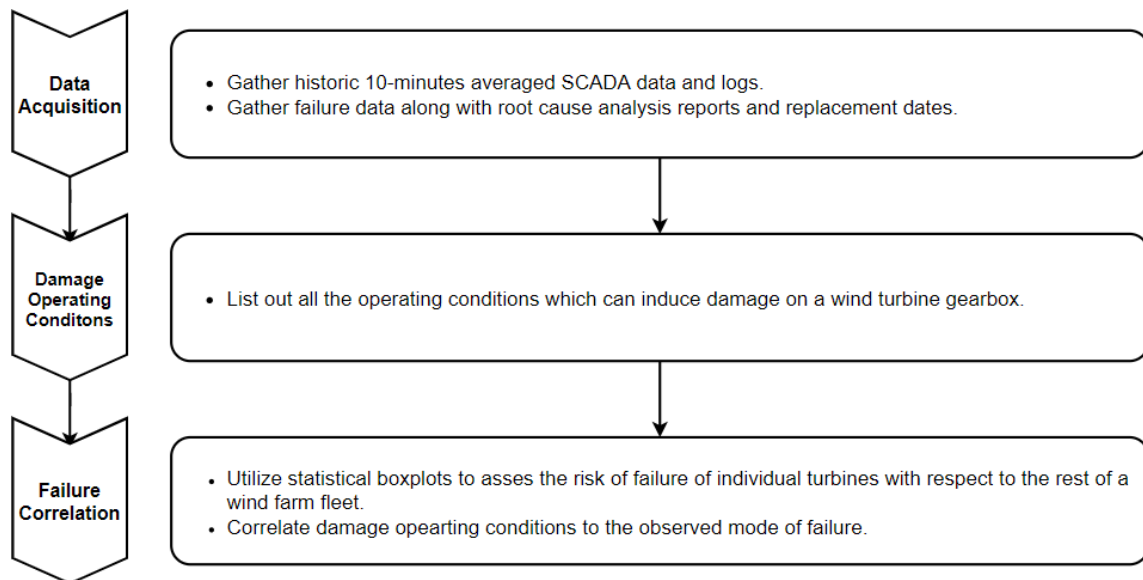


Figure 2.15: Overview of proposed condition monitoring method.

Figure 2.16: **A** - Proposed condition monitoring technique using only SCADA data.

An artificial neural network based on recurrent neural network is used for anomaly detection. The model is trained to predict the normal behaviour gear bearing temperature and gearbox oil temperature. The neural network is trained on fault free dataset, usually data recorded 2 years prior to failure. This is known as Normal Behaviour Modeling (NBM). The

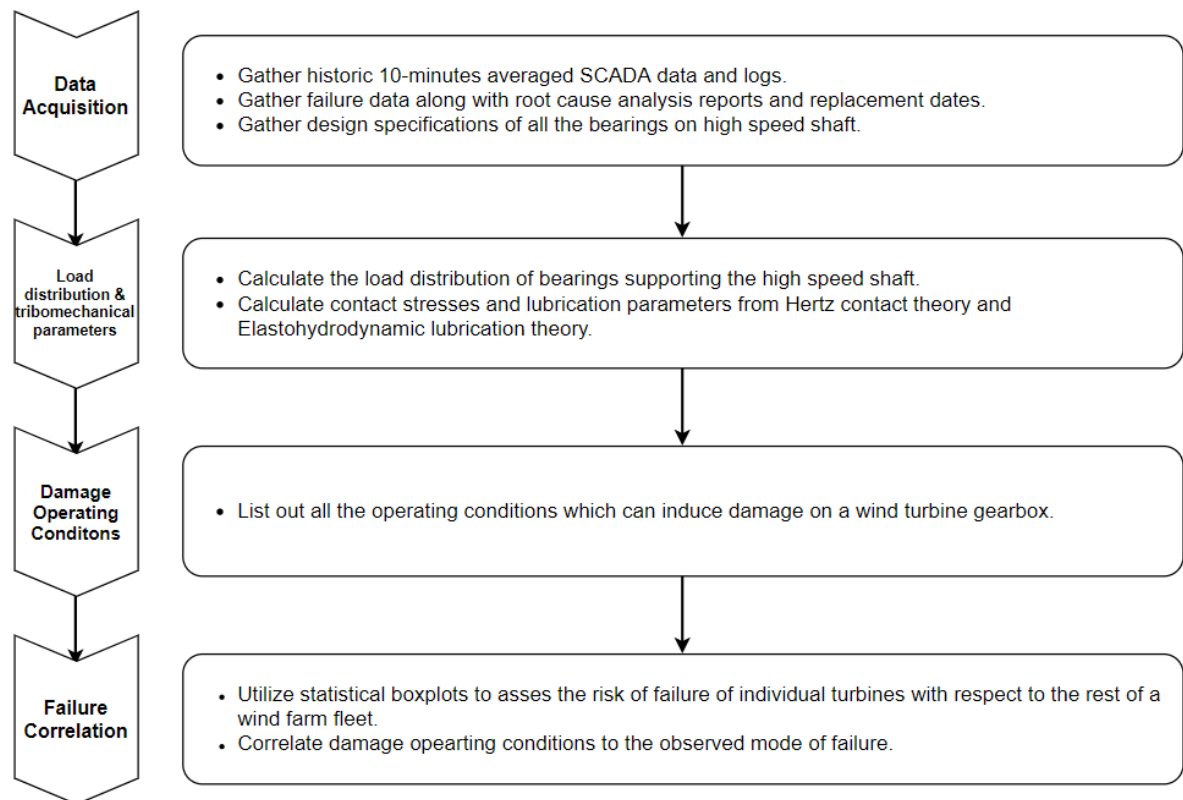


Figure 2.17: **B** - Proposed methodology to find the correlation between White Etching cracks and damage operating conditions in the High Speed Stage of gearbox.

residual or the difference in the ANN predicted and the SCADA measured temperature is used to flag anomalies if it crosses a certain pre-determined threshold. This is shown in Figure 2.18.

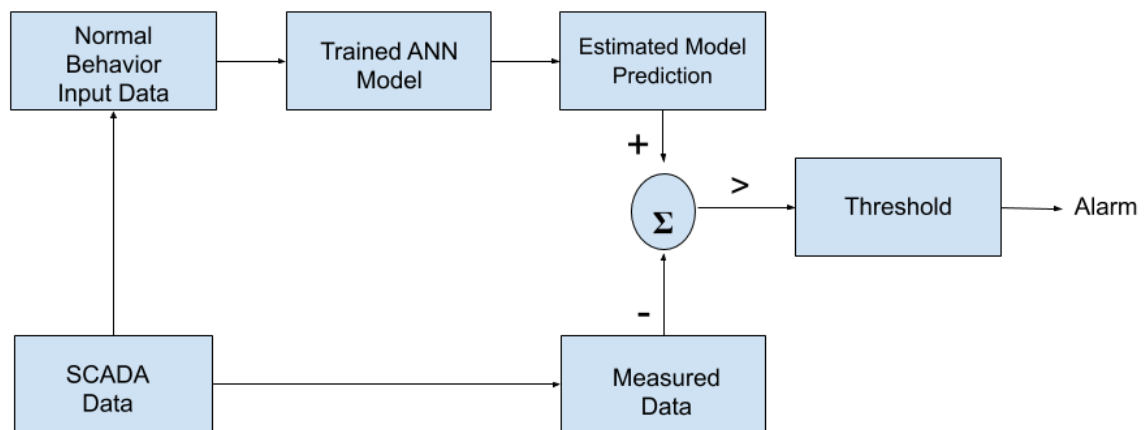


Figure 2.18: **C** - ANN based condition monitoring technique (Normal Behaviour modeling).

Failure Risk Assessment of Gearbox using Physics of Failure Technique

Wind turbine gearboxes fail due to a plethora of reasons, some of which are quantifiable and others which are still a mystery. The development of condition monitoring based on operational data is highly dependent on historical data to validate and test the proposed algorithms. In this chapter, various bearing failure modes observed in wind turbines are discussed, along with the basics of tribology and contact mechanics as most of the bearing failures are due to tribo-mechanical reasons. The usage of 10-minutes averaged SCADA operational data to quantify the gearbox damage is explored. The operational data of each individual turbine is used in quasi-state analytical models to determine bearing loads and is used to correlate observed failure to the damage-inducing operating conditions. This forms the basis of physics of failure approach. Physics of Failure (PoF) is a technique under the practice of Design for Reliability that leverages the knowledge and understanding of the processes and mechanisms that induce failure to predict reliability and improve product performance.

A comprehensive literature review on the unconventional but dominant failure mode seen in high speed shaft bearings, *White Etching Cracks (WECs)* is presented in the later half of the chapter. An effort is made correlate the operational data to the proposed WEC failure hypothesis by NREL Gearbox Reliability Collaborative (GRC) research and SKF research.

3.1. Introduction

In this study, the historic operational data of each individual turbine of an offshore wind farm is compared with the rest of the fleet. Most approaches to condition monitoring so far have ignored this fact and used only data from one turbine. Each turbine in a wind farm experiences slightly different loads due to wind shear, wake, and other such dynamic effects. Also, there might be small differences in the manufacturing and assembling of each individual turbine. Regardless of these small differences, all the turbines in a wind farm are subjected to the same seasonal variations and other generic weather-related conditions.

The motivation for exploring statistical-based condition monitoring of gearbox has been derived from some promising research done by Christopher Gray and Simon Watson [10], Jannis Tautz Weinert [44] and Gonzalez et al., [45]. Although the latter three were promising approaches to improve condition-based maintenance strategies, there is further potential in exploring the tribo-mechanical aspect of the operational data to further assess the risk of failure of wind turbines in a farm, as the access to the historic data (from commencement of wind farm) of the wind turbines is available.

3.2. Gearbox Specification

The offshore wind farm of interest has 2 MW rated geared wind turbines. The gearbox arrangement of the turbine is shown in Figure 3.1. The gearbox has 3 stages - a low speed stage consisting of planetary gear, an intermediate stage with a pair parallel helical gear and a high speed stage with another pair of parallel helical gear.

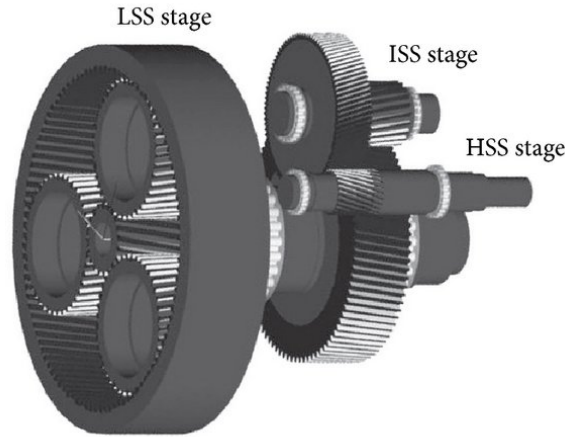


Figure 3.1: Wind turbine gearbox arrangement. Adapted from [46].

Based on the field failure data available, a vulnerability map of the gearbox is shown in Figure 3.2. The most vulnerable component of the gearboxes are the high speed stage (HSS) bearings, followed by low speed stage (LSS) bearings and gears. When a HSS bearing is damaged, the replacement of the whole gearbox is unnecessary as replacement of just the HSS bearings and IMS - B bearing suffices. A complete gearbox change is necessary when there is damage in the low speed stage of the gearbox.

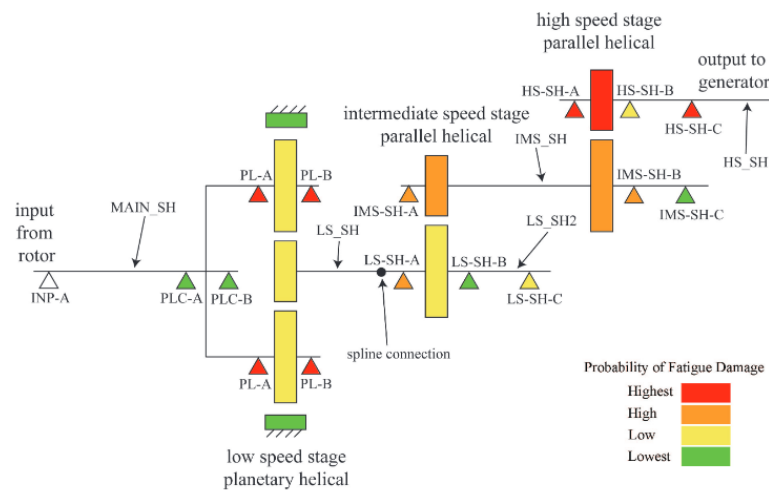


Figure 3.2: Wind turbine gearbox vulnerability map. Adapted from [47].

Although the components of the gearboxes are designed to last for at least 20 years, premature failures within 5 years of service occurs in HSS bearings, while some premature failure occurs in LSS stage within 7-10 years of operation [48]. As bearing manufacturers are not

part of wind turbine operating staff, they most often do not have full information regarding the complex loadings bearings have to endure, notably with unpredictable transient loadings [49].

Hence, in the first half of the chapter, the correlation of the damage operating conditions to the failure leading to complete gearbox overhaul, failure in the LSS and IMS gears and bearings is explored. In the later half, only HSS bearing failures are explored.

3.3. Damage Operating Conditions for Bearing Failures

The physics of failure approach by Gray and Watson, 2010 [10] looked retrospectively at statistical parameters such as average wind speed, hours at rated power, hours at rated speed etc. and could prove that failing bearings had seen higher cumulative loading. In this section, an analysis is made to see whether it is possible to monitor various kinds of failure modes with statistical parameters.

The following critical operating conditions are used to relate the failure seen in the field with the SCADA. Some of the parameters used are similar to the ones used by Gray et al., [10], along with certain additions.

- Rated power - Total number of hours where electrical power exceeded 90% of rated.
- Wind speed - Mean wind speed over the operating time of the turbine.
- Turbine standstill (0 rpm) - The number of hours that the turbine was standstill i.e rotor at 0 rpm at wind speeds above 10 m/s.
- Rated rpm - Total number of hours where rotor speed exceeded 90% of rated.
- Bearing temperature - Total number of hours when the bearing temperature is above 70°C.
- Rotor starts – Event count of rotor starts.
- Yaw movement – Total yaw movement in degrees.
- Emergency stops – Number of emergency stops during operational life, as recorded in the SCADA logs.

These parameters are plotted for every turbine until their failure and compared with the rest of the fleet with the help of boxplots. The calculation results for all 60 turbines are compared, with all values normalised relative to the sample mean. Box plots are used to show the distribution of results for the whole sample, with the minimum, 25th percentile, median, 75th percentile and maximum values indicated. In this report, box plots are mainly used to assess the turbines at higher risk of failure and the potential mode of failure due to which it might fail. The turbines with operating conditions in the upper region of the boxplot that is **above 75th percentile** is considered to be at a higher risk when compared to the rest of the turbines in the farm. With the help of Table 3.1, the relation between observed failure modes and corresponding operating conditions of the turbine is explored further.

MATLAB function *normalize*, is used to normalize the data. This function calculated the z-score of the data with center 0 and standard deviation 1. For a random variable X , z-score is calculated as shown in Equation 3.1. x is a data point in variable X .

$$z\text{-score} = \frac{x - X_{mean}}{X_{std}} \quad (3.1)$$

Table 3.1: Identification of failure root causes and critical operating conditions - Bearings

Failure mode(s)	Subfailure mode(s)	Cause of failure(s)	Damage operating condition(s)	Additional Comment(s)
Fatigue	Subsurface initiated fatigue	Hertzian fatigue/repeated cycles of high stress	1. Longer hours of operation at rated power/high rotational speed 2. High wind speed/turbulence	Under normal operating conditions, pure subsurface fatigue does not occur frequently/prematurely
	Surface initiated fatigue	Inadequate lubrication / high asperity contacts leading to material removal	1. High turbulence plus standstill rotor (i.e., 0 rpm) 2. Low rpm turbine idling	More common than subsurface initiated fatigue / causes microspits and microspall on surface
Wear	Abrasive wear	Oil degradation/oxidation/aging/contamination due to improper sealing	1. Prolonged operation without oil change/highly contaminated operating environment	Relatively uncommon in wind turbines with regular maintenance history
	Adhesive wear	1. High load in boundary/mixed lubrication regime 2. Sudden load variation causing microslips/skidding in bearings	1. High turbulence / gusts 2. Sudden acceleration from low rotational speed under high bearing load i.e., start/stop operation	Also known as smearing, causes surface microspits
Corrosion	Moisture corrosion	Formation of corrosive rust	1. Moisture content in lubricant 2. Operating in highly corrosive environment with high salt content	
	Frictional corrosion	Loss of lubricant at contact, microslip due vibration loads causing material loss at contacting surface	1. High turbulence plus standstill rotor (i.e., 0 rpm)	
Plastic deformation	Overload	1. Static/shock loads leading to plastic deformation/indentations	Emergency stops/ sudden shut-down/grid connection losses	High von misses stresses below surface leading to subsurface crack
	Indentation from debris	1. Rolling of bearing surfaces over debris from contamination	1. Prolonged operation without oil change/contaminated lubricant	
	Indentation while handling	Improper handling during bearing installation/transportation	N.A	
Fracture and cracking	Forced fracture	Improper handling / using hammer during installation	N.A	
	Fatigue fracture	Fatigue strength exceeded due to cyclic bending		
	Thermal cracking	1. Friction from sliding causing heating of bearing rings 2. Bearing not correctly seated in housing	1. High amount of bearing sliding due to various causes like low loads / sudden acceleration leading to local hot spots	

Table 3.1 contains details about various bearing failure modes as defined by ISO 15243:2017 [50] and the potential damage operating conditions as observed in wind turbines.

3.4. Case Study - Part I

In this section, 4 failure case studies with different recorded failure modes are described. Most of the cases presented in this section are related to LSS and IMS gearbox damages which requires complete gearbox overhaul. A thorough analysis of bearing lubrication and contact stresses for the LSS and IMS could not be done, due to initial unavailability of the bearing make and specifications. Also, the root cause analysis reports are unclear for the LSS and IMS bearings and gears. For example, there is no mention if the spalling failures originated due to surface or subsurface initiated causes.

3.4.1. Turbine 26 - Planetary stage bearing spalling

The first investigated gearbox replacement took place in April 2010, just 2 years from commencement of operations. This is a classic case of premature gearbox failure. After a thorough inspection, spalling on one of the planetary stage bearing was reported as the root cause for failure. In the inspection reports, a lot of scuffing and standstill marks in the IMS pinion, and spalling on the sun pinion is also reported.

Here, the reason for spalling in turbine 26 might be due to a high number of emergency stops and higher amount of time when the turbine is in standstill position, leading to overloading in the bearings and gears. This can be seen in Figure 3.3. Here, the operating conditions of turbine 26 is compared with the rest of the fleet consisting of 60 turbines.

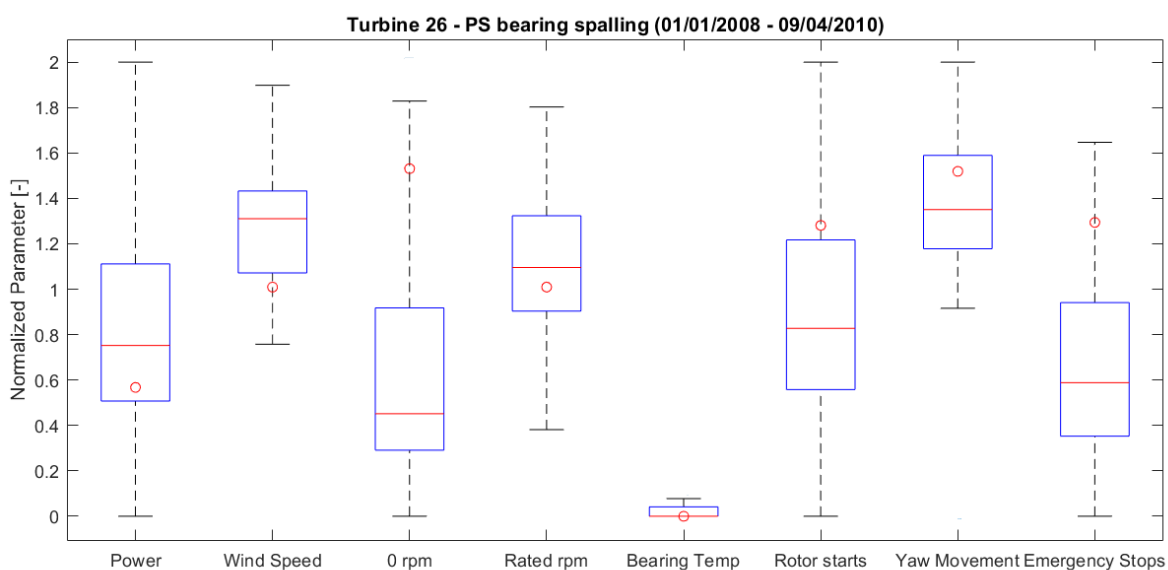


Figure 3.3: Turbine 26 - PS bearing spalling.

3.4.2. Turbine 32 - PS planet pinion pitting

The gearbox replacement of turbine 32 took place in August 2014. The root cause analysis reported pitting on planet pinion as primary cause of failure. Also, micropitting on almost all the gears and indentation from debris was found in almost every bearing of the wind turbine.

In figure 3.4, it can be seen that the wind turbine had been operating at high load, compared to the rest of the fleet prior to failure. Also, the turbine has had high number of rotor starts

and emergency stops when compared to the rest of the fleet. From the boxplot, it is easy to assume that the cause of failure might be due to high cycle fatigue, as the turbine is operating at higher loads for longer periods of time. The high loads only facilitate expediting the damage, while the primary cause of failure may be due to lubrication issues. Therefore, it is necessary to include tribological phenomenon in such boxplots for failure analysis.

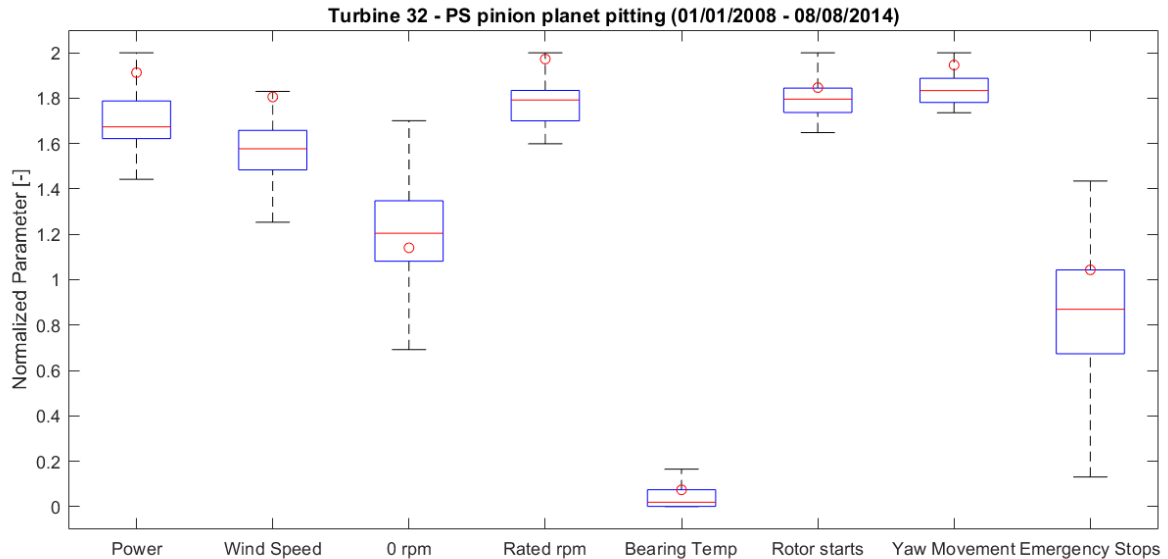


Figure 3.4: Turbine 32 - PS planet pinion pitting.

3.4.3. Turbine 2 - PS gear ring fracture/cracking

Turbine 2 gearbox failure was reported in May 2015. The primary cause of failure is a fracture/cracking in the gear ring of the planetary stage. It is also reported that this might have happened due to debris from a needle bearing of a mechanical pump, which is not a part of the drivetrain, but it is not ascertained. There are higher chances of a crack occurring in gears when the number of turbine start-stop cycles are high as the lubricant thickness might be less than optimum during the start of turbines [51].

Apart from the gear ring fracture, indentation and scuffing marks were also found in other gears and bearings. Fretting corrosion was also reported in one of the IMS bearings. From the boxplot as shown in Figure 3.5, it can clearly be seen that the turbine has many damage operating conditions in the upper 75th percentile, which means it was always at a higher risk of failure ¹.

3.4.4. Turbine 53 - IMS bearing spalling

The failure in turbine 53 was reported in April 2014. The main cause for replacement of the gearbox was spalling/flaking in the IMS bearing. From the boxplots in Figure 3.6, no clear indication of the failed turbine being at risk prior to failure is visible. All the damage operating conditions are below well below 75th percentile.

Flaking failures are mainly caused due to overloads or poor lubrication [52]. The number of emergency stops is well below 25th percentile; emergency stops are one of the major causes for overloading in bearings. This failure shows the importance of considering lubrication effects such as lambda ratio as one of the damage operating conditions. This is further discussed in the following section.

¹Bearing temperature parameter is omitted in this case study, as it was realized that this parameter gave no correlation to any failure. This is explained in later sections.

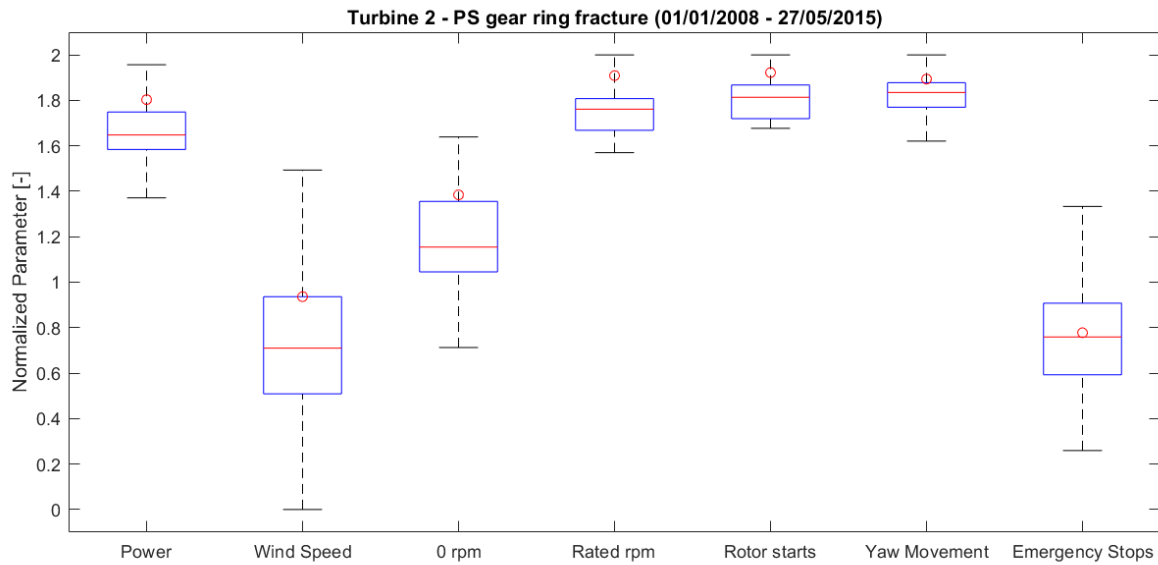


Figure 3.5: Turbine 2 - PS gear ring fracture/cracking

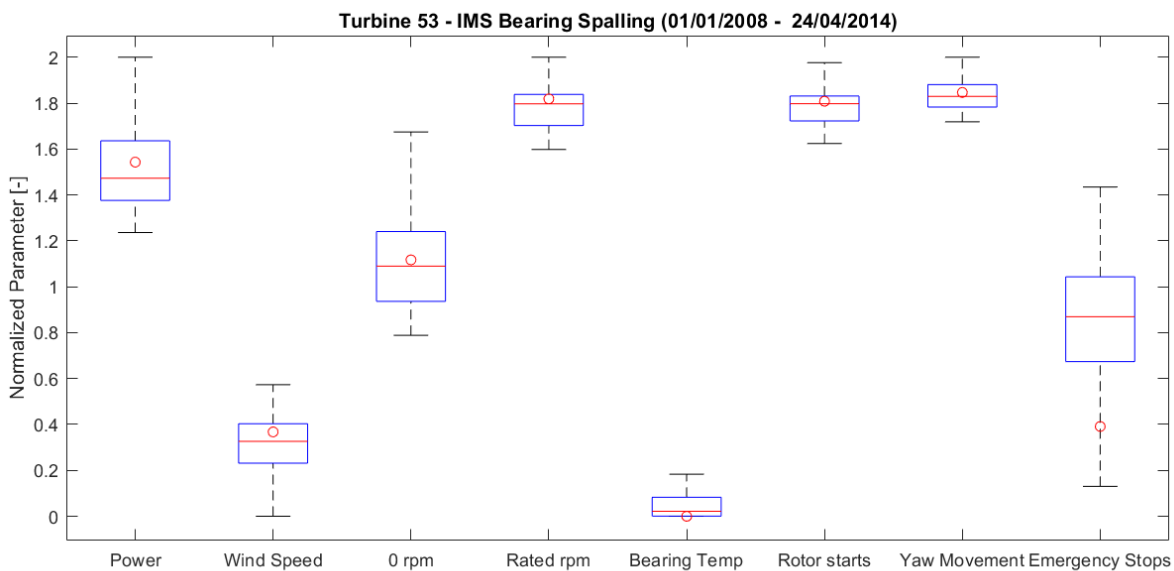


Figure 3.6: Turbine 53 - IMS bearing spalling.

The above damage operating conditions approach to assess turbine failure risk can perform better if the influence of elasto-hydrodynamic lubrication (EHL) and contact mechanics of the bearings is included.

The EHL and contact stresses are different for each individual bearing, dependent on the bearing load capacity, size, and loads on bearings. Apart from major gearbox overhauls, there have been several cases of HSS bearing failures, wherein only the bearings on the HSS and one of the IMS bearings are replaced. Therefore in this thesis, the focus is placed on the specific failures in the HSS bearings. All the HSS bearing failures observed in the wind farm of study is axial cracking or white etching cracks. In the following sections, the basics of contact mechanics, contact lubrication, white etching cracks and their causes are discussed.

3.5. HSS Bearing Configuration and Load Distribution

The high-speed shaft of the gearbox is coupled with the generator. They usually operate at high speeds, approximately 1500-1700 rpm. The torque transfer from the IMS shaft to the HSS shaft is done with a pair of parallel helical gear. The pinion on the HSS induces torque onto the shaft. The HSS is supported by three anti-friction roller bearings. The rotor side consists of a cylindrical roller bearing (CRB), while the generator side is supported by a pair of tapered roller bearings (TRB). The TRB closer to the pinion is also known as upwind TRB, while the other TRB away from the pinion can also be referred as downwind TRB.

The CRB can only support radial loads, while TRBs are designed to support both radial and axial loads. But in particular, in wind turbines, the TRB closer to the pinion supports most of the radial load apart from CRB, while the TRB away from the pinion takes the axial loads. A pair of TRB is used to accommodate axial loads in both directions, for instance during torque reversal events. The bearing design and properties are detailed in Appendix A.

An analytical model to assess the forces on the bearings is developed. The free-body diagram of the high-speed shaft is shown in Figure 3.7. The following assumptions were made in developing the free-body diagram:

1. The shaft is rigid and weightless.
2. The weight of the disc-brakes and pinion is negligible.
3. The CRB does not carry any bending moment or react to axial force.
4. The upwind TRB does not support any axial loads.
5. The TRBs do not carry bending moments, because their radial stiffnesses are much greater than their tilting stiffnesses.
6. The generator coupling does not transmit moments or axial force, i.e., forces due to torque reversals are excluded.

Most of the assumptions are valid for normal operating conditions, and was verified experimentally by NREL GRC [53], [54]. The bending moment is ignored as it cannot be determined from SCADA or design considerations, and also the bending moment imparted by generator coupling due to misalignment can be considered negligible during normal operation.

To determine the loads acting on bearings that support the shaft, it is important to first analyze the forces acting on the shaft due to pinion. The loads acting on the pinion can be divided into three main components - tangential, radial and axial.

The torque in the HSS shaft can be determined with turbine power output and generator rpm using Equation 3.2 where T is the HSS torque, ω is the generator RPM. Figure 3.8 shows the linear relationship between power output and torque.

$$T = \frac{P \cdot 60}{2 \cdot \pi \cdot \omega} \quad (3.2)$$

The tangential load on the shaft is only due to the torque. The tangential load on shaft can be determined using the pitch circle diameter (PCD) of the pinion and HSS torque using Equations 3.3 and 3.4. Here, F_m is the tangential load on shaft and β is the helix angle of the pinion.

$$PCD = \text{Number of teeth} \times \text{module} \quad (3.3)$$

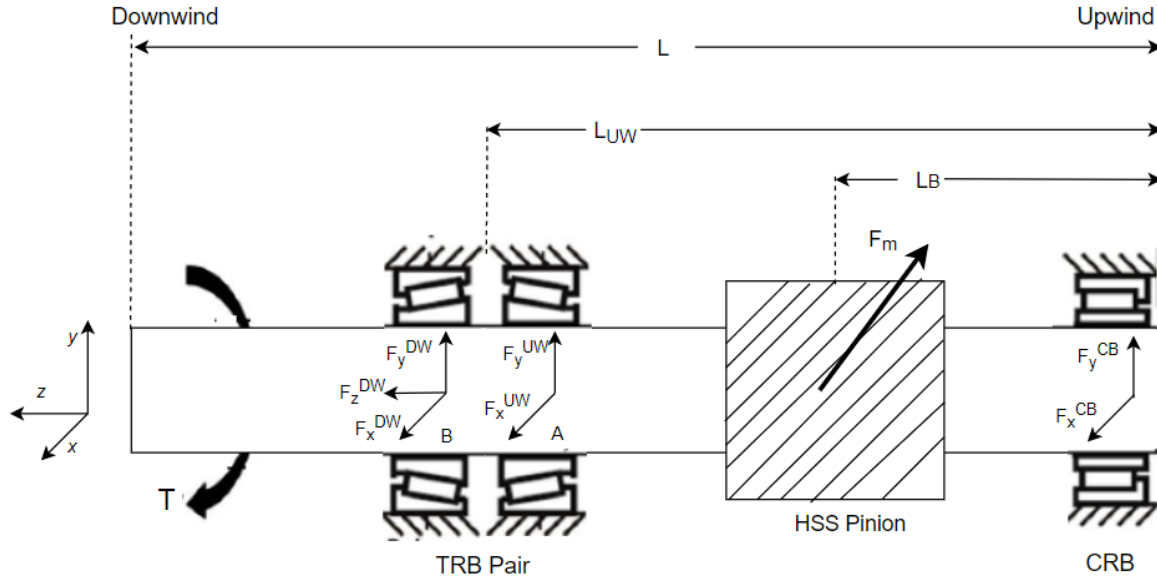


Figure 3.7: HSS free-body diagram.

$$F_m = \frac{T}{0.5 \cdot PCD \cdot \cos \beta} \quad (3.4)$$

The radial gear load, F_s on the shaft is calculated using equation 3.5. Here, α_p is the pressure angle of the pinion.

$$F_s = F_m \cdot \frac{\tan \alpha_p}{\cos \beta} \quad (3.5)$$

The axial load, K_a acting on the shaft is calculated using equation 3.6.

$$K_a = F_m \cdot \tan \beta \quad (3.6)$$

The right angle shaft/resultant load as a consequence of radial and tangential load on the shaft is equated using equation 3.7.

$$K_r = \sqrt{F_m^2 + F_s^2 + K_a^2} \quad (3.7)$$

After calculating the loads on the shaft, the following Equations 3.8-3.13 can be used to calculate the load sharing between the bearing supporting the high speed shaft.

Load sharing: The radial load due to resultant load K_r is not evenly distributed between the CRB and TRB pair. As the CRB lies closer to pinion, around 60% of the radial load acting on the shaft is carried by them. The remaining radial load of around 40% is carried by the upwind TRB and downwind TRB [53]. When the tapered roller bearing are placed in face-face configuration, the radial load acts in the center of the bearings as shown in Figure 3.9.

The radial load sharing between the CRB and the TRB pair is calculated using Equations 3.8 and 3.9 respectively. The value of pressure angle, α is 10° and the helix angle β is 14° .

$$F^{CB} = \sqrt{(F_x^{CB})^2 + (F_y^{CB})^2} = \left(\frac{L_{UW} - L_B}{L_{UW}} \right) K_r \quad (3.8)$$

$$F_r^{TRB} = \sqrt{(F_x^{UW})^2 + (F_y^{UW})^2} = \left(\frac{L_B}{L_{UW}} \right) K_r \quad (3.9)$$

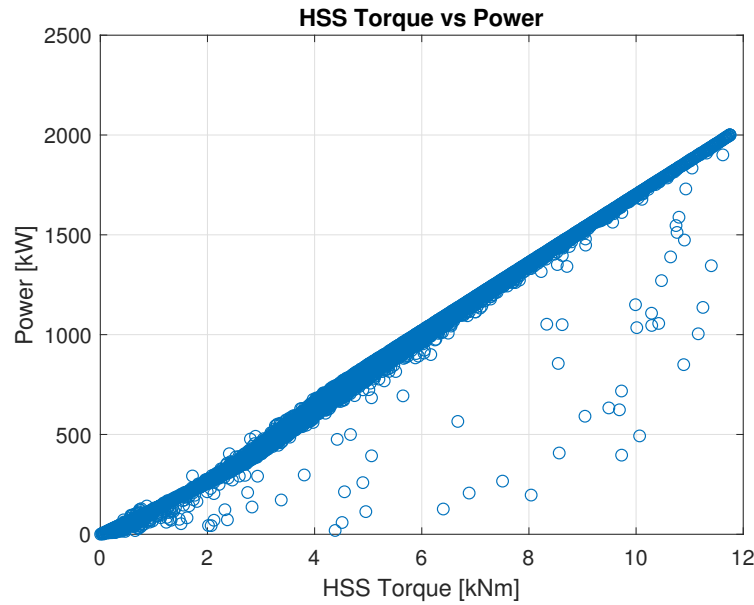


Figure 3.8: Plot of Power versus HSS Torque for one year of SCADA data.

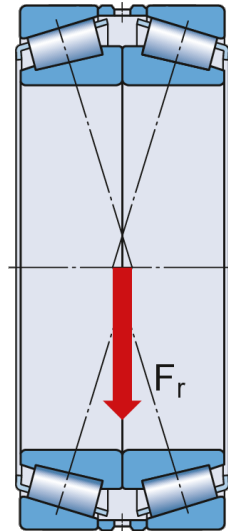


Figure 3.9: TRB face-face matched pair radial load sharing [55].

When a radial load is applied to a single row tapered roller bearing, the load is transmitted from one raceway to the other at an angle to the bearing axis and an internal axial load is induced. This internal axial load must also be considered along with the gear mesh induced axial force to calculate the equivalent axial force acting on the TRB. Also using TRBs in pairs, axial loads in both the directions can be supported. For this reason, these bearings are used in pairs (face-to-face arrangement). The face-to-face arrangement can accommodate larger tilting moments.

In case of TRB arranged in face-to-face configuration, the axial loads on the bearings can be calculated using equations 3.12 and 3.13. Here, F_{aA} and F_{aB} are the axial loads on upwind and downwind TRB respectively. In Figure 3.10, A is the upwind TRB and B is the downwind TRB. Y_A and Y_B are the axial load factor, which depends on bearing size, load capacity and taper angle of rollers. The load case expressed in equation 3.10 and 3.11 has to be satisfied before calculating the axial loads. We assume that there is no preload applied in the TRBs, as

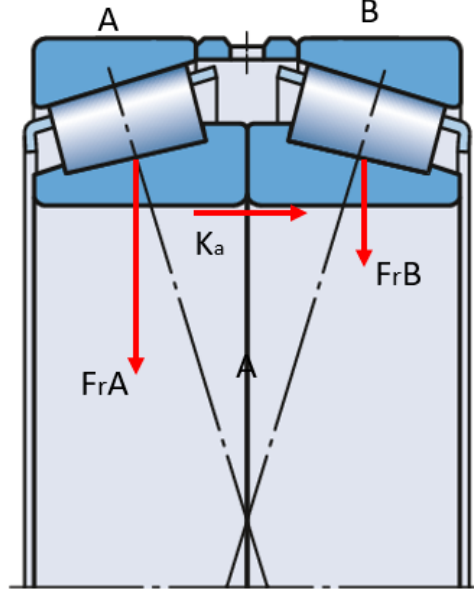


Figure 3.10: TRB face-face matched pair axial load sharing [55].

there is no sufficient information available. This quasi-steady analytical model was validated experimentally and using multibody simulation software, SIMPACK by NREL GRC [54].

$$\frac{F_r A^{TRB}}{Y_A} < \frac{F_r B^{TRB}}{Y_B} \quad (3.10)$$

$$K_a \geq 0.5 \times \left(\frac{F_r A^{TRB}}{Y_A} - \frac{F_r B^{TRB}}{Y_B} \right) \quad (3.11)$$

$$F_a A = \frac{0.5 \cdot F_r A^{TRB}}{Y_A} \quad (3.12)$$

$$F_a B = F_z^{DW} = F_a A + K_a \quad (3.13)$$

The histogram of torques and various loads on the roller bearings are shown in Figures 3.11-3.14. It gives a representation of the magnitude of loads acting on the bearings. As expected, most of the radial load is supported by CRB. The radial load on CRB at rated power is around 145 [kN]. The radial and the axial loads on the TRB pair at rated power is approximately 34 [kN] and 45 [kN] respectively. The HSS load distribution on shaft and bearings for a 750[kW] wind turbine was simulated in Transmission 3D to validate the analytical model developed by NREL GRC [53]. The results obtained in the finite element model in Figure 3.15 also suggests that most of the radial load is supported by the CRB and upwind TRB. The downwind TRB is loaded on almost all the rollers, which suggests that it takes almost all the axial load when operating at rated power. Therefore, the downwind TRB experiences higher loading when compared to upwind TRB. All the calculations are done using the operational data from wind turbine - 48 for the year 2014.

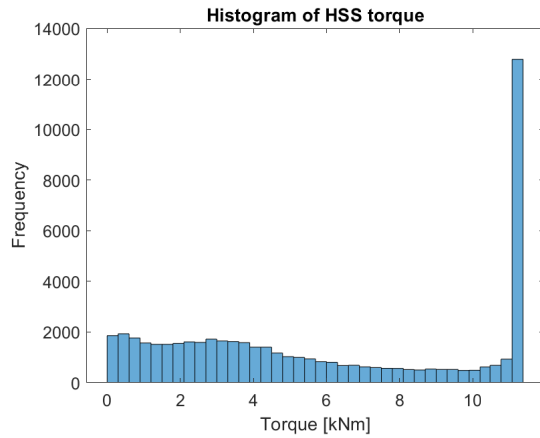


Figure 3.11: Histogram of HSS torque taking one year SCADA data.

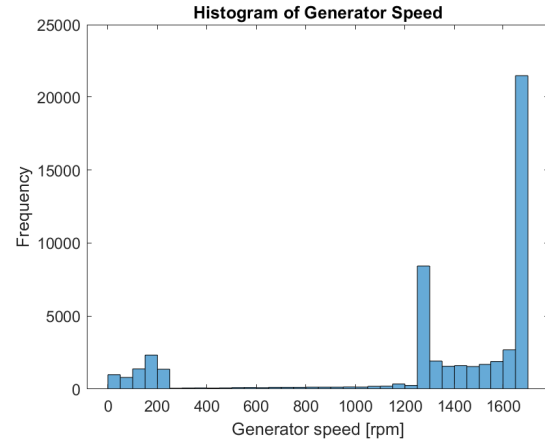


Figure 3.12: Histogram of generator speed taking one year SCADA data.

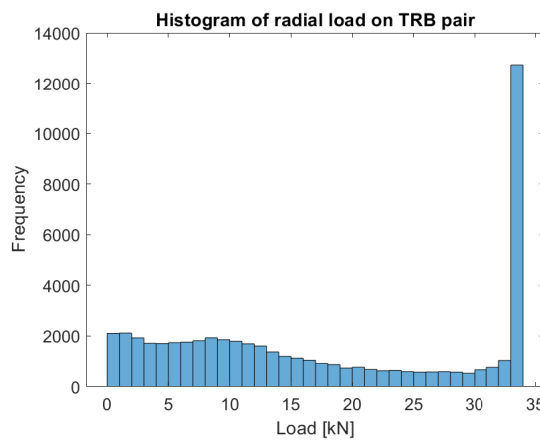


Figure 3.13: Histogram of radial load on TRB pair taking one year SCADA data.

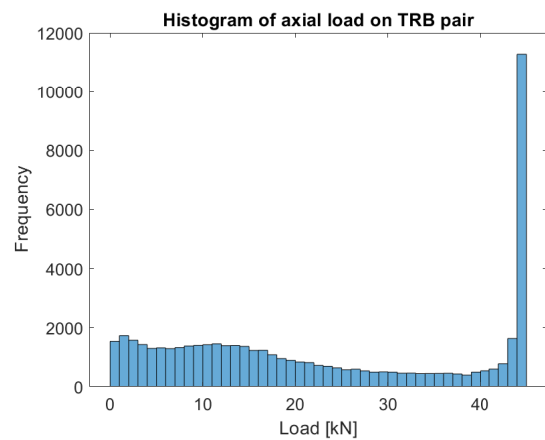


Figure 3.14: Histogram of axial load on TRB pair taking one year SCADA data.

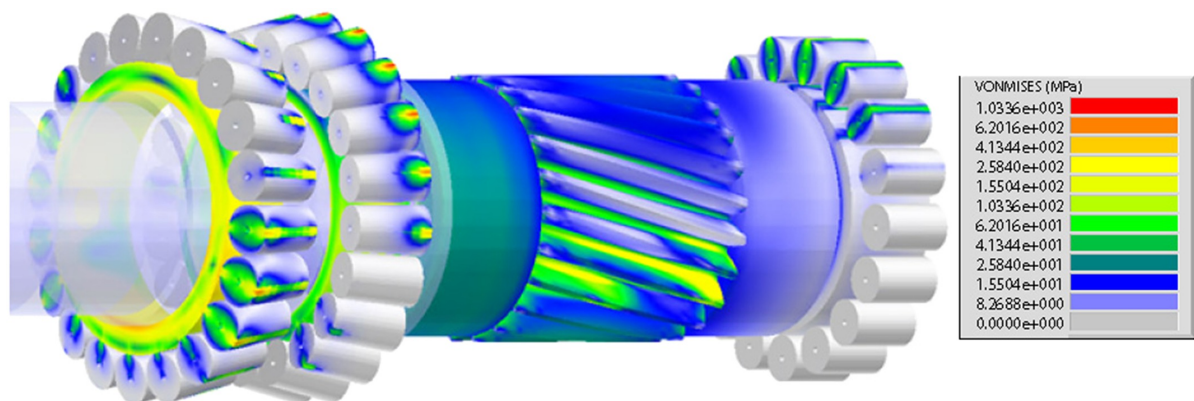


Figure 3.15: High speed shaft and bearing roller loads at 100% power (750 KW) calculated using the Transmission3D model. The contact areas on the rollers are highlighted. Adapted from [53].

3.6. Bearing Rating Life

The fatigue life of an individual bearing is the number of revolutions that a bearing operates before the first signs of metal fatigue (rolling contact fatigue (RCF) or spalling). The bearing

rating life, L10 is the measure of reliability, where 90% of the bearing population survives. Or, conversely, 10% of bearings will have failed in the L10 number of service hours. L10 is usually expressed in millions of cycles or number of hours.

In this section, calculation of L10 rating life, based on the latest standards, ISO 281:2007 [56] is assessed to verify whether axial cracks observed in HSS bearings took place due to material fatigue. ISO 281:2007 does not cover the influence of wear, corrosion and electrical erosion on bearing life as they do not take any surface parameters such as roughness, hardness, tribo-chemical aspects, sliding of the contacting surfaces into consideration. The L10 reliability is purely based on the assumption that cracks appear in the subsurface of the material due to high cycle fatigue.

The L_{10} according to ISO 281:2007 is calculated using Equation 3.14 -

$$L_{10} = a_1 \cdot a_{ISO} \cdot \left(\frac{C}{P_{dy}} \right)^p \quad (3.14)$$

where,

- a_1 is the life adjustment factor for reliability
- a_{ISO} is the life modification factor.
- C is the basic dynamic load rating.
- P_{dy} is the equivalent dynamic bearing load.
- $p = 10/3$ for roller bearings, 3 for ball bearings.

3.6.1. Equivalent Dynamic Bearing Load (P)

The equivalent dynamic bearing load is a hypothetical load with a constant magnitude and direction, that acts radially on radial bearings and axially and centrically on taper bearings. This load is considered to be equivalent to the actual loads acting on the bearings. If a radial and axial load having a constant direction and magnitude is applied simultaneously on a bearing, the equivalent dynamic bearing load can be defined using Equation 3.15.

$$P_{dy} = X \cdot F_r + Y \cdot F_a \quad (3.15)$$

Here, X and Y are the radial and axial bearing load factor for the bearing respectively. F_r and F_a is the actual radial load and axial load on the bearing.

For a CRB which can support only radial loads, Equation 3.15 is reduced to $P = F_r$. For a TRB pair, calculating the dynamic equivalent load is more complex. It is dependent on the ratio of axial load to the radial load supported by the bearing pair.

If $F_a/F_r \leq e$, then $P_{dy} = F_r + Y_A \cdot F_a$. If $F_a/F_r > e$, then $P_{dy} = 0.67 \cdot F_r + Y_B \cdot F_a$. The value of e is 0.43 for the used bearing as provided by the manufacturer.

3.6.2. Life Modification Factor a_{ISO}

The life modification factor is a function of the following factors-

- the fatigue load limit in relation to the acting bearing equivalent load (P_u/P_{dy})
- the lubrication condition (viscosity ratio κ)
- contamination level in bearing (η_c)

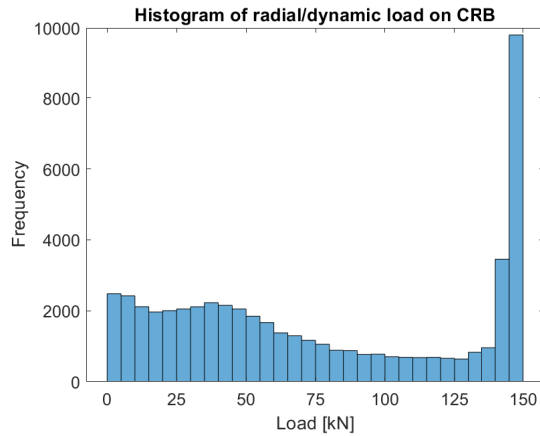


Figure 3.16: Histogram of radial load on CRB taking one year SCADA data.

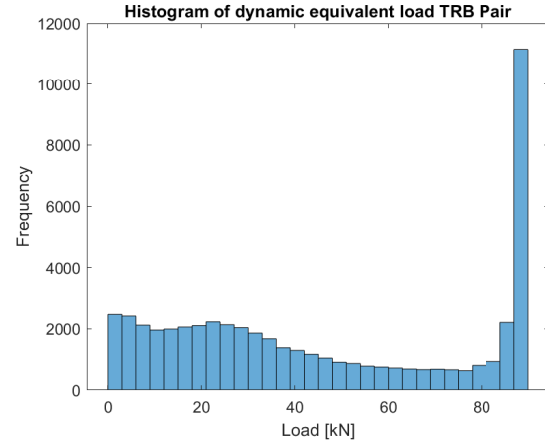


Figure 3.17: Histogram of dynamic equivalent load on TRB pair taking one year SCADA data.

The viscosity ratio κ is defined as the ratio between the kinematic viscosity to the rated viscosity of the lubricant used in the bearings. This ratio gives information about metal-to-metal contact in bearings. The κ values range between 0.1 and 4, where 0.1 refers to full metal-to-metal contact and a value of 4 refers to operation where the lubricant film around the contacting surface is fully developed and there is no metal-metal contact leading to surface wear.

$$\kappa = \frac{\nu}{\nu_1} \quad (3.16)$$

The kinematic viscosity, ν , is a function of lubricant density and operating temperature. Usually, the lubricant manufacturers provides the values for kinematic viscosity at 40°C and 100°C. The kinematic viscosity at any given operating temperature can be calculated from these values using the Ubbelohde-Walther equation adapted in ASTM D341-17 [57]. Equation 3.17 is used to calculate viscosity at any temperature T in Kelvins. A and B are constants.

$$\log_{10} \log_{10}(\nu + 0.7) = A - (B \times \log_{10}(T)) \quad (3.17)$$

The rated viscosity ν_1 is function of bearing speed (n) and pitch diameter (D_{pw}) of the bearing, as shown in equation 3.18 [56].

$$\nu_1 = \begin{cases} 45000 \cdot n^{-0.83} \cdot D_{pw}^{-0.50}, & \text{if } n < 1000 \text{ rpm} \\ 4500 \cdot n^{-0.50} \cdot D_{pw}^{-0.50}, & \text{if } n \geq 1000 \text{ rpm} \end{cases} \quad (3.18)$$

The contamination factor η_c describes the level of solid contamination of the lubricant. Contamination factor influences the fatigue life of bearings. A η_c value of 1 means perfectly clean conditions without any indentations and a value closer to 0 means severely contaminated conditions resulting in pronounced indentations. From ISO 281 recommendation for roller bearings with diameter greater than 100 [mm] and high cleanliness conditions, a cleanliness factor η_c value of 0.8 is chosen [56].

A function $\eta_c(P_u/P)$ is defined to choose the a_{ISO} value. Figure 3.18 shows plot of the life modification factor a_{SKF} which is equivalent to a_{ISO} for radial roller bearings.

3.6.3. L10 Life based on Field Operating Conditions

In wind turbines, the operating conditions such as wind speed, temperature, speed, lubrication conditions are continuously changing. An adaptation of Palmgren-Miner rule is used to calculate L10 life when the operating conditions are transient and changing continuously.

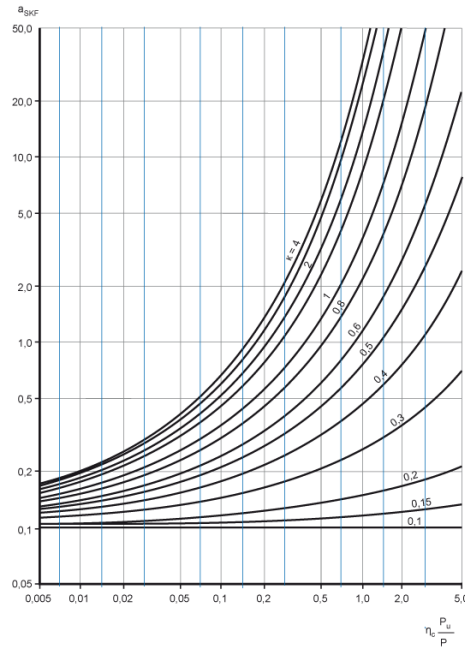


Figure 3.18: Life modification factor, a_{SKF} . Adapted from [58].

If N_1 equals the number of revolutions required under the load condition P_1 , and N is the expected number of revolutions for the completion of all variable loading cycles, then the cycle fraction $U_1 = N_1/N$ is used by the load condition P_1 , which has a calculated life of L_{10mi} . Under variable operating conditions, bearing life can be rated using Equation 3.19.

$$L_{10} = \frac{1}{\sum_{i=1}^n \frac{U_i}{L_{10mi}}} \quad (3.19)$$

The calculated L_{10} life is compared with the actual L_{10} life in Table 3.2. As observed most of the HSS bearing failure is due to the downwind TRB axial cracking. Here NRS NRS refers to downwind TRB. The bearings whose failure root cause analysis is not available is mentioned as N.A.

As it can be inferred from the Table 3.2, the actual L_{10} life observed in axial cracking failure is around 10-25% of the expected lifetime of the bearings. Similar results were observed by Evans [59]. Therefore, it becomes necessary to look at contact stresses and lubrication parameters to understand premature failures in bearings. With improvements in bearing manufacturing and design, now a days bearing very rarely fail due to subsurface fatigue as accounted by ISO 281:2007. With good lubrication and design, the bearing could outperform the design lifetime by over three to six fold [60]. Therefore, employing Palmgren-Miner rule to calculate the bearing damage using SCADA data might be futile, if bearing surface properties and bearing kinematics is not considered.

3.6.4. Minimum Requisite Load

Every roller bearing should be subjected to a minimum required load to ensure uniform rolling of the roller elements. When this minimum load is not met due to various reasons, the roller might start skidding instead of rolling. This causes smearing in the contacting surfaces which leads to bearing damage. The minimum requisite load is more important in applications where the bearings are subjected to rapid accelerations or rapid starts and stops, which can all be seen in wind turbines.

Table 3.2: Actual vs Calculated L10 life of failed bearings.

Turbine No.	Bearing failure	Cause	Calculated L10 (hours)	Actual L10 (hours)	Actual L10 / Calculated L10 %
1	N.A	N.A	N.A	65160	-
2	RS inner ring	Spalling - Inner ring	318940	35064	10.99
3	N.A	N.A	N.A	62784	-
7	NRS NRS inner ring	Axial hairline crack.	299783	48744	16.26
10	NRS NRS inner ring	Axial hairline crack.	291651	58680	20.12
13	N.A	N.A	N.A	75144	-
15	N.A	N.A	N.A	66792	-
18	NRS NRS inner ring	Spalling - Inner ring	284974	65544	22.99
19	NRS NRS inner ring	Spalling - Inner ring	325018	36792	11.32
21	N.A	N.A	N.A	70104	-
23	NRS NRS inner ring	Axial hairline crack	357016	44520	12.47
27	NRS NRS inner ring	Axial hairline crack	-	46920	-
30	NRS NRS inner ring	Axial hairline crack	-	51648	-
36	N.A	N.A	N.A	66144	-
38	NRS NRS inner ring	Axial hairline crack	304352	36408	11.96
39	NRS NRS inner ring	Axial hairline crack	278462	59424	21.34
41	NRS NRS inner ring	Axial hairline crack	303949	56808	18.70
44	NRS NRS inner ring	Axial hairline crack	318732	52680	16.52
46	NRS NRS inner ring	Axial hairline crack	303904	33432	11.01
50	N.A	N.A	300738	75192	25.00
51	NRS NRS inner ring	Spalling - Inner ring	280296	50832	18.14
55	N.A	N.A	N.A	69312	-
57	N.A	N.A	N.A	75456	-
59	NRS NRS inner ring	Axial hairline crack	301872	51168	16.96

For a CRB, Equation 3.20 as provided by SKF is used to determine the minimum requisite load, where k_r, n_r are load and speed constants depending on the bearing size. Here, n is the bearing speed and d_m is the mean bearing diameter. For TRBs the minimum load should be around 2% of its dynamic load capacity.

$$F_{rm}^{CRB} = k_r \left(6 + \frac{4n}{n_r} \right) \left(\frac{d_m}{100} \right)^2 \quad (3.20)$$

3.7. Contact Mechanics

Bearing failures are dependent on contact stresses. Rolling element bearings are typical mechanical components that operate under concentrated-contact conditions. The Hertz theory of elastic contact deformation is used to analyze the bearing contact pressure and subsurface von mises stresses. The theory makes use of several assumptions as the exact determination of contact stresses for complicated surfaces is a difficult process -

- The contacting surfaces are smooth and frictionless.
- The contact area is small compared to the size of the bodies.
- The bodies are isotropic and elastic.

For roller bearings, it is assumed that there is a line contact between the contacting surfaces as shown in Figure 3.19. Theoretically, the the contact formed is rectangular with half-width b and length B , as shown in Figure 3.19.

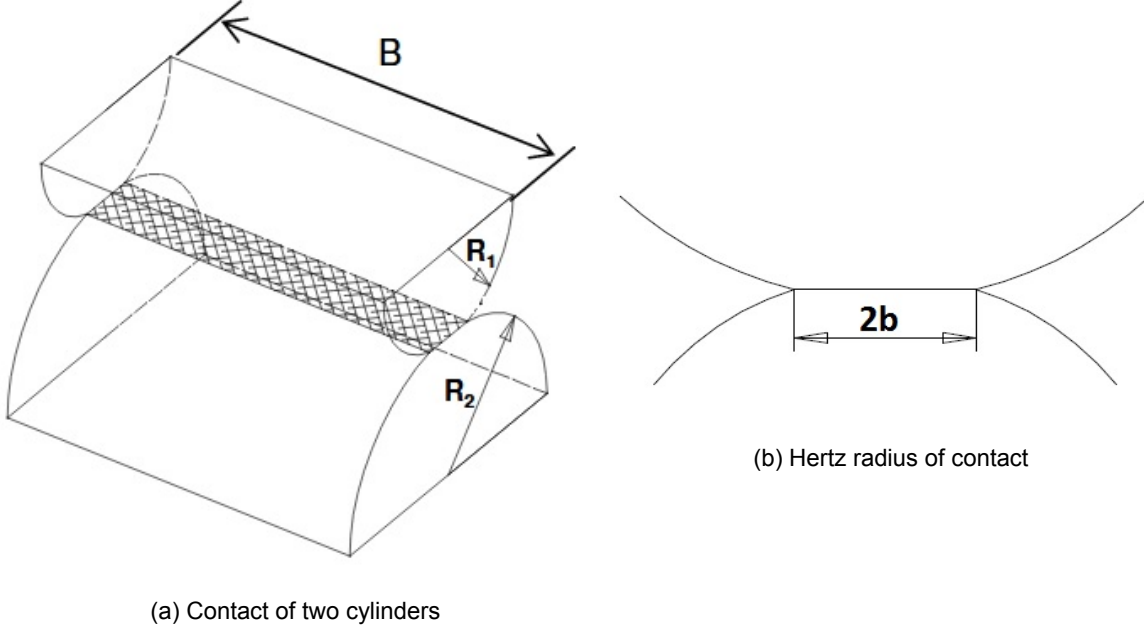


Figure 3.19: A pictorial representation of data removed from the training data set due to filtering.

The procedure to obtain the contact pressure and stresses is provided in this section. The same calculation can be used for all roller bearings (for CRBs and TRBs) [61]. Assume, R_1 and R_2 the radii of the contacting surfaces, ν_{p1} and ν_{p2} (~ 0.3) is the Poisson's ratio of materials in contact. E_1 and E_2 (~ 208 GPa) are the elastic modulus of the surfaces. F is the dynamic equivalent load acting on the surfaces. The half width b , is calculated using Equation 3.21.

$$b = \sqrt{\frac{2 \cdot F}{\pi \cdot B} \frac{(1 - \nu_{p1})^2/E_1 + (1 - \nu_{p2})^2/E_2}{1/2R_1 + 1/2R_2}} \quad (3.21)$$

The pressure distribution is rectangular and the maximum contact pressure can be determined from Equation 3.22.

$$P_{max} = \frac{2 \cdot F}{\pi \cdot b \cdot B} \quad (3.22)$$

The principal stresses on the z -axis can be determined using the following equations, where z is the distance below the contacting surfaces -

$$\sigma_x = \sigma_1 = -2\nu_p P_{max} \left[\sqrt{1 + \frac{z^2}{b^2}} - \left| \frac{z}{b} \right| \right] \quad (3.23)$$

$$\sigma_y = \sigma_2 = -P_{max} \left[\frac{1 + \frac{z^2}{b^2}}{\sqrt{1 + \frac{z^2}{b^2}}} - 2 \left| \frac{z}{b} \right| \right] \quad (3.24)$$

$$\sigma_z = \sigma_3 = \frac{-P_{max}}{\sqrt{1 + \frac{z^2}{b^2}}} \quad (3.25)$$

The 3 principal stresses are represented by the dotted lines below the surface in Figure 3.20. From the principal stresses, maximum von mises stress and the shear stress below the contacting surfaces can be calculated. According to the Hertz theory, maximum von mises stress occurs at a distance of $0.786 \times b$ below the contacting surfaces [61].

Von mises stress or equivalent stress is a part of equivalent stress failure theory, used to predict yielding in a ductile material. There is some microplasticity or small plastic deformations in a material whose yield strength is exceeded by the von mises stress. The von mises stress or the equivalent stress can be calculated from the principal stresses using Equation 3.26. Using all the above equations, it is estimated that the sudden application of brakes/emergency braking when the turbine is rotating at rated power, can lead to von mises stresses greater than 250-300 GPa on high speed bearings. This can lead to some plastic deformation in bearings if they are not in the hydrodynamic lubrication regime, which is discussed further in Section 3.8.

$$\sigma_v = \sqrt{\frac{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2}{2}} \quad (3.26)$$

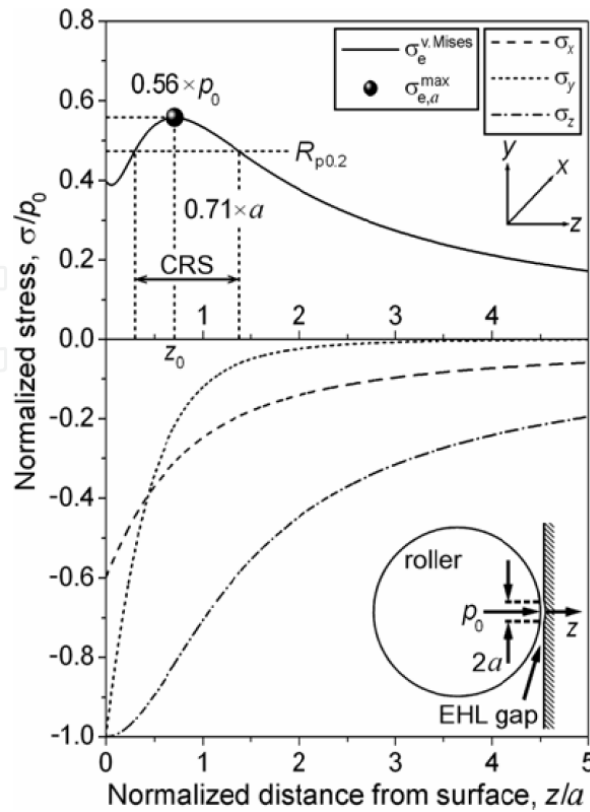


Figure 3.20: Normalized plot of the depth distribution of the σ_x , σ_y , and σ_z main normal and of the v. Mises equivalent stress below the center line of the Hertzian contact area [62].

The variation of maximum contact pressure with von mises stress and applied load is shown in Figures 3.21 and 3.22.

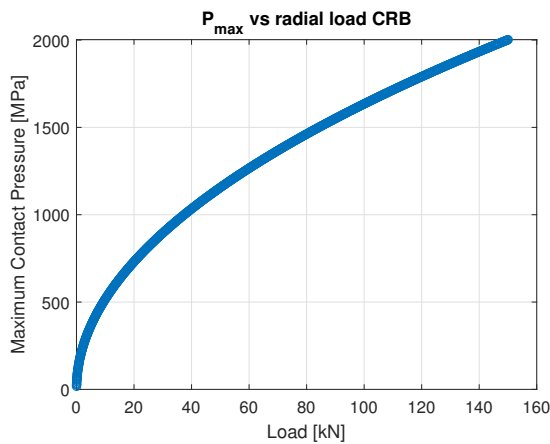


Figure 3.21: Plot of maximum contact pressure versus radial load on CRB.

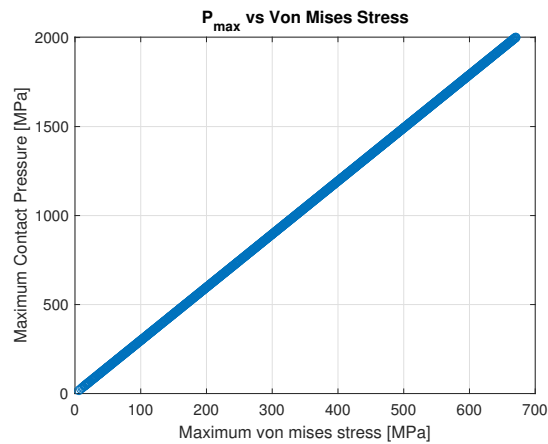


Figure 3.22: Plot of maximum contact pressure versus maximum von mises stress on CRB.

3.8. Contact Lubrication

Rolling bearing elements of wind turbines are subjected to high contact pressures and rolling velocities as mentioned in section 3.7, which could lead to scuffing failures if the steel surfaces were to come in full metal to metal contact. Hence, lubrication is recognized as a key parameter to increase the efficiency and durability of bearings. In the present section, the different lubrication regimes and tribo-chemical aspects will be briefly presented as they greatly influence axial cracking or *White Etching Cracks* failure modes.

3.8.1. Elastohydrodynamic Lubrication (EHL) Theory

Elastohydrodynamic lubrication (EHL) is the most widely used and accepted model for assessing the lubrication for friction pairs having elastic contacts like gears, bearings, and so on. In EHL theory, the contacting surface undergo elastic deformation and lubricant viscosity changes with contact pressure.

One of the main criterion for evaluating the lubrication regime is lubricant thickness between the two load transmitting contacting surfaces. Figure 3.24 gives the list of formulas to determine the minimum and the central film thickness.

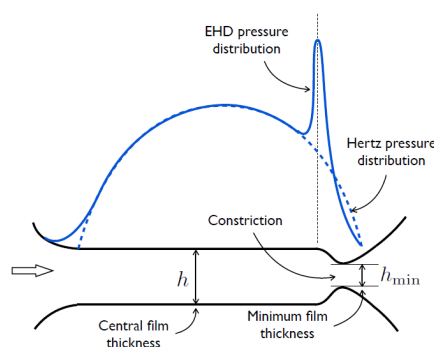


Figure 3.23: Pressure distribution and film-thickness variation in an EHD contact.

As it can be observed in Figure 3.23, there is a peak in the EHD pressure when the lubricant film thickness is at its minimum. At the exit of the roller from the load zone, due to sudden pressure drop and constant mass flow, a peak in EHD pressure is observed. In

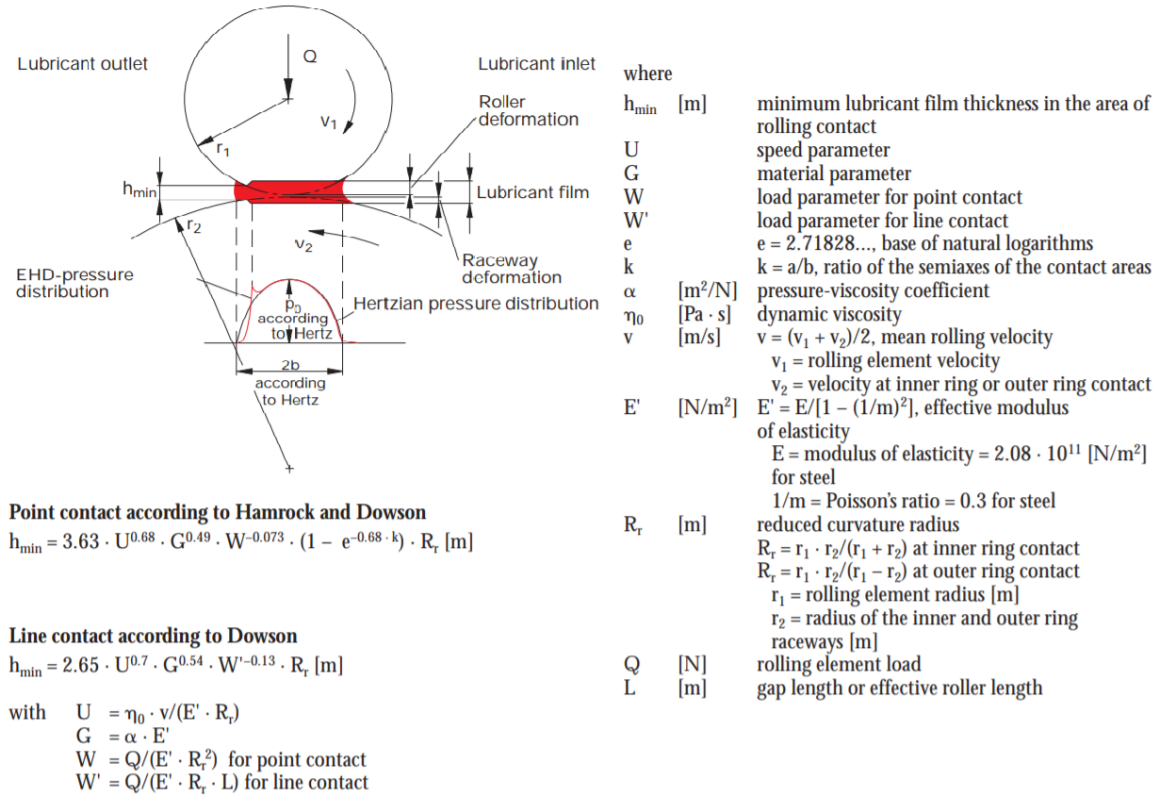


Figure 3.24: Lubricant film thicknesses for point contact and line contact. Adapted from [63].

Figure 3.24, there are few formulas which are not explicitly defined. Figure 3.24 represents Dowson EHL lubrication theory [64]. Equations from various sources used for the missing terms are included -

- The pressure-viscosity coefficient is a function of kinematic viscosity (ν) of the lubricant. According to ISO/TR 1281-2, 2008 the pressure viscosity (α) equation in (cm²/s) can be determined using Equation 3.27 [65].

$$\alpha = 0.1122 \cdot \left(\frac{\nu}{10^4} \right)^{0.163} \quad (3.27)$$

- Dynamic viscosity (μ) can be determined using kinematic viscosity (ν) and density (ρ) of the lubricant as shown in Equation 3.28.

$$\mu = \rho \cdot \nu \quad (3.28)$$

Apart from minimum lubricant thickness, the roughness of the contacting surfaces also play a crucial role in predicting the lubrication regime a roller bearing is operating in. As the operators do not the access to the roughness parameters of the bearings, an empirical formula, equation 3.29 developed by Baalman [66] is used to estimate the RMS roughness value for composite bearings based on the diameter of the bearings.

$$R_q = 1.2 \cdot 10^{-5} \cdot D_{pw}^{0.55} \quad (3.29)$$

Lubricant Lambda ratio: In the EHD traction calculations described above, it is assumed that the contacting surfaces are smooth and a lubricant film always separates the contacting surfaces. But in reality, there are always some asperities present on the contacting surfaces.

Unlike smooth Hertzian contact, the contact area is now a myriad of micro-contacts of highly irregular shapes within the macro-contact. The contact pressure at these micro-contacts can be high enough to cause localized property changes both in contact bodies and in lubricant. Lubricant lambda ratio, λ is generally used to establish the lubrication regime between the contacting surfaces. It is defined as the ratio between the minimum film thickness to the square root of the roughness RMS of the contacting surfaces, as shown in Equation 3.30. Here R_{q1} and R_{q2} are the RMS roughness values of the roller and the raceway respectively. The estimated values of RMS roughness from Equation 3.29 are $0.19 \mu\text{m}$ and $0.21 \mu\text{m}$.

$$\lambda = \frac{h_{min}}{\sqrt{R_{q1}^2 + R_{q2}^2}} \quad (3.30)$$

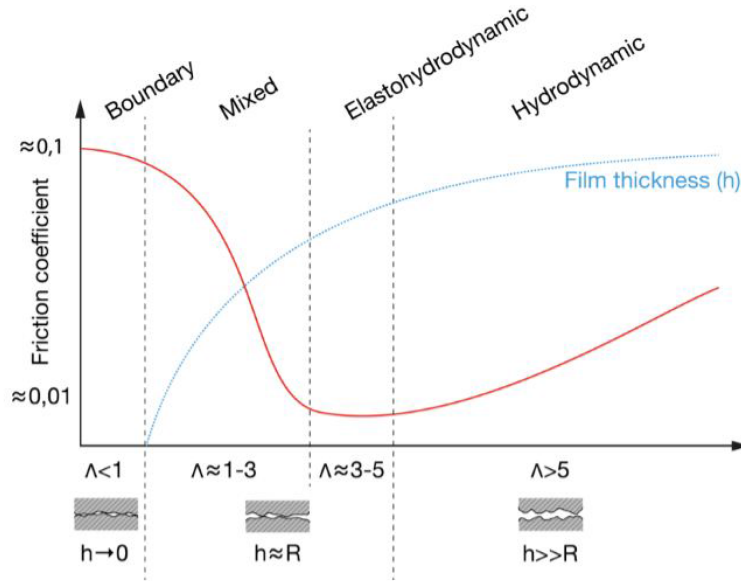


Figure 3.25: Typical Stribeck curve of a lubricant under constant load, temperature and slide-to-roll ratio and its correlation to the film thickness curve under the same operating conditions. Adapted from [67].

The lubrication regime can be classified into four categories based on the lambda values [64].

- **Boundary Lubrication ($\lambda \leq 1$):** Boundary lubrication is the most severe regime characterized by high friction and wear. All the load is supported by the asperities. Full metal-metal contact occurs in this regime. Usually, low viscosities, low velocities and/or high contact pressures lead to boundary lubrication regime. Usually, during turbine start-ups, due to low temperatures and speed, lubricant might operate in the boundary regime. It can also occur during sudden turbine shut-downs and idling.
- **Mixed Lubrication ($1 \leq \lambda \leq 3$):** A transition between the elastohydrodynamic and boundary regime occurs through mixed lubrication regime. In this regime, the load is supported partially by the lubricant film and partially by the asperities, leading to surface wear and damage. The coefficient of friction is drastically reduced. The film thickness increases with a decrease in lubricant viscosity. This regime might occur when the turbines are idling at low speeds.
- **Elastohydrodynamic Lubrication ($3 \leq \lambda \leq 5$):** The coefficient of friction is the lowest in this regime. This is achieved with full separation of the surfaces, combined with the low amount of fluid subjected to shearing. There is no asperity contacts observed until a λ value of around 4, and this regime is a particular case of highly loaded non conforming contacts where the deformation/strain of the contacting surfaces are greater than film thickness, leading to full separation of the contacts.

- **Hydrodynamic Lubrication ($\lambda \geq 5$):** As the lubricant viscosity increases due to higher rolling speeds, a thick film is formed. In this regime, there is no contact between the asperities. The coefficient of friction increases due to higher lubrication shearing only, as the film thickness and viscosity increases.

In Figure 3.26, a plot of lambda ratio versus the generator rpm for a year of SCADA data is shown. It can be inferred from the plot that at lower rpm, i.e., lesser than 150 rpm the value of λ is lower than 4. The rpm value is usually low when the turbine is idling due to low wind speeds or during shutdowns. At this stage, the turbine bearings, especially the HSS bearings can undergo wear and damage. At higher speeds, the lubrication film is fully developed and operated in the hydrodynamic regime. A histogram of λ is shown in Figure 3.27, to give an estimation of the amount of time in an average year the bearing operates in the boundary or mixed lubrication regimes.

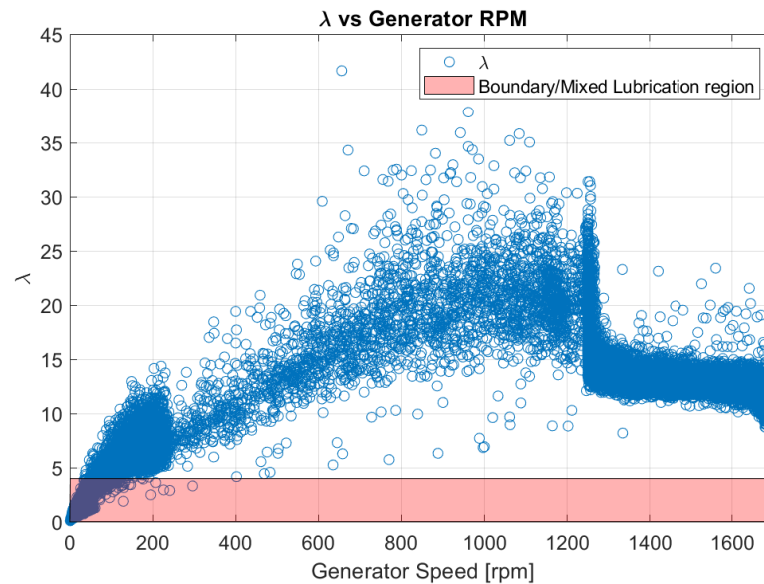


Figure 3.26: Scatter plot of λ vs Generator RPM of CRB for one year SCADA data.

3.8.2. Contact Friction

Bearing friction is not always constant. It depends on various parameters like - lubricant thickness, bearing loads, speed. The SKF model of calculating frictional moment is used to calculate the lubrication regime, to confirm if the lambda ratio calculated in section 3.8.1 is valid. A detailed explanation of the equations and theory to calculate frictional moment can be found in [68]. Figure 3.28 represents the change in frictional moment as a function of speed, for a given bearing and lubricant. Figure 3.29 shows a scatter plot of frictional moment versus generator speed of CRB. It can be conclusively seen that at lower speeds, the frictional moment is high which is due to high asperity contacts. Usually the coefficient of friction during normal operation of roller bearings is around 0.0010-0.0030. At lower rpm (less than 100 rpm) coefficient of friction in the HSS bearings can rise upto around 0.1 - 0.3.

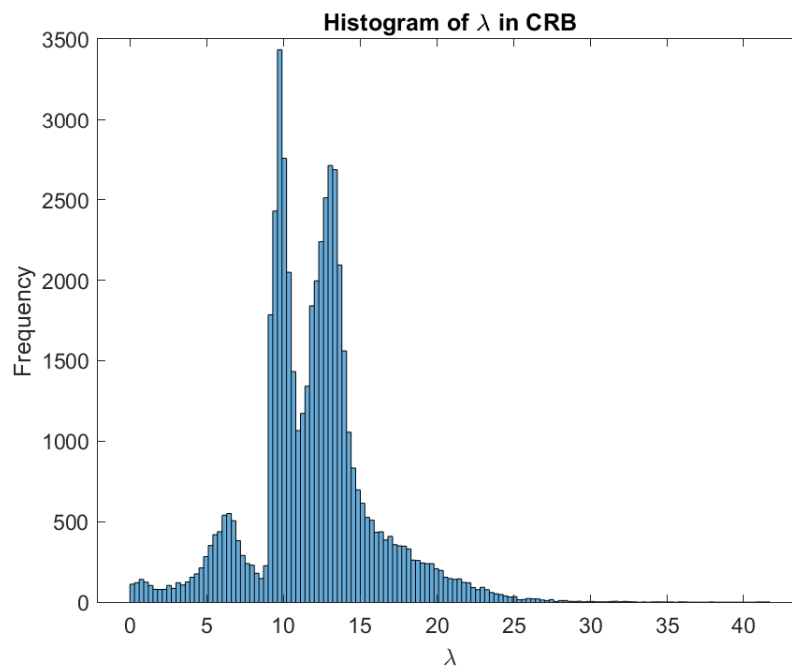


Figure 3.27: Histogram of λ in CRB for one year of SCADA data.

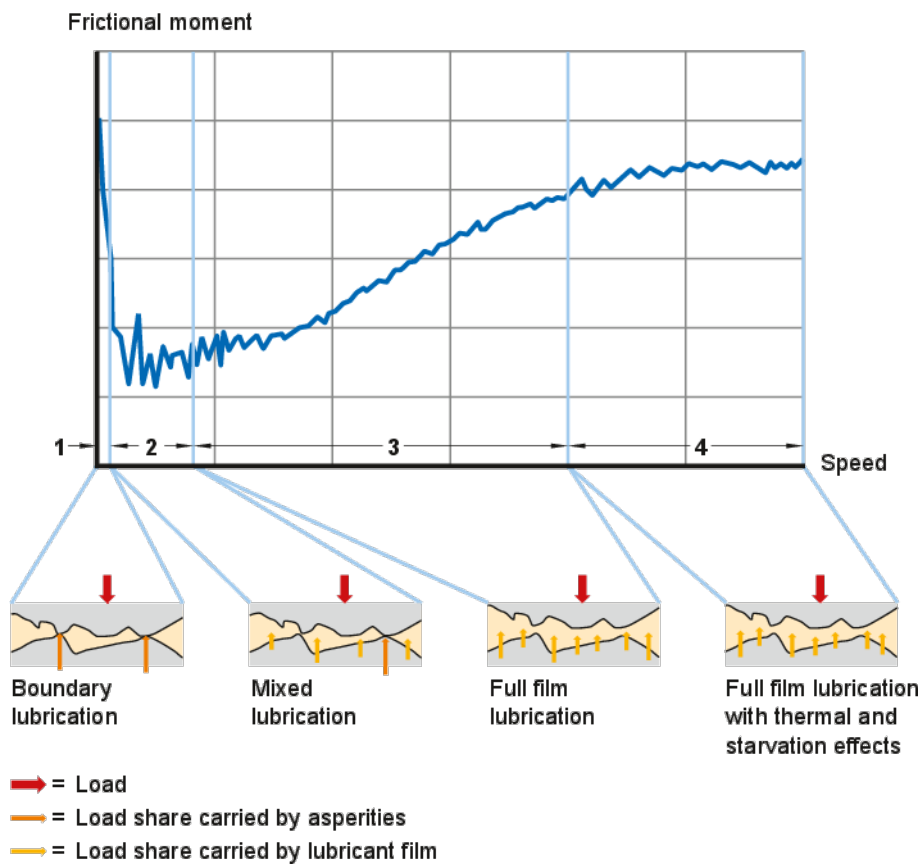


Figure 3.28: Bearing frictional moment as a function of speed. Adapted from [68].

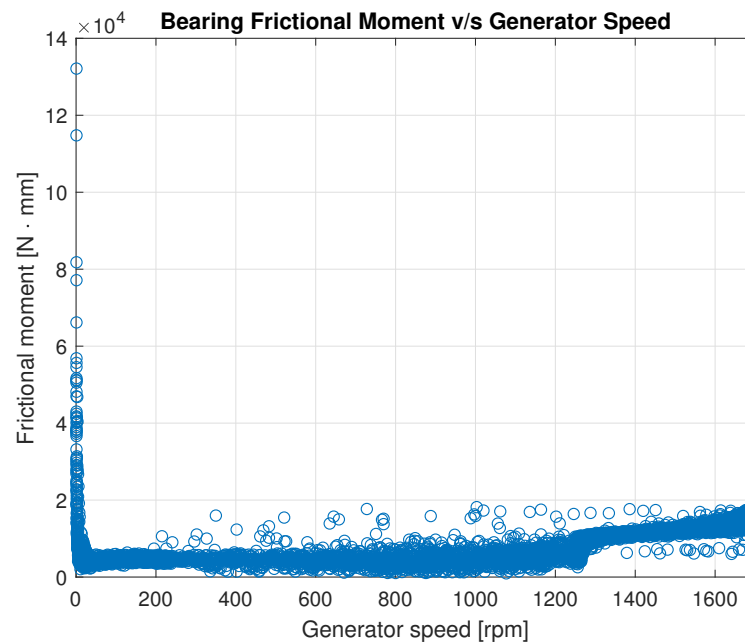


Figure 3.29: Bearing frictional moment as a function of speed using SCADA data.

3.9. White Etching Cracks (WECs)

White Etching cracks or axial cracks are leading cause of failures in wind turbine gearboxes, which are prominently seen in the HSS and IMS bearings. They are non-classical and non-conventional failures. As seen in Table 3.2, most of the reported failures in the HSS bearings in the wind farm of reference is due to axial cracking and spalling. It reduces the calculated L10 life of bearing by almost 70-80%. It eventually leads to irreversible brittle flaking, spalling, or radial cracking of the affected bearing component. White etching refers to the appearance of the altered steel microstructure when polishing and etching a microsection. In Figure 3.31, a representation of the axial/white etching cracks is shown. WECs are known to occur occasionally in industrial bearings, marine propulsion systems but occurs more frequently in wind turbines. This suggests that the stochastic and extreme conditions under which the wind turbines operate have a direct relation to WECs. Although focused research by researchers, wind turbine OEM's & bearing manufacturers is being conducted for the last two decades, there is no consistent evidence or theory to explain the root cause of WECs. Based on some of the hypothesis proposed, as mentioned in Section 3.9.1, an attempt is made to correlate the observed field failures due to WECs to their respective operating conditions with respect to the rest of the fleet.

3.9.1. Potential Root Causes of WEC

In this section, the potential root causes of WECs or axial cracks from various sources are summarized -

- Purely Loads/Stress factors:** Brief periods of heavy loads can cause microplasticity in the subsurface and near subsurface regions leading to the formation of white etching areas. In an operational wind turbine, such stresses can arise from overload events like emergency braking, grid connection losses, which can lead to contact stresses of above 3.5 GPa, even with good lubrication conditions [71]. Lai et al., [71] set up an experiment to observe WECs. First, the bearing was subjected to short heavy loads and later subjected to moderate loads of about 1.5 [GPa] at moderate-good lubrication condition. After 1.9×10^7 cycles 2 of the 4 bearings had shown WECs.



Figure 3.30: Failure appearance: a) straight cracks; b) straight cracks and small spalls; and c) spalls. [69]

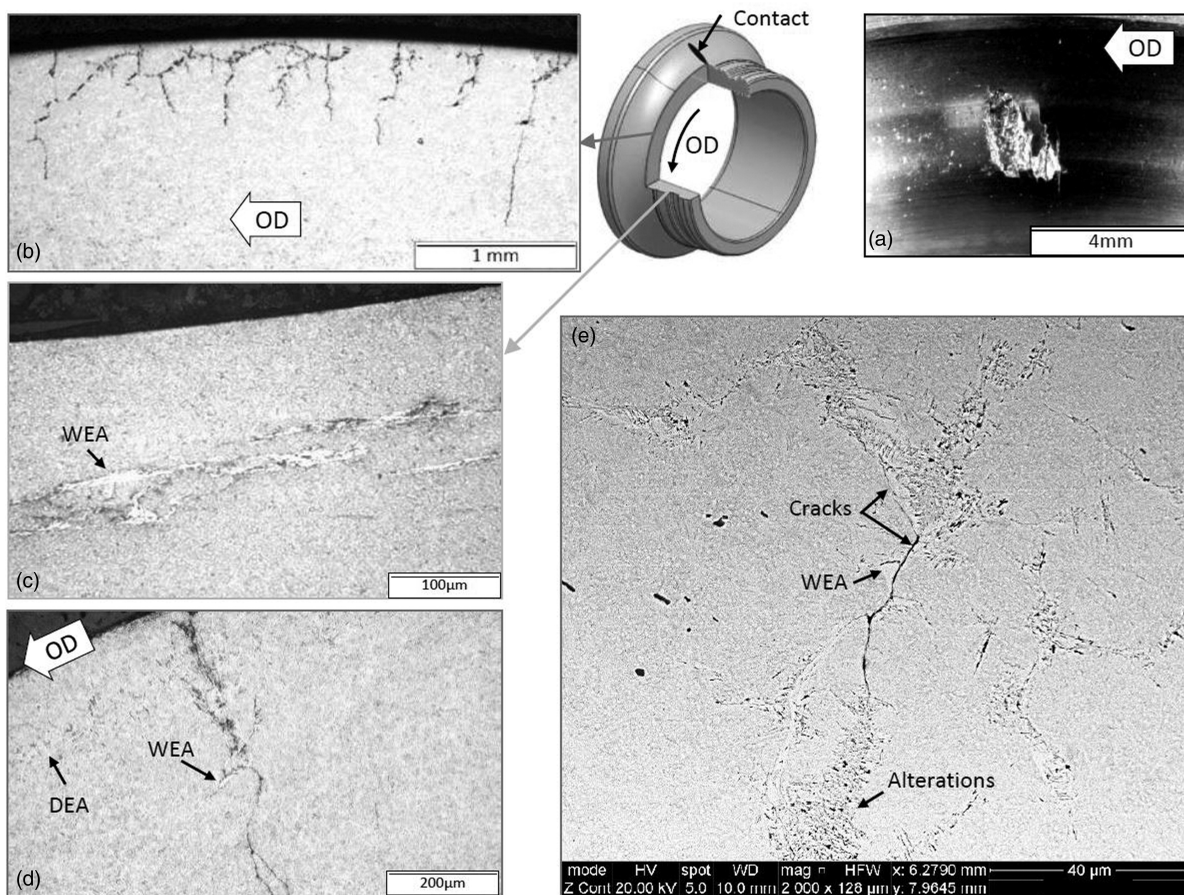


Figure 3.31: (a) Raceway top view of a typical WEC-associated spall; (b) circumferential LOM with discrete WEC networks displaying vertical links to the surface and a step-wise propagation; (c) axial LOM with WECs laying parallel to the surface in accordance with the respective steps in (b); (d) circumferential LOM closer to the DEA region; (e) SEM analysis revealing ultra-thin cracks, adjacent refined microstructure and darker alterations. LOM cross sections are all Nital 2% etched [70].

- Lubrication issues:** According to SKF research [51], most of the axial cracks are initiated on the surface or near subsurface (0-150 μm) and propagate through the surface due to low lubrication conditions and is accelerated by corrosion fatigue, leading to high contact pressures. Bearings operating at moderate loads and in the boundary or mixed lubrication regime showed signs of axial cracking [62]. Some evidence has also been found that any disturbance in the bearing kinematics due to high vibrations during standstill position can lead to an acceleration in the corrosion fatigue process [69].

Figure 3.32 shows a schematic of the axial crack initiation due to mixed lubrication conditions.

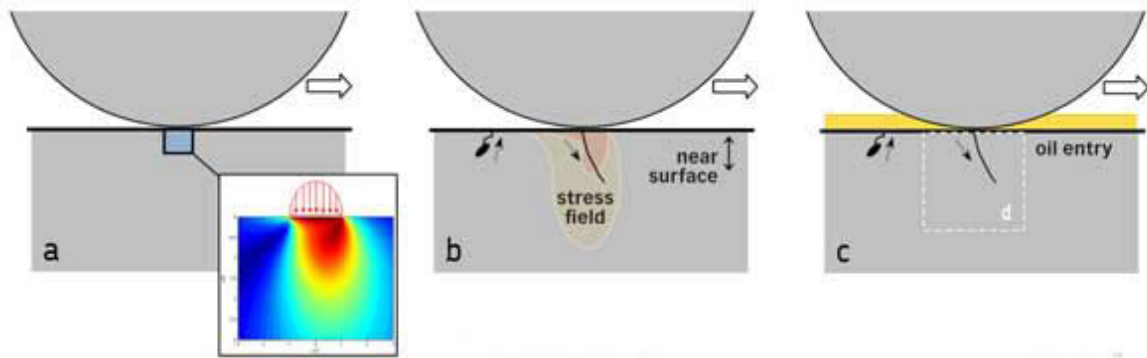


Figure 3.32: (A) Roller-raceway contact with areas of local high traction due, for example, to local mixed friction, leading to tensile stresses that can, b) lead to damage such as a small crack; c) surface crack or crack connected to the raceways allows the entry of oil. Adapted from [69].

- Roller skidding/sliding:** Recently, a lot of evidence has been produced to suggest that WECs develops due to roller skidding during high speed transient events, and also skidding because the minimum required load for rolling is not achieved [53]. During some transient loading like gusts, brake applications, grid connection losses along with high loads, there are torque reversals occurring where the load application changes instantly from loaded zones to unloaded zones and occurs until it is dissipated, as shown in Figure 3.33. This causes skidding due to a loss in the contact pressure and low loading. A representation of the torque and speed changes in the high-speed shaft obtained through experimentation is shown in Figure 3.34. Higher chances of skidding can be expected in a grid loss event where the torque is reduced by almost 60%, while the speed does not reduce much. Another important parameter to assess roller sliding is the slide-roll ratio (SRR), which needs specialized instruments to determine, hence ignored in this thesis.

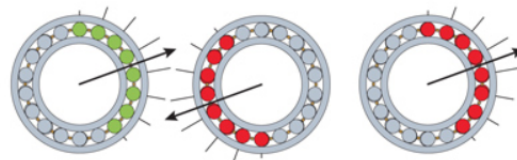


Figure 3.33: Schematic representation of rapidly evolving torque reversals.

- Non-metallic Inclusions:** Various experiments have observed formation of WECs in bearings with certain non-metallic inclusions like MnS (Manganese Sulphide), large carbides, voids, so on which act like stress concentration regions. Usually, the crack propagation rate is enhanced when these inclusions are below the surface where the maximum shear stress can occur [72],[73]. Figure 3.35 explains the importance of sliding and inclusions for WEC formation and propagation.
- Low material strength:** There are few theories which suggest that hydrogen ingress promotes formation of subsurface cracks leading to WECs. In the presence of hydrogen and high contact pressure, the crack propagation is rapid which results in early bearing failure [75].

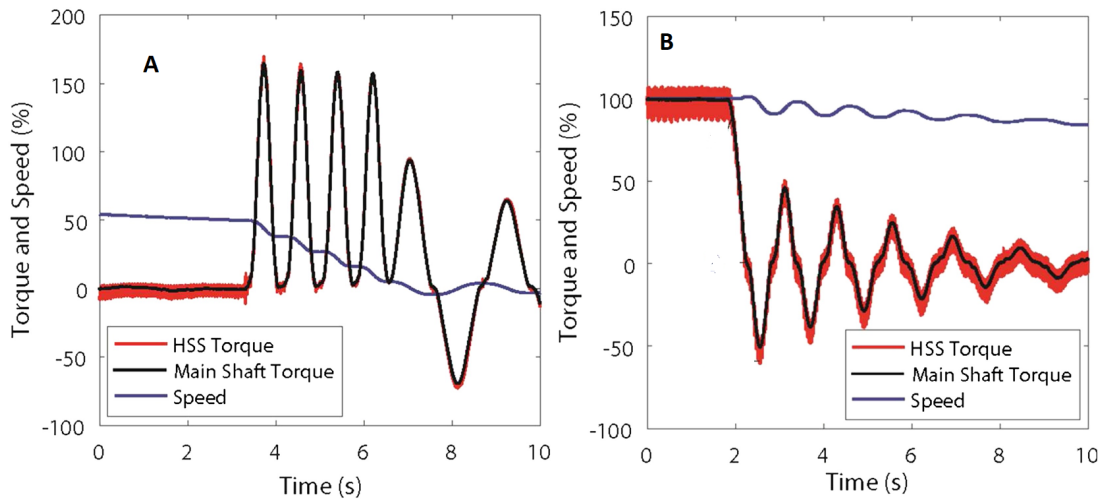


Figure 3.34: Dynamometer test results on HSS (A) during braking events (B) during grid loss events [53]



Figure 3.35: Importance of surface traction (not to scale): (a) pure rolling: critical stress threshold exceeded relatively deep—micro-cracking at inclusion may never reach surface and cause failure; (b) rolling and sliding: critical stress field shifted closer to surface—micro-cracking at inclusion propagates to surface, leading to failure [74]

Figure 3.36, is a schematic representation of the conditions that lead to WECs as explained above. The WECs driving factor may be only one of the above mentioned factors, or due to a combination of factors. To make it more complex, many factors are correlated. This makes it difficult to quantify the damage.

3.10. Case Study II - HSS Bearing Failure

Based on all the above discussions in the chapter, it is decided to include the following operating conditions that could potentially induce damage on wind turbine HSS bearings -

- Low lambda - Total number of hours when the lubricant lambda value is less than 4.
- Low load - Total number of hours when load on bearing is less than the minimum requisite.
- High speed brake application - Event count of brake application at high wind speeds and rated power, excluding emergency braking.

Some of the damage inducing factors like hydrogen charging, electric current discharge, non-metallic inclusions cannot be assessed using 10-minute averaged SCADA data. Also, it is challenging to determine the exact number of grid outages from the SCADA logs, hence number of grid loss events is not included as one of the damage operating condition in this

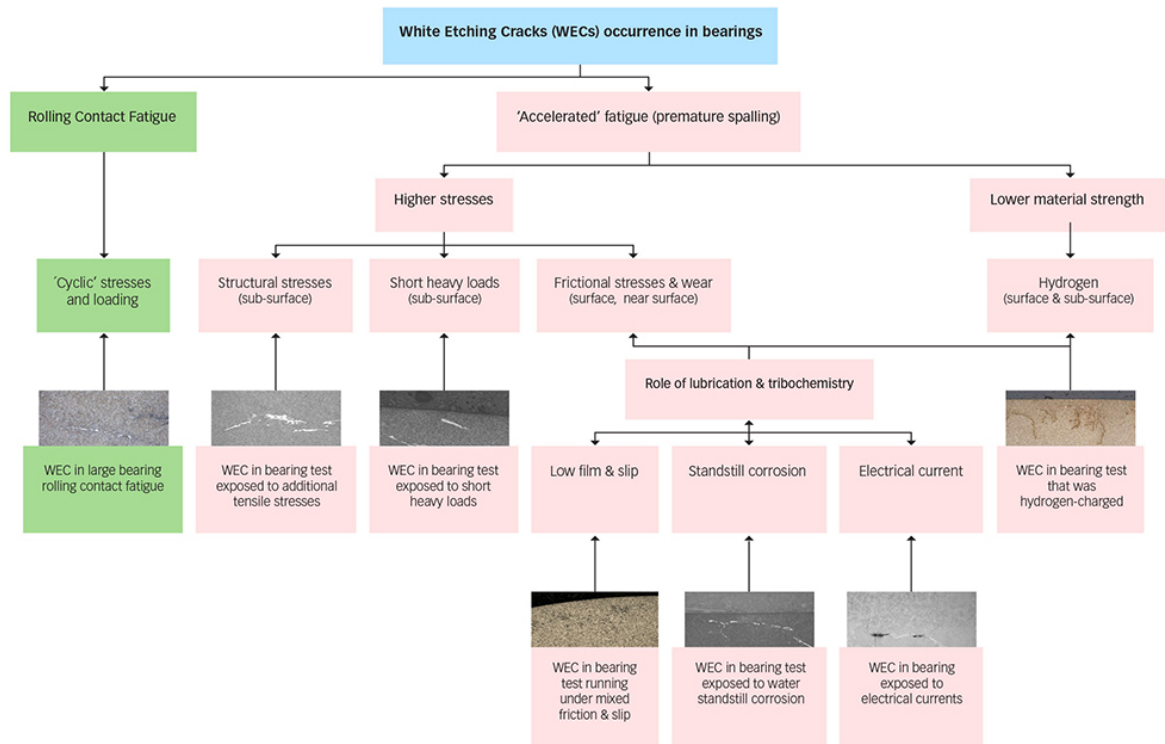


Figure 3.36: White etching crack occurrence (simplified) in rolling element bearings. [69]

work. Also previously assumed conditions namely, *bearing temperature and yaw movement* is excluded in further studies, as no correlation is found between gearbox and these conditions.

A comparison of the failed bearings with the rest of the fleet is done using boxplots, as their advantages in assessing the risk of failure was shown in Section 3.4.

3.10.1. Turbine 2 - HSS CRB axial cracking

The CRB of turbine 2 failed due to axial cracking which manifested into spalling in the year 2012. Turbine 2 is compared against 60 other turbines in the wind farm using boxplots as mentioned earlier in the chapter. Such boxplots can also help in root cause analysis of the failure.

The red points on the boxplot in Figure 3.37 represents turbine - 2 with respect to the rest of the fleet. As it can be seen from the below figure, the failed CRB had to endure lower load operation which in turn leads to low lambda ratio for longer periods of time which could lead to skidding and eventual failure. The number of rotor start - stops is also above the 75th percentile mark.

3.10.2. Turbine 8 - No Failure

There was no damage reported in turbine 8 until 2019. A case of no failure, yet high risk according to box plots is included in Figure 3.38. This plot shows the limitation of physics of failure technique while using low frequency data. As we use basic statistical plots to check whether the turbines are at a high or low risk of failure, there tends to be more than one turbine operating at high risk. This is the case as observed with Turbine 8. Even though the turbine is operating at low load lambda and a higher number of emergency stops as compared to the mean to the fleet, no apparent failure occurred. Hence, it is necessary to have additional monitoring techniques, and use physics of failure model to complement them so as to make the condition monitoring system more robust.

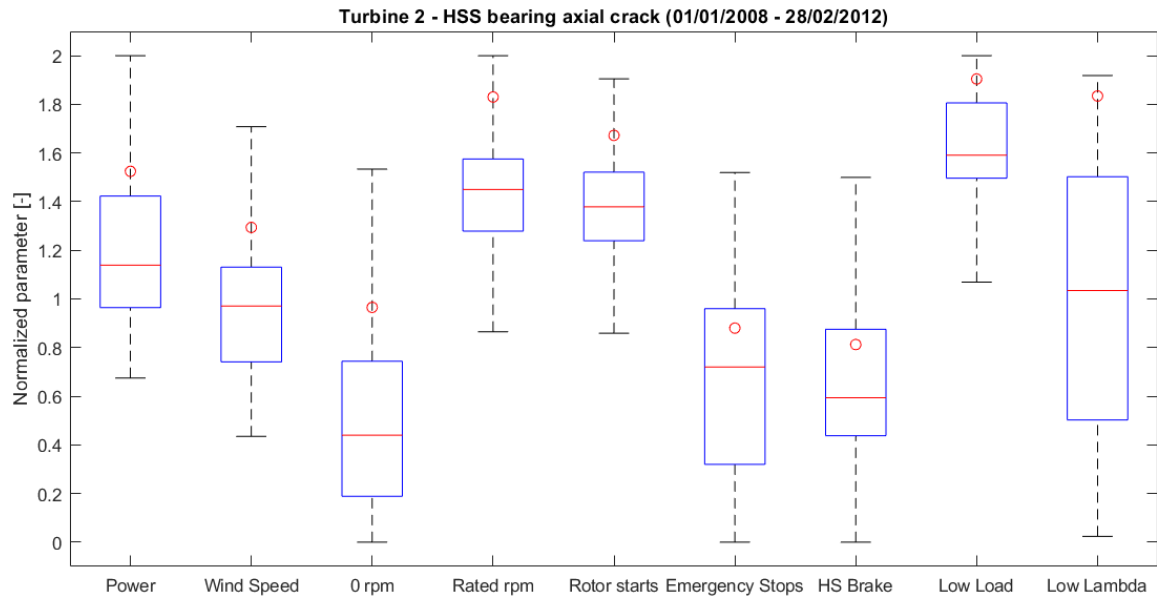


Figure 3.37: Turbine 2 - HSS bearing axial cracking

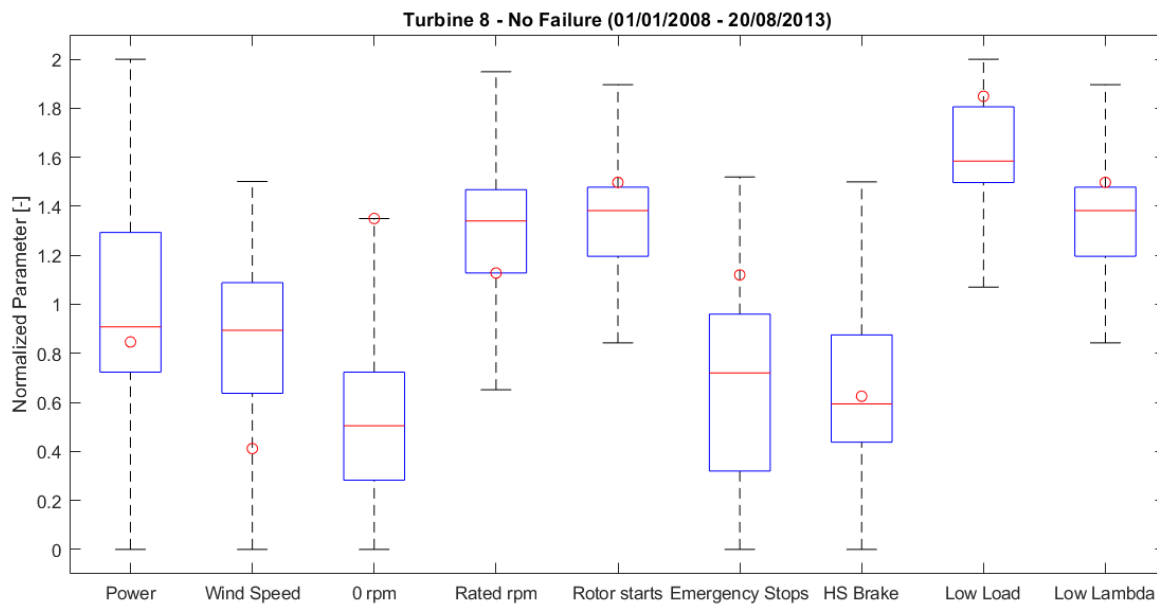


Figure 3.38: Turbine 8 - No failure

3.10.3. Turbine 14 - No Failure

Turbine 14 is a case of low risk turbine which has not had any components of the gearbox replaced at least until 2014. As seen in Figure 3.39, all the points in the boxplots lie below the 75th percentile, indicating a low risk of failure.

More case studies relating to physics of failure of axial cracking is presented in Chapter 5 along with the temperature based anomaly detection by ANN model.

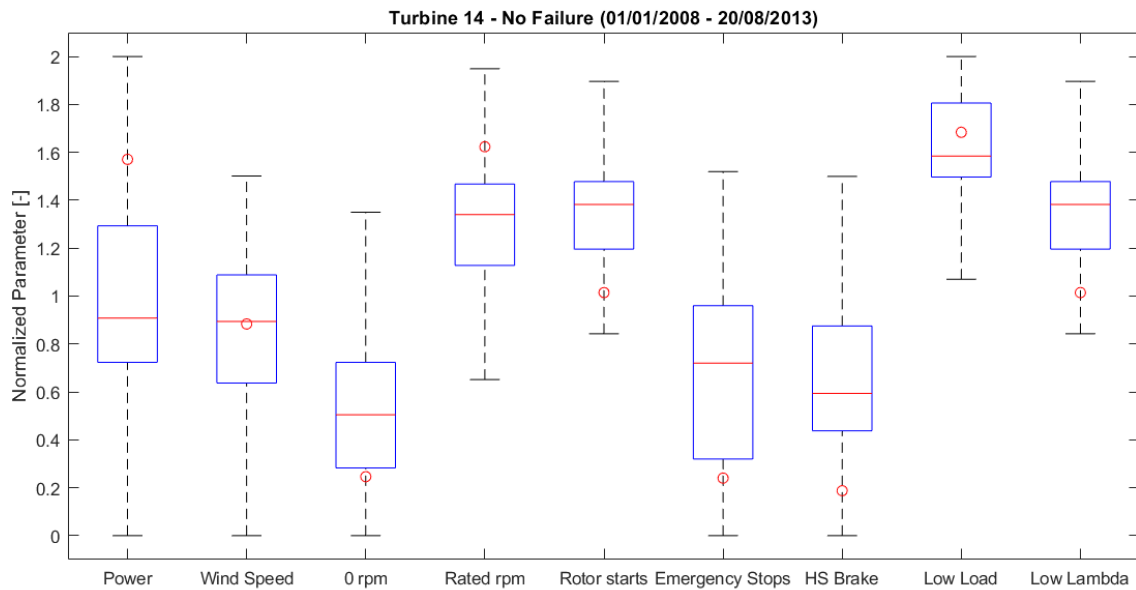


Figure 3.39: Turbine 14 - No failure

3.11. Limitations

There are certain limitations of the physics of failure models.

- **Low frequency data** - The SCADA data is recorded at 10 minute resolution. It is very challenging to build a high fidelity physics based model using 10-minutes averaged data, as start-stop, transient events can happen within few seconds which is not captured in the low resolution data.
- **Risk assessment** - Boxplots are good at assessing risks/outliers. It is clear that there lies more than one turbine above the 75th percentile. It is important to note here that, *all the turbines which failed were operating at high risk, while not all turbines operating at high risk failed*. Hence, this approach can be used in complement with vibration monitoring/ thermal monitoring systems to improve the confidence in condition monitoring of wind turbines, as physics-based models take into account the historic operating conditions which are ignored by other monitoring techniques. This method is best suited for risk assessment of the entire fleet.

To overcome the limitations of the physics of failure models and to complement them, a temperature based condition monitoring technique based on artificial neural networks is also developed.

Temperature based Anomaly Detection using Artificial Neural Networks

This chapter describes the methodology of using temperature signals from SCADA for anomaly detection in wind turbine gearbox. Various issues with the ANN models are discussed and suitable mitigation measures are presented. The RNN model architecture and hyperparameter tuning are also discussed in the chapter.

4.1. Normal Behavior Modeling

The effectiveness of normal behavior models (NBM) for the purpose of anomaly detection has been demonstrated in [36], [38], and [39]. In this thesis, a NBM based on Gated Recurrent Unit (GRU) is proposed and compared with NARX based model as described in [39]. The schematic representation of the method is shown in figure 4.1.

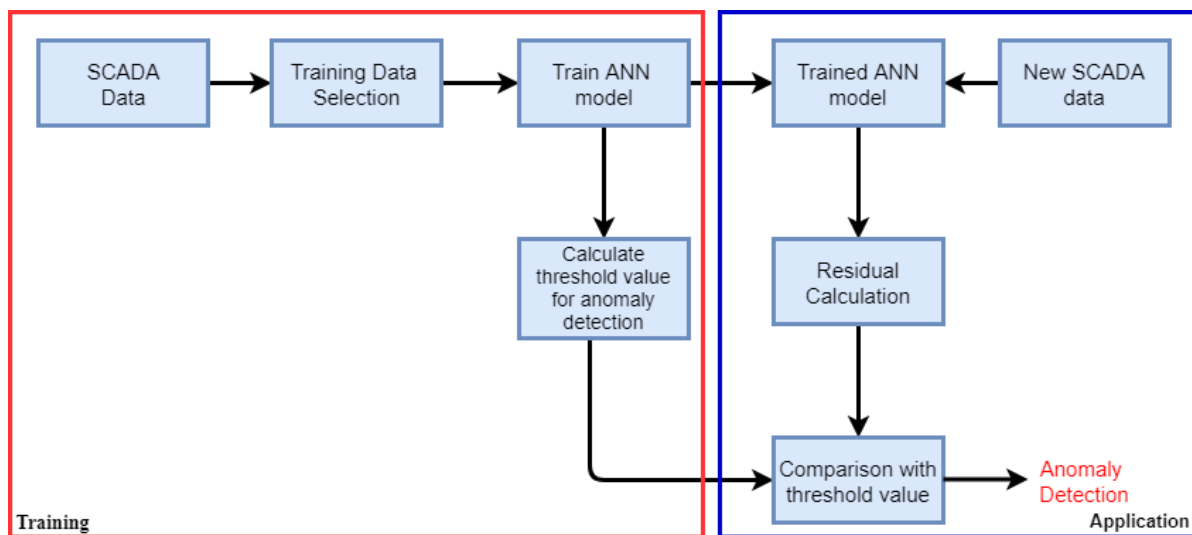


Figure 4.1: A schematic representation of an ANN based condition monitoring method. Adapted from [39]

The development of NBM can be divided into two parts. As seen in the left block of Figure 4.1, the ANN model trained using only data consisting of normal operation (fault free period) of the wind turbine. This is a one time process. The trained model is then used to predict the modeled parameter values, which is then compared with the actual values recorded in the

SCADA. This is shown in the right-hand block of Figure 4.1. If the residual value, that is the difference between the predicted and measured value crosses a predetermined threshold, an anomaly is flagged.

The following steps are important for development of normal behavior models which are further detailed in the ensuing sections -

1. Input and output parameter selection.
2. SCADA data preprocessing.
3. ANN configuration/Model Architecture.

4.2. Input and Output Parameter Selection

The selection of input and output parameters is critical for the performance of ANN models for condition monitoring. A typical SCADA system records more than 100 variables and it becomes essential to select only a few variables for modeling to avoid overfitting/underfitting. Overfitting can lead to false alarms and inaccurate anomaly detection. Since ANN modeling is a data-driven approach, which lacks any physical understanding of the system, the selection of model parameters has to be done carefully based on domain knowledge and correlation analysis.

The selection of output parameters for condition monitoring of gearbox is straightforward, as there are only two temperature-based parameters which enable failure detection. Potential component failures should manifest themselves in the chosen measurement, to enable failure detection-

- Gear bearing temperature.
- Gearbox oil temperature.

The selection of input parameters is more complex. Therefore, a combination of correlation analysis between various variables and domain knowledge is used to select suitable input parameters which help in accurate estimation of output parameters. In contrast to the target or output parameter selection, there is usually a big number of potential input measurements to choose from.

A correlation matrix is shown in Figure 4.2. 2 years of wind turbine SCADA data is used to analyze the correlation between various variables. This figure shows the correlation coefficients along with the histogram of the respective parameters. As mentioned earlier care must be taken while selecting model input parameters using such correlation coefficients. Selecting all the input parameters based on statistics only, with high correlation coefficients can lead to overfitting, and can lead to problems during the anomaly detection phase. Hence, it is essential to consider input parameters based on domain knowledge and physical understanding of wind as well.

The ambient and nacelle temperature are directly related to the gear bearing temperature and the lubrication oil temperature as all the related sensors are placed inside the nacelle. According to the first law of thermodynamics, there exists a thermal equilibrium between the above-mentioned temperatures under normal operating conditions[13]. Hence, nacelle and ambient temperatures are chosen as one of the inputs to the ANN model. Apart from these, power and rotor RPM is chosen as the other input parameters to the ANN model as their correlation coefficient with respect to the output parameters is high. Also, power and rotor rpm is directly proportional to the chosen output parameters. Higher output power leads to higher bearing temperatures and vice versa.

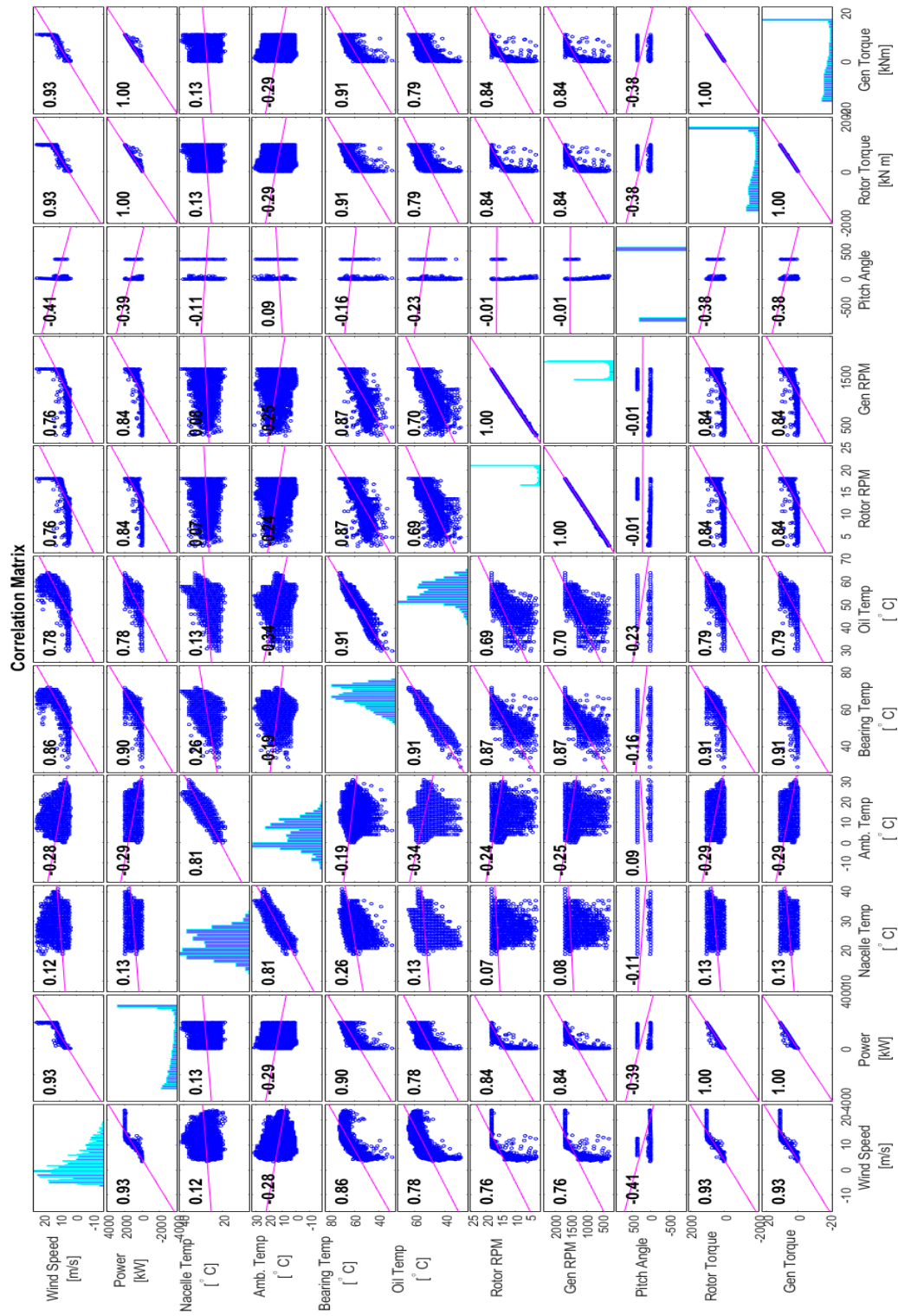


Figure 4.2: Correlation Matrix of SCADA variables.

To test the hypothesis that additional calculated physical parameters from the recorded parameters can lead to better performance of the ANN model as stated in [19], generator torque is used as an additional input parameter to the ANN model. The correlation coefficient between generator torque, gear bearing temperature and gearbox oil temperature is also high. The temperature sensors are placed on the bearing in the high-speed shaft which is coupled to the generator. Therefore, abnormal changes in the generator torque and the chosen output parameters can lead to better anomaly detection.

Two sets of inputs and output parameters as tabulated in Table 4.1 similar to the one used in [39] and Table 4.2 is compared using NARX and RNN based ANN models.

Table 4.1: Input/Output parameters for ANN [Set-1]

Output Parameters	Input Parameters
Gear Bearing Temperature	Power
	Rotor RPM
Gearbox Oil Temperature	Ambient Temperature
	Nacelle Temperature

Table 4.2: Input/Output parameters for ANN [Set-2]

Output Parameters	Input Parameters
Gear Bearing Temperature	Power
	Rotor RPM
Gearbox Oil Temperature	Ambient Temperature
	Nacelle Temperature
	Generator Torque

4.3. Data Preprocessing

Data preprocessing is a data mining technique used to convert raw data into clean data that can be used for machine learning projects. The recorded raw SCADA data is not always clean as it can have some missing values due to communication errors, can be noisy i.e., containing errors or outliers and so on. The SCADA data thus has to be preprocessed to remove all the faulty data before using it.

The phrase "garbage in, garbage out" is particularly applicable for data-driven analysis as the performance of the machine learning model depends purely on the data used to train it. Data preprocessing becomes more essential while developing a normal behavior model as only data during the fault-free period of the turbine has to be filtered out for training the ANN model which is a major challenge.

Three filters similar to the one implemented in [39] are used with minor changes. The applied data filters are robust and have been successfully implemented before. All the filters are applied sequentially as explained. The purpose and the algorithms for the filters are detailed in the following sections.

1. General filter.
2. Cluster filter.
3. Missing data filter.
4. Masking/padding for RNN models.

4.3.1. General Filter

Data vectors with missing datapoints considering both the input/output parameters are filtered out in this step. This filter is applied to both training and application sets. In equation 4.1, x_t is the data vector at time t where one of the datapoints is missing in dataset x_{ti}

$$\text{delete } x_t \text{ if } [x_t | x_{ti} = \text{not available}] \quad (4.1)$$

Secondly, filter out all data vectors when the wind turbine is not producing any power. Finally, remove data vectors where one of the recorded values appears to be higher than the threshold values. These threshold values are tabulated in Table 4.3. These threshold values are based on manufacturers specifications. The data points removed after applying the general filter is shown in Figure 4.3a.

Table 4.3: Boundaries for parameters of the gearbox model

Gearbox Parameters	Rated Value	Boundary
Power [kW]	2000	[0,2000]
Rotor RPM	18.1	[0,18.2]
Ambient Temperature [° C]	-	[-20,40]
Nacelle Temperature [° C]	-	[-20,70]
Gear Bearing Temperature [° C]	-	[-20,90]
Gearbox Oil Temperature [° C]	-	[-20,90]

4.3.2. Cluster Filter

The clustering technique for SCADA data mining was first introduced by Kusiak et al., [36], and was later extended by Bangalore et al., [39]. The cluster filter is used to remove data outliers and data corresponding to curtailment conditions from the training data set. It is essential to remove the datavectors corresponding to curtailment, as it cannot be considered as normal behavior nor a fault. The algorithm version presented by Bangalore et al., is used in this thesis and is described in Algorithm 1.

Algorithm 1 Cluster filter algorithm [39]

- 1: Decide the maximum number of clusters, N .
 - 2: Assign a cluster number $n \in \{1, \dots, N\}$ to each input data vector D_i , $i \in \{1, \dots, \text{length (Dataset)}\}$ in training dataset using k-NN clustering method.
 - 3: Find the centroid, C_n for each cluster.
 - 4: Calculate the Mahalanobis distance MHD_i of each data vector from its cluster center C_n .
 - 5: Estimate the probability distribution for the Mahalanobis distances in the vector MHD .
 - 6: Eliminate data vectors whose probability of occurrence is less than a threshold value.
-

The training dataset is divided into 12 clusters using k-means clustering method, unlike using Ward's minimum variance algorithm as used by Bangalore et al., [39]. The number of clusters is decided based on the operating conditions of the wind turbines and their behavior. A probability threshold of 5% is chosen, and all the data vectors below this threshold is eliminated. A low threshold value suffices as curtailment is not done frequently. This cluster filter is applied only to training dataset and not to application dataset.

4.3.3. Missing Data Filter

Missing data filter is applied to both training and application datasets. The missing information filter is used to correct the issue, which is intended to guarantee that at least 1 h of continuous data is available for a parameter vector to be considered during the training and application phases. Stray data vectors which are not continuous are eliminated.

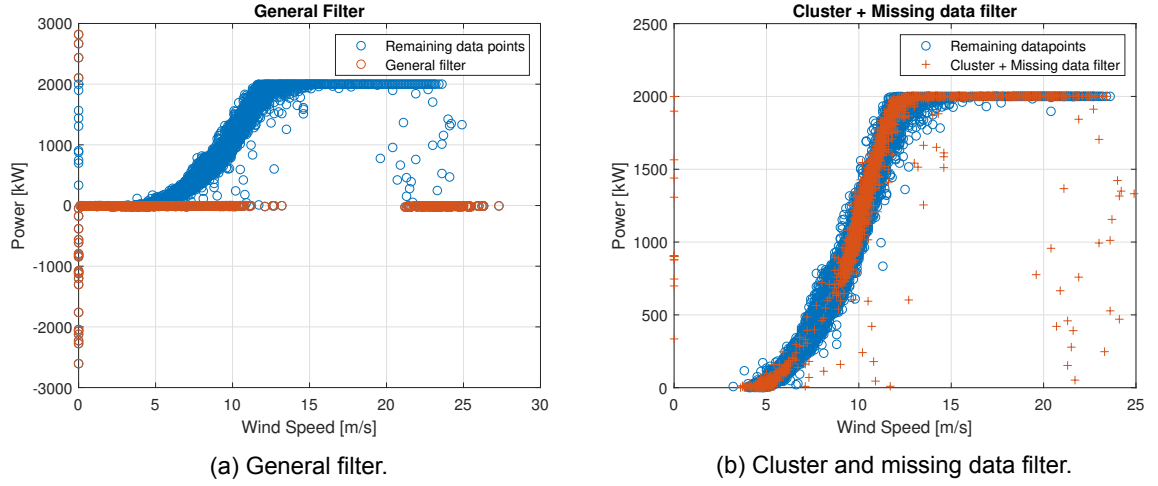


Figure 4.3: A pictorial representation of data removed from the training data set due to filtering - Power curve

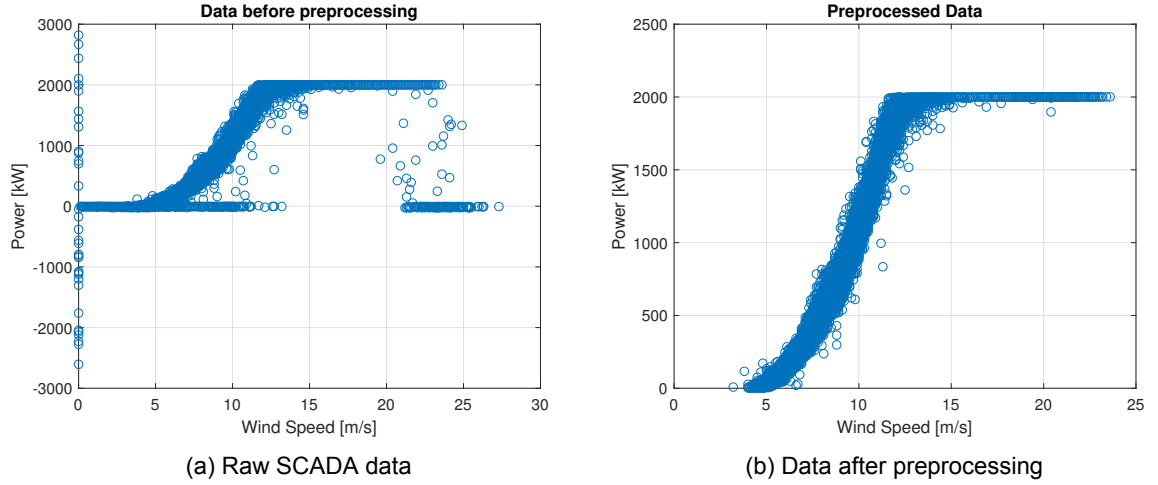


Figure 4.4: Power curve

4.3.4. Masking for RNN

RNN networks are primarily used for sequence modeling, i.e., they perform best with fixed length inputs or in other words evenly spaced time series. For LSTM input/output parameters unlike NARX data vectors that do not fit into the bill of normal behavior are *masked* instead of being removed. Masking is a process where all the data vectors that were supposed to be removed using data filters above is replaced with the value '0'. In general, with neural networks, it's safe to input missing values as 0, with the condition that 0 is not already a meaningful value. The network will learn from exposure to the data that the value 0 means missing data and will start ignoring the value during the training phase [76]. The mask layer can be applied using the line of code of *Keras* as shown below.

$$model.add(Masking(mask_value = 0., input_shape = (timesteps, features))) \quad (4.2)$$

For an input time series data \mathbf{X}_T if x_t is the missed element, which equals to 0, the training process at the t -th step will be skipped, and thus, the calculated cell state of the $(t-1)$ -th step will be directly input into the $(t+1)$ -th step. Figure 4.5 is a representation of the function of a mask layer.

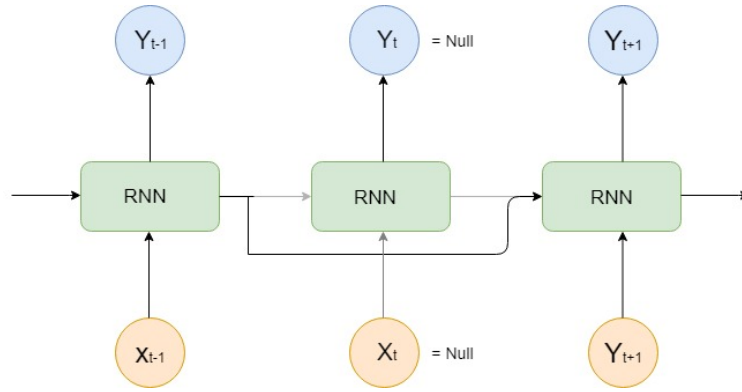


Figure 4.5: A pictorial representation of masking operation in LSTM models [77]

4.4. Artificial Neural Network Model Architecture

Data preprocessing step is followed by development of ANN configuration. There is no rule of thumb to decide the best possible configuration. All good performing ANN models are a result of trial and error. Within this thesis, model architecture was selected based on the NARX model used by Bangalore et al., [39] and also a novel stacked RNN based model as presented by [40]. The application of stacked RNN has not yet been tested for wind turbine condition monitoring yet, and has a potential to outperform existing models[43]. The following sections gives a detailed description of the specific model architecture and their performance based on fault free data.

The metric used to analyze the performance of the neural network model is Mean Squared Error (MSE). MSE is a common metric used to evaluate regression problems as it might be penalizing large errors more, because if your error is large, its square is much larger. Therefore, lower the MSE better the predictive performance of the model. The MSE of a predictor is defined as shown in Equation 4.3. Here n is the number of data points in the variable, Y_i and \hat{Y}_i are the measured and the predicted values respectively. In simple words, MSE is the average of square of errors.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4.3)$$

4.4.1. NARX model

The NARX model used in the thesis is similar to the one used by Bangalore et al., [39]. This model is used as benchmark, and the newly developed model in this thesis is compared against NARX model. Additionally the same model is used with additional inputs as shown in Table 4.2 to check for improvement in model performance if any.

The summary of the closed loop NARX is tabulated in Table 4.4. The model has one hidden layer with 20 neurons all densely connected, and one neuron for the output layer. The non linear activation function, *Sigmoid* is used as default. The pictorial representation of the implemented NARX model is shown in Figure 4.6.

Figure 4.7 shows a plot of modeled versus the measured bearing temperature on the application data. As it can be observed, the temperature predicted by the NARX model is almost similar to the measured temperature. As the weight initialization is done randomly, the mean squared error (MSE) of each model run will be slightly different. The average MSE of 30 models run on the same data is tabulated in Table 4.7. For reference, the MSE plot for one of the model is shown in Figure 4.8.

Table 4.4: Specification of NARX model.

ANN Model	NARX	
Layer	Hidden	Output
Neurons	20	2
Activation Function	Sigmoid	Linear
Input Parameters	1. Power 2. Rotor RPM 3. Nacelle Temperature 4. Ambient Temperature 5. Generator Torque 6. Missing Data Filter	
Output Parameters	1. Bearing Temperature 2. Gearbox Oil Temperature	

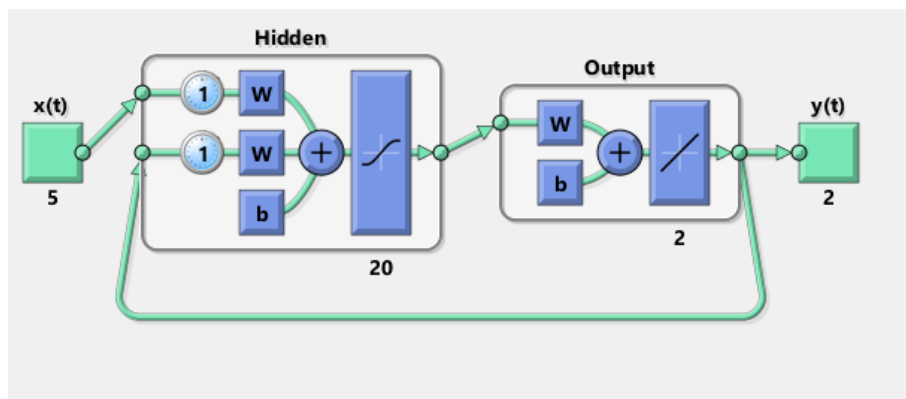


Figure 4.6: Implemented NARX Neural Network.

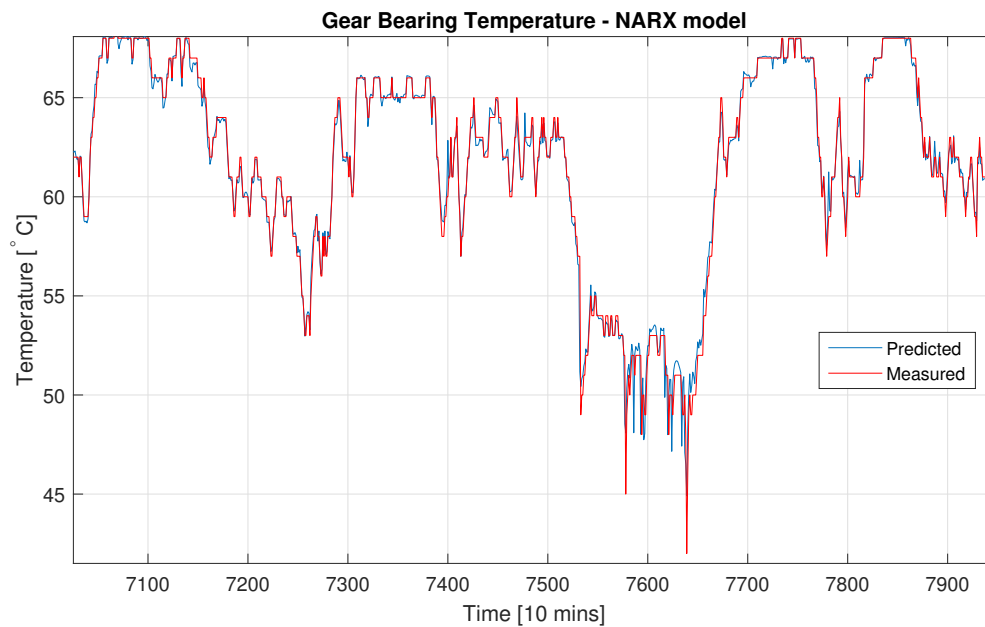


Figure 4.7: Bearing temperature - modeled versus measured NARX.

The error histogram of one such training output of NARX model is shown in Figure 4.9. The errors during the training phase is low, with 90% of the errors close to 0. Hence, it was concluded, that the errors are truly random and do not show a trend or are shifted by

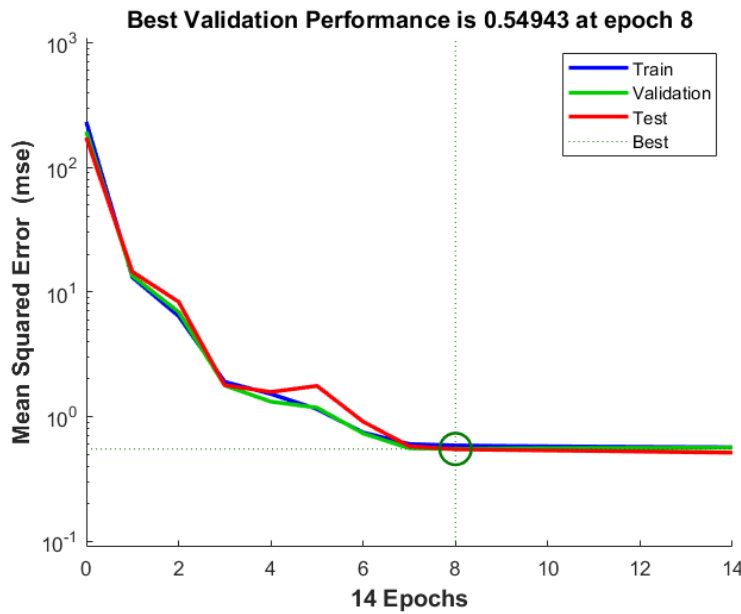


Figure 4.8: Mean squared error for NBM.

a functional shortcoming of the model itself. The errors on the training, validation and test set is almost comparable and do not show significant differences. Also, the training and validation losses almost converge as seen in Figure 4.8. This conclusively shows that there is no overfitting and the model has good generalization abilities.

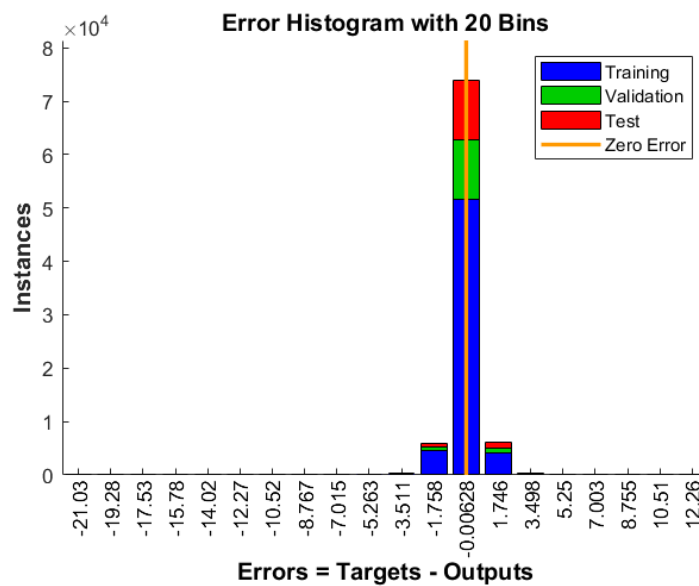


Figure 4.9: Training error histogram for bearing temperature.

4.4.2. Stacked RNN model

The RNN was designed using an open source framework called Keras [78]. Keras is a neural networks Application Programming Interface (API) written in Python, it runs on top of TensorFlow, a low level API for running machine learning projects. As RNNs require more memory and computation resources compared to a NARX model, the neural network was run on *Google Colaboratory*, which provides free Nvidia Tesla K4 GPUs for faster training of the neural networks.

In this thesis, both LSTM and GRU networks were tested. The performance of the GRU network was on par with the LSTM network with respect to the mean squared error as detailed in Section 4.4.3. A representation of a stacked RNN network is shown in Figure 4.10. It can be observed that the output of each GRU layer fed back into as input for the next timestep prediction. A stacked RNN architecture can extract higher level temporal information as compared to a single layer network.

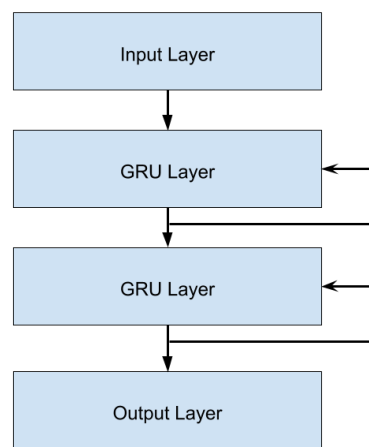


Figure 4.10: Representation of stacked GRU model.

As the RNN is trained on a time series data, it is quite natural to use a stateful mode of RNN. After dividing the training samples into batches as explained later in the section, it becomes to pass on the information from one batch to another while training an RNN. Stateful flag in Keras is introduced to circumvent these problems during training and make the model remember what happened in the previous batch by passing states from the previous batch to the next batch. If a stateless mode of RNN is used, the information from one batch is not passed on to another, and all the states are initialized to 0 after each batch.

4.4.3. Data scaling and hyperparameter tuning for RNN

A machine learning model is defined as a mathematical model with a number of parameters that need to be learned from the data. Some parameters are directly learnt from the training data such as weights in an ANN. A model hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data. Hyperparameter tuning is the process of determining the optimum values of these hyperparameters. A RNN model has many hyperparameters that has to be tuned manually to optimize the performance of the model. In the rest of this section, the process of tuning the hyperparameters is explained.

Data Scaling

The input and output data to the neural network has to be scaled to have all the features within the same range. This helps in faster training and improves the convergence of steepest

descent algorithms, which do not possess the property of scale invariance. Without data scaling, the weights of the network can grow very large leading to unstable networks and increasing generalization error.

MaxAbsScaler, a variant of *MinMaxScaler* is used to scale the input and output vectors. The scaled features are in the range $[-1,1]$. This scaler works best for cases where the distribution is not Gaussian or the standard deviation is very small. *MaxAbsScaler* feature scaling techniques do not affect the sparsity of the dataset, which also helps in implementing the masking layer without any hassles. Scaling the value between $[-1,1]$ for a LSTM/GRU layer is also preferred as the final output of these layers are squashed between $[-1,1]$ by the *tanh* function as shown in Figure 2.9 and 2.14. In this way, the network learns and converges faster. After training and doing predictions on the dataset, the scaling can be inversed to obtain the actual outputs.

Equation 4.4 is used by the *MaxAbsScaler*, where x_{scaled} is the scaled value of x and \mathbf{X} is column vector of the feature to be scaled -

$$x_{scaled} = \frac{x}{\max|\mathbf{X}|} \quad (4.4)$$

It is important to note here that, data scaling should always be done after splitting the dataset into training and test sets to prevent any data leakage. Data Leakage is the creation of unexpected additional information in the training data, allowing a model or machine learning algorithm to make unrealistically good predictions. For example, if the dataset is split after scaling, the training set can have a peek into the future already which leads to such incredibly good predictions. But in the real world, the machine learning model has no access to the test dataset.

Hidden Layers and Nodes

The two main hyperparameters that control the topology of an ANN are - the number of hidden layers and the number of nodes in each hidden layer. There is no fixed formula or a rule of thumb to determine the number of hidden layers and the number of nodes in these layers. An experimentation technique or manual tuning is done to determine the number of hidden layers. Various network (including LSTMs and GRUs) configurations with 1, 2 and 3 hidden layers with different permutation and combination of hidden nodes were tested. A GRU network with 2 hidden layers with 10 and 5 nodes in the 1st and 2nd hidden layer respectively, provided the minimum MSE. The prediction performance of LSTM networks was equivalent to that of GRUs but took more computation time. Hence, for anomaly detection, a GRU network is utilized. The MSE losses of the tested networks configurations are further detailed in Table 4.7.

Activation Functions and Weight Initialization

The purpose of an activation function is to add non-linearity to the output of the hidden neuron. A neural network without an activation function is essentially just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks. Activation functions is perhaps the most important hyperparameter to be tuned.

A *Rectified Linear Unit* (ReLU) is used for as activation function for the proposed neural network model in this thesis. Mathematically, ReLU can be defined using Equation 4.5 and visually as shown in Figure 4.11. The performance of the network using *tanh* (MSE = 0.1509°C) activation function was identical to *ReLU* (MSE = 0.1464°C).

$$R(z) = \begin{cases} 0, & \text{if } z < 0. \\ z, & \text{if } z \geq 0. \end{cases} \quad (4.5)$$

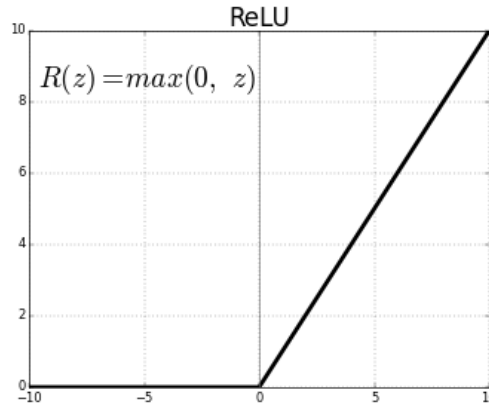


Figure 4.11: Representation of ReLU activation function.

ReLU can also help solve vanishing gradient problem[23]. Another important reason to use *ReLU* activation function is because it induces model sparsity (which is different compared to data sparsity) in the hidden units of the neural network, as explored by Glorot et al., 2011 [79]. Hence, *ReLU* was preferred over *tanh*.

Glorot (or Xavier) uniform initialization [80] is used for initializing the weight matrices.

Optimizer and Learning rate

The choice of an optimizer and learning rate is paramount to ensure stable performance of a neural network.

Optimizers help in minimizing the cost function $\mathbf{J}(\theta)$. In our case, $\mathbf{J}(\theta)$ is the mean squared error metric. A variant of gradient descent, Adam (**A**daptive **M**oment Estimation) [81] is used as optimizer. Adam optimizer is one of the most popular gradient descent optimization algorithms which is computationally efficient and has very little memory requirement [81]. Gradient descent is an iterative learning algorithm that uses a training dataset to update a model.

Learning rate controls the update speed of parameters. The weights of a network are updated through backpropagation as explained in section 2.5.3. The amount that the weights are updated during training is referred to as the step size or the “learning rate.” Generally, a large learning rate allows the model to learn faster, at the cost of arriving on a sub-optimal final set of weights. A smaller learning rate may allow the model to learn a more optimal or even globally optimal set of weights but may take significantly longer to train. Various learning rates - 1E-5, 1E-4, 0.001, 0.01 and 0.1 were experimented with to determine the right learning rate. A learning rate of **0.001** gave optimal solutions with reasonable computation times as compared to other rates with similar losses. The variation of MSE with various learning rates is shown in Figure 4.12.

Batch Size

The batch size defines the number of samples that will be propagated through the network before the weights are updated. For example, if all the training samples are used simultaneously to train the network, then the batch size is equal to 1. In most deep learning applications, the training samples are divided into smaller batches before feeding it into the network. There are some advantages and limitation of using such mini-batches for training deep networks. The main advantages of using mini-batches are the requirement of less memory to train the network. The network also trains faster with mini-batches. This is because the weights are updated after each batch propagation. In Figure 4.13, the direction of the mini-batch gradient (green color) fluctuates much more in comparison to the direction of the full batch gradient (blue color).

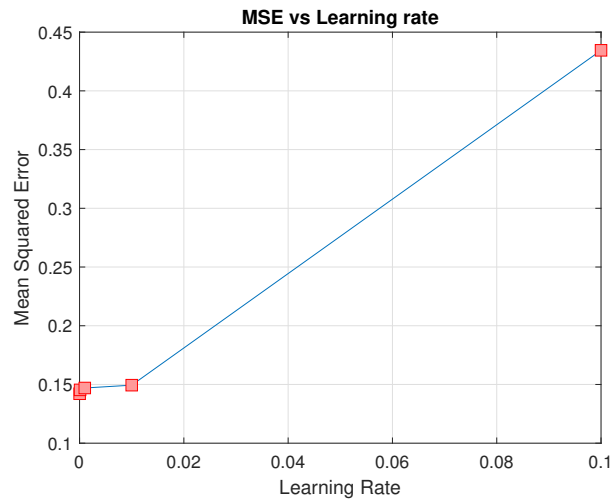


Figure 4.12: MSE vs learning rate for final GRU network.

Similar to learning rate, smaller samples in batches leads to increased computation time, while large batches require more memory and do not always converge. Various batch sizes - 32, 64, 128, 144, 256 and 512 were tested to tune the network. It should also be noted that the batch size should always be a multiple of the number of training samples. Also, the test set must be a multiple of the batch size. After many trial and errors, a batch size of 144 is found to give an optimum trade-off between computation time and network performance. The other advantage of having a batch size of 144 is that the average temperature values can be updated everyday to monitor differences between the average measured and average predicted temperatures, as explained in Chapter 5. As SCADA records 10-minutes averaged data, a total of 144 data points are obtained everyday.

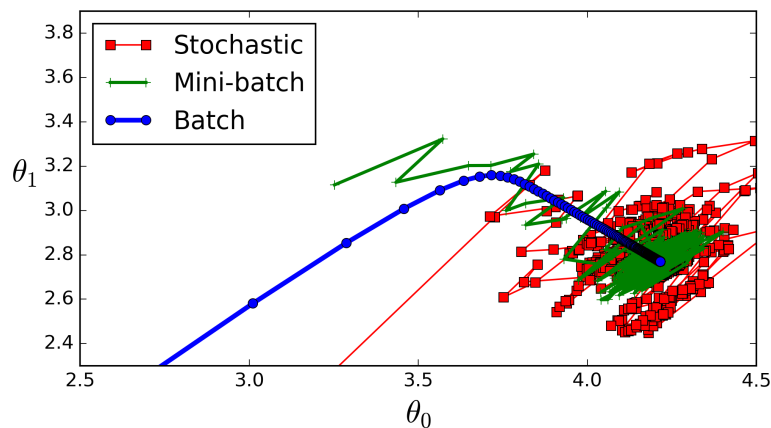


Figure 4.13: Gradient descent batch size and convergence.

Regularization

Regularization techniques are mainly used to reduce the overfitting of the neural networks. The 3 main regularization techniques generally used in machine learning projects are L1, L2 and dropout. In specific, recurrent dropout is used in this thesis. *Dropout* refers to randomly dropping connections between network nodes during training. This is a specific, built-in way to use dropout to fight overfitting in recurrent layers. A regular dropout is not applied to the input or output layer as it worsened the performance of the model. This might be due to the following reason - the logic of dropout is for adding noise to the neurons in order not to be dependent on any specific neuron. By adding a regular dropout for RNN cells, there is a

chance for forgetting something that should not be forgotten. A recurrent dropout of 0.3 and 0.1 in the first and second hidden layer respectively resulted in the lowest MSE.

Callbacks

It is necessary to train the neural networks with optimum number of epochs. One epoch is one forward pass and one backward pass of all the training examples through the network. Too many epochs can lead to model overfitting while less number of epochs can lead to underfitting without good generalization of the model. Early stopping requires that a validation dataset is evaluated during training. Early Stopping is a callback method used to stop the training once the performance of the validation dataset stops improving or starts deteriorating. In this thesis, early callback with a patience of 10 epochs is implemented to make sure the neural network has good generalization. This is similar to the callback applied in MATLAB NARX network. A modern way of avoiding overfitting, is to use a dropout layer along with early stopping [23], which is successfully implemented in this thesis.

The values of the hyperparameter used for training and testing the SCADA data is tabulated in Table 4.5. The GRU network architecture in which the MSE was minimal with respect to other architectures is tabulated in Table 4.6. In case of both NARX and RNN networks, additional input of generator torque improved the predictive performance of the algorithms.

Table 4.5: Final tuned hyperparameters for GRU network.

Hyperparameter		Value(s)
No. of hidden layers		2
No. of nodes in hidden layers	1st Layer	10
	2nd Layer	5
Activation Function		ReLU
Optimizer		Adam
Learning rate		0.001
Batch Size		144
Recurrent dropout in hidden layer	1st Layer	0.3
	2nd Layer	0.1

Table 4.6: Specification of GRU network.

ANN Model	Gated Recurrent Unit (GRU)		
Layer	1st hidden layer	2nd hidden layer	Output
Neurons	10	5	2
Activation Function	ReLU	ReLU	Linear
Input Parameters	1. Power 2. Rotor RPM 3. Ambient Temperature 4. Nacelle Temperature 5. Generator Torque		
Output Parameters	1. Bearing Temperature 2. Gearbox Oil Temperature		

Figure 4.14 shows the network configuration for gear bearing temperature prediction model. An identical model is also run in parallel to predict gearbox oil temperature. As seen, the total params represents the number of weights that has to be adjusted to fit the training data, which is 726 in this case.

Figure 4.15 shows the measured and predicted bearing temperatures for 5 days. The predictions seem to fit the measured data quite well. Additional prediction results are presented

Layer (type)	Output Shape	Param #
masking_1 (Masking)	(144, 1, 5)	0
gru_1 (GRU)	(144, 1, 10)	480
activation_1 (Activation)	(144, 1, 10)	0
gru_2 (GRU)	(144, 5)	240
activation_2 (Activation)	(144, 5)	0
dense_1 (Dense)	(144, 1)	6
activation_3 (Activation)	(144, 1)	0
Total params: 726		
Trainable params: 726		
Non-trainable params: 0		

Figure 4.14: Network configuration and number of weight parameters to be tuned - 726.

in Appendix B. To check for overfitting, the training and test loss curves is shown in Figure 4.16. The training and the test loss curves are almost converging, confirming that the model has not overfitted on the training data. The MSE of the proposed GRU model for gear bearing temperature is 0.1464°C . The MSE of the lubrication oil temperature prediction model is 0.2375°C , a little higher than bearing temperature model. This is because the oil temperature is highly volatile and changes more rapidly than the gear bearing temperature.

Table 4.7 presents the performance of few of the various network architectures tested. The best prediction performance model had 2 hidden GRU layers with 10 and 5 nodes respectively. Based on the prediction performance of the tested networks, GRU was chosen as apt for anomaly detection task.

Table 4.7: MSE metric results for some of the configurations.

ANN Model	Hidden Layers	Neurons in hidden layers		Mean Squared Error [$^{\circ}\text{C}$]
		Layer - 1	Layer - 2	
NARX	1	20	-	0.5494
LSTM	1	10	-	0.3792
		20		0.2312
		50		0.2984
LSTM	2	10	5	0.1478
		20	5	0.1611
		50	20	0.2981
GRU	1	10	-	0.3672
		20		0.2452
		50		0.2876
GRU	2	10	5	0.1464
		20	5	0.1543
		50	20	0.3241

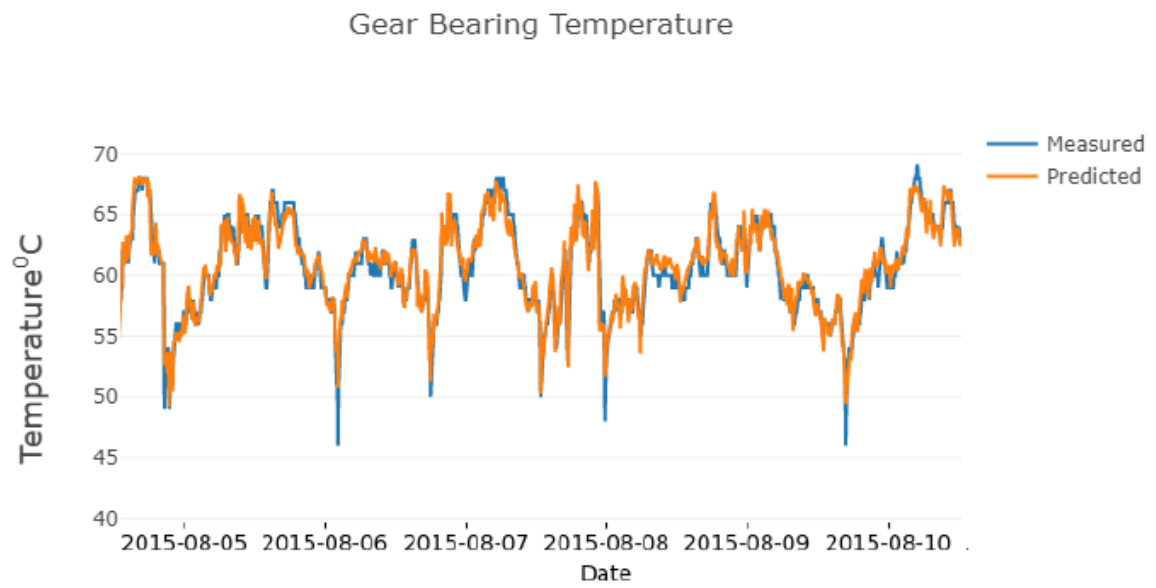


Figure 4.15: Measured and GRU predicted gear bearing temperature.

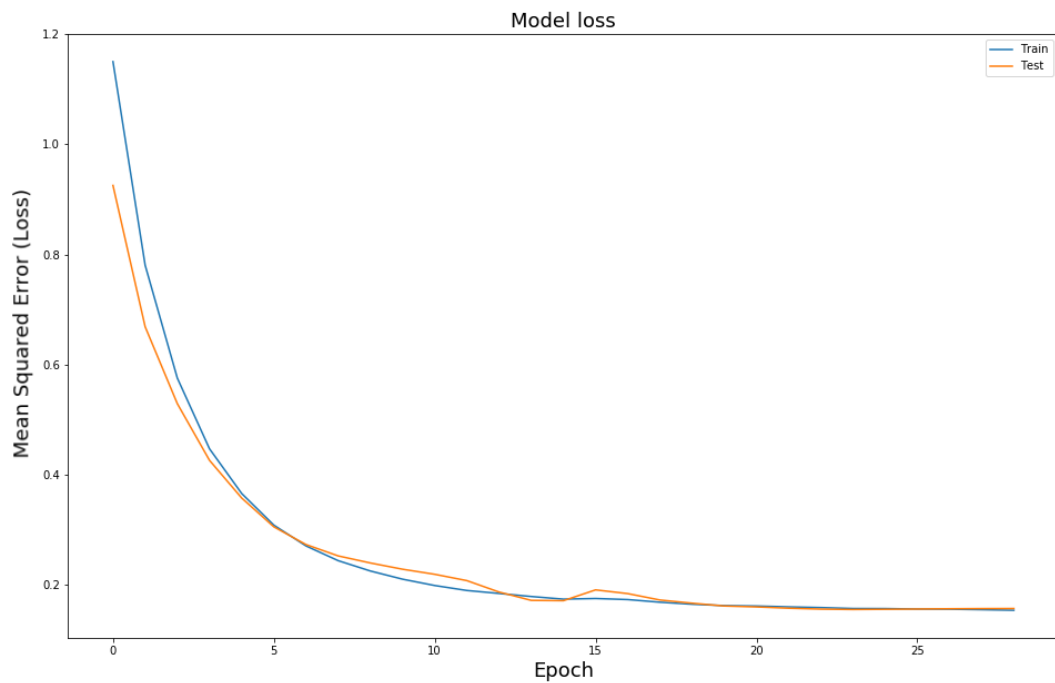


Figure 4.16: GRU network performance with learning rate 0.001 for gear bearing temperature model.

5

Case Studies

In this chapter, 5 case studies on the implementation of the proposed condition monitoring technique are presented. Also, the limitations of temperature based anomaly detection model are detailed.

Before delving into the case studies it is important to understand the technique used for data post-processing to flag anomalies. It is also important to note here that only two temperature sensors are available in the wind turbine gearbox under investigation, unlike modern turbines which have multiple sensors. One of the sensors is on the tapered roller bearing of the high-speed shaft and the other in the gearbox oil sump.

5.1. Data Post-processing

The purpose of developing normal behavior models is to be able to detect anomalies in the components being modeled. Various methods have been suggested by different researchers for anomaly detection using ANN models. Bangalore et al., 2017 [39] use Mahalanobis distance between the measured and predicted temperature. An anomaly is flagged if the Mahalanobis distance crosses a certain threshold. Zeher et al., [35] used an increase in the frequency of errors between the predicted and measured parameter values to flag an anomaly. Kusiak et al., [36] made use of an arbitrarily considered hourly averaged error residual to flag anomalies. A similar method is used by DNV-GL Wind Gemini [82] digital twin model to flag anomalies.

In this thesis, a similar method used by DNV-GL Wind Gemini as mentioned earlier is used to flag anomalies. An anomaly is flagged when the daily averaged residual crosses a threshold of 3°C . A moving window length of 144 is chosen, that is moving average over a day as it helps make a clearer distinction between anomalies and fault-free operation. Different window lengths of 36, 72, 144, 288 were tried. Apart from this threshold, an increasing trend and more frequent occurrence of alarms also suggest an impending fault in the wind turbine. Taking a moving average of residual helps in removing the effect of noise. This threshold limit is based on the average residual observed during the normal behavior period as explained later in the case studies. It is important to mention that the analysis presented in this chapter is targeted to identify and predict over-temperature cases only. An anomaly is flagged only if the average residual is above $+3^{\circ}\text{C}$.

The predictions of neural networks on unseen data are stochastic. Neural networks use randomness by design to ensure they effectively learn the function being approximated for the problem. In the case of normal behavior modeling, the predictions on anomalous data are not uniform and slightly different for each run. This is because the ANN model is not trained

on faulty data and also because the initial weight initialization is random. The randomness is also due to optimization which is based on stochastic gradient descent. The optimizer can sometimes get stuck at local minima. To counter these effects, the output is averaged over the 10 best ANN models selected from a total of 20 trained models. Though this process is tedious, it always results in better predictions for condition monitoring.

Finally, sometimes the ANN can give false positives, in other words, some false alarms due to various reasons such as excessive noise in the test data or if the cooling fans in the nacelle are not working properly and so on. To avoid this, the proposed physics of failure can be used. When an anomaly is detected in the ANN-based models, the historical damage operating conditions of the anomalous turbine can be compared against the rest of the fleet in the same wind farm. If the turbine is found to be operating under more damage-inducing conditions, then the flagged anomaly is most certainly a true positive.

5.2. Case studies of HSS bearing failures

In this section, the turbines in which the high speed shaft bearings failed are studied. The input and output SCADA parameters used for anomaly detection for all case studies is presented in Appendix B along CMS vibration analysis reports. The proposed ANN + PoF is compared with vibration based monitoring systems. It is important to note here that there are 4 levels of severity in vibration based monitoring system as described in Table 5.1.

Table 5.1: Vibration monitoring system alarms.

Severity Level	Type	Description	Recommended Action
1	Danger	Severe progressing alarm	Immediate action
2	Alert	Considerable progressing alarm	2-4 weeks
3	Alert	Progressing alarm	2-4 months
4	Alert	Small or none progressing alarm	None
5	System	Hardware system problem	None

The summary of the alarm trigger and gearbox component failure dates is tabulated in Table 5.2.

Table 5.2: Summary of all case studies with alarm trigger dates.

Turbine No.	Failure Date	Root Cause of Failure	Gear Bearing Alarm	Gearbox Oil Alarm	CMS alarm (Severity 3) Stage - Date	CMS alarm (Severity 2) Stage - Date
7	5 Apr 2013	Spalling in HSS TRB	30 Oct 2012	2 Oct 2012	High speed stage- 19 Sep 2012	High speed stage- 3 Apr 2013
23	19 Mar 2013	Axial crack in HSS TRB	31 Oct 2012	31 Oct 2012	High speed stage- 21 Dec 2012	High speed stage- 4 Feb 2013
48	21 May 2015	Spalling on HSS bearings and PS bearing	21 Oct 2014	5 Nov 2014	Planetary stage - 11 Jul 2014	N.A
32	1 June 2014	Micropitting on all gears and indentations on few bearings	No alarm	14 Mar 2014	High speed stage- 13 Mar 2014	N.A

5.2.1. Turbine - 48

The complete gearbox of Turbine-48 was replaced on 21st May 2015. According to root cause analysis reports, there was spalling observed in all 3 bearings of high speed shaft and some micropitting in IMS stage bearings. The ANN model is trained on data from the year of 2012 and applied for anomaly detection during the period of 21 May 2014 to 21 May 2015. The fault free data of the calendar year 2013 is used as test data to check whether the model gives false alarms. The output is presented for both the gear bearing and the gearbox lubrication oil models.

The residuals of bearing and oil temperature of gearbox averaged over 10 models for the year 2013 when the turbine was operating normally is presented in Figure 5.1 and 5.2. In this case, it is observed that the residual temperature is always below $+3^{\circ}\text{C}$ suggesting that the turbine is in a healthy state.

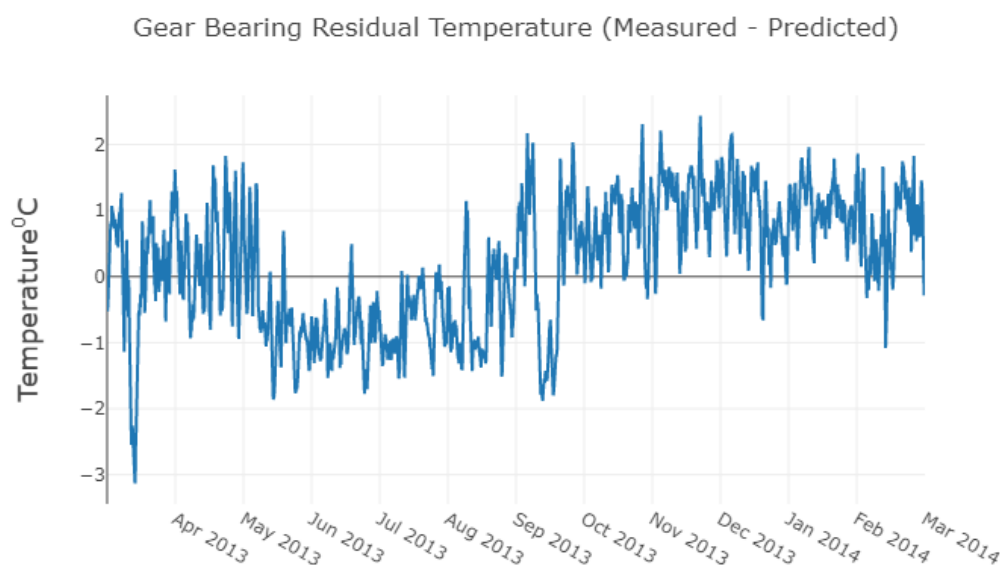


Figure 5.1: Gear bearing residual temperature for a healthy turbine.

The residuals of bearing and oil temperature of gearbox averaged over 10 models for the year 2014-15 when the turbine was operating abnormally is presented in Figure 5.3 and 5.4. In this case, it can be observed that there are instances when the residual temperature is above $+3^{\circ}\text{C}$ suggesting that the turbine is operating in a faulty state. The first alarm in the gear bearing and gearbox oil temperature model occurs on 21 October 2014 and 5 November 2014 respectively, roughly around 7 months before bearing replacements. According to CMS vibration analysis reports, an alarm of severity level 3 on a fault in the planetary stage was received on 11th July 2014. No vibration analysis report of the high speed shaft bearings was available.

The input and the output parameters for the turbine is shown in Figure B.1 and B.2. It can be observed that the input and output parameters for the month of September and October, when the anomaly was detected are within the limits of the data provided to the ANN model during the training process. Therefore, it can be inferred that the anomaly detected is due to abnormal behaviour of the turbine gearbox.

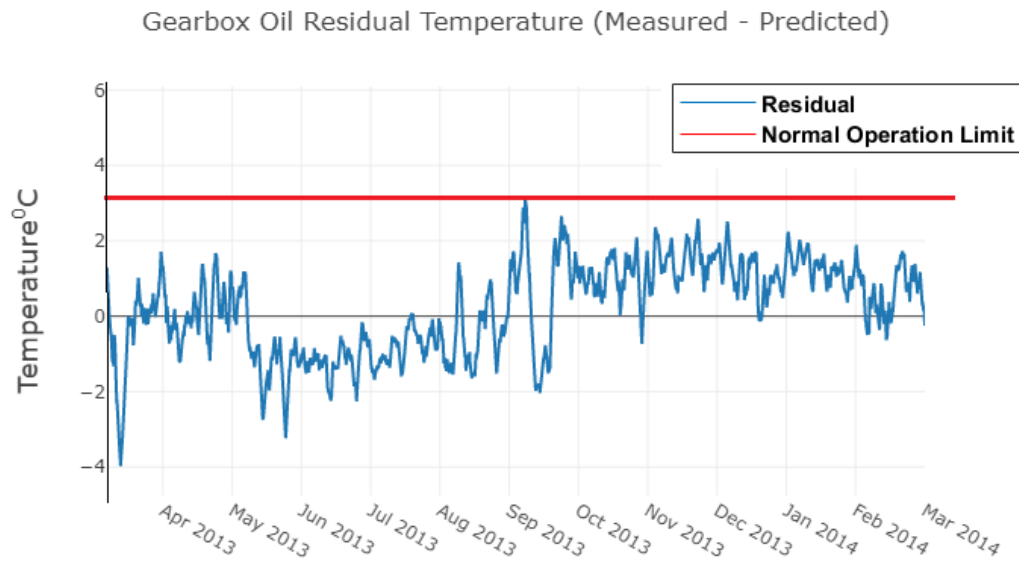


Figure 5.2: Gearbox oil residual temperature for a healthy turbine.

Physics of failure model output: A boxplot representing turbine 48 against the rest of the fleet with respect to damage operating conditions as described in Section 3.10 is shown in Figure 5.5. The red points represent the position of turbine - 48 with respect to 60 wind turbines present in the wind farm. Turbine 48 has relatively experienced more number of braking events at high wind speeds and rated power, and also has operated for a longer periods of time at low load and lambda value with respect to the entire fleet. This signifies that the failed turbine was operating at a high risk of failure before breakdown based on its operating conditions. The input and output SCADA parameters used for anomaly detection is presented in Appendix B.

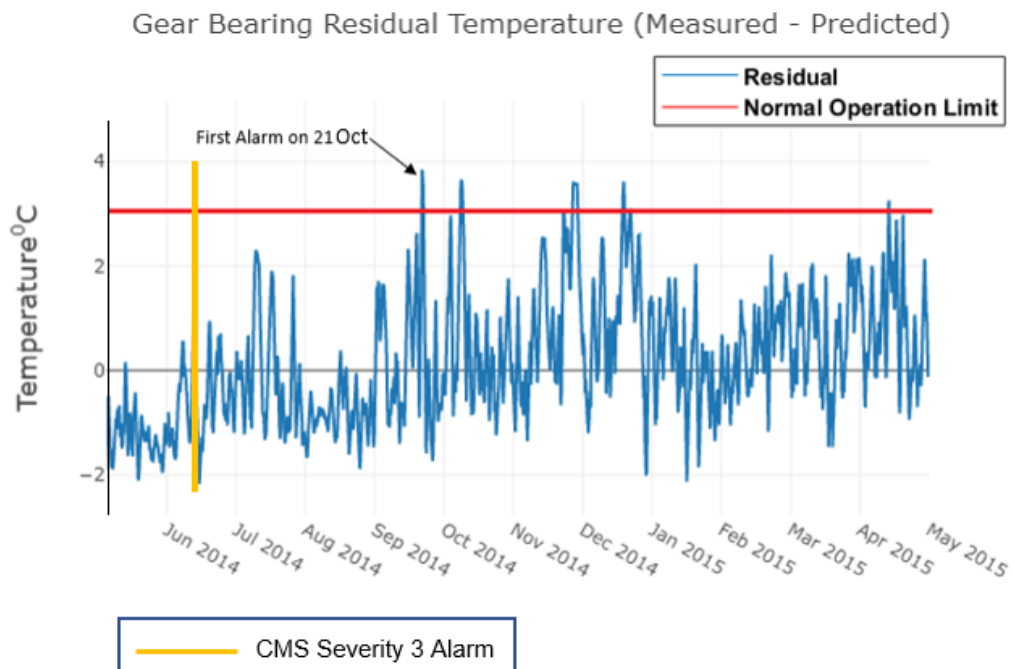


Figure 5.3: Gear bearing residual temperature for a faulty turbine.

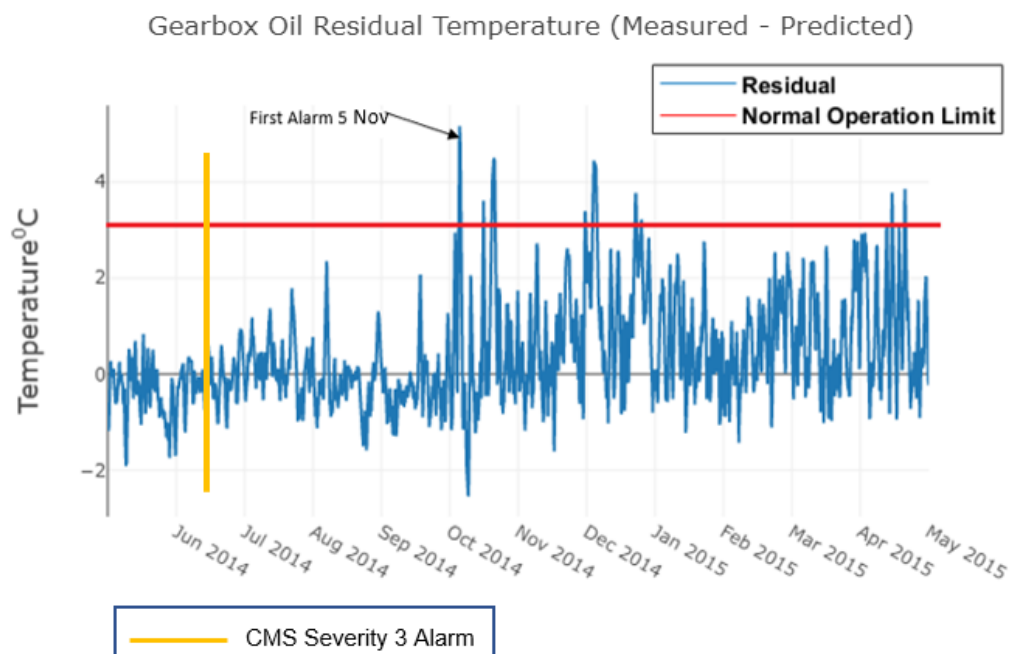


Figure 5.4: Gearbox oil residual temperature for a faulty turbine.

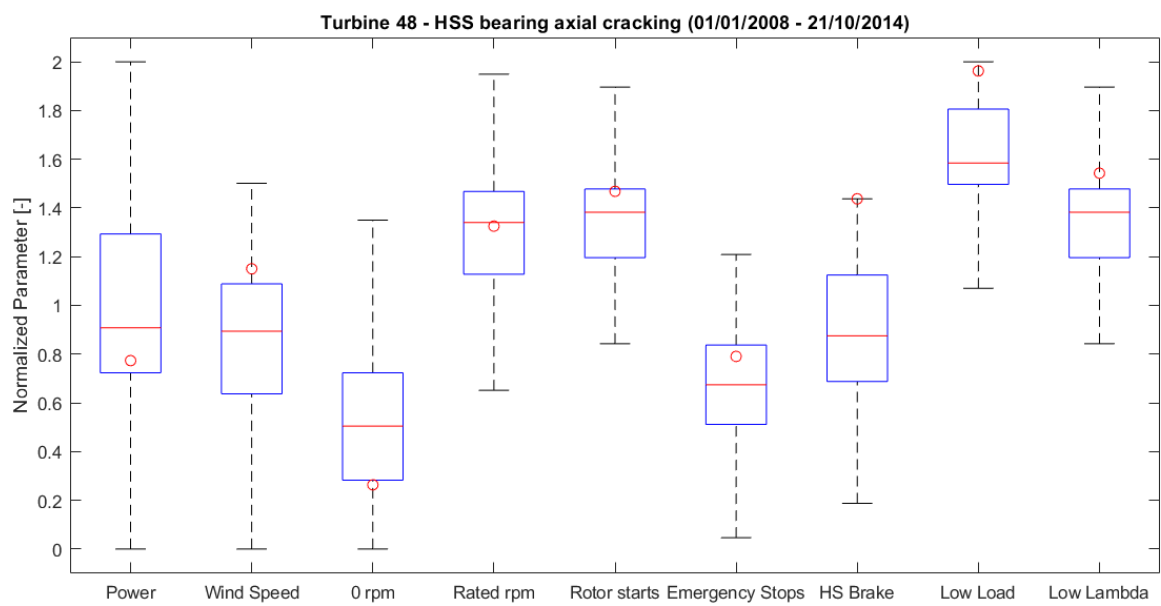


Figure 5.5: Turbine 48 - HSS bearing axial cracking

5.2.2. Turbine - 23

The high-speed shaft and the bearings of the Turbine - 23 failed on 19th March 2013. There were axial hairline cracks observed in the inner ring of the tapered roller bearings. The ANN model was trained on data from the year of 2010 and applied for anomaly detection from June 2012 to March 2013. The fault-free data of the calendar year 2011 was used as test data to check whether the model gives false alarms. The output is presented for both the gear bearing and the gearbox lubrication oil models.

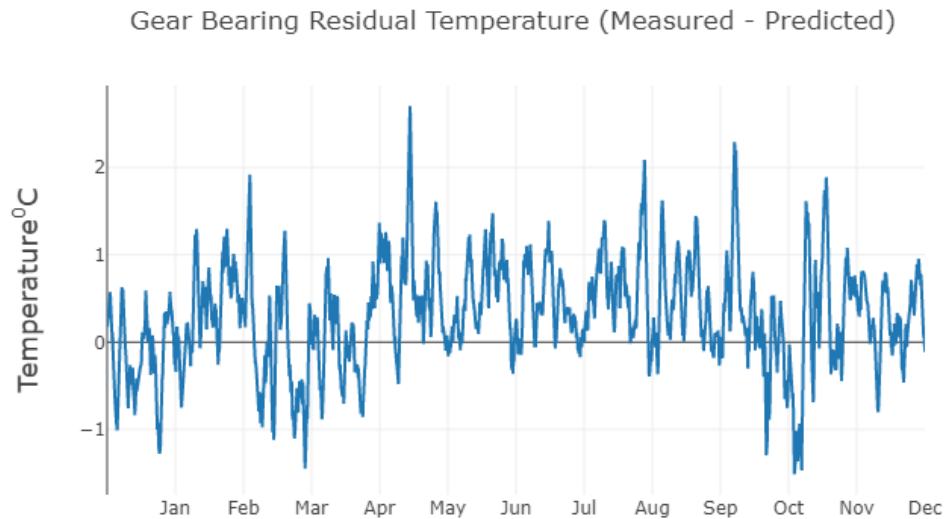


Figure 5.6: Gear bearing residual temperature for a healthy turbine - Year 2011.

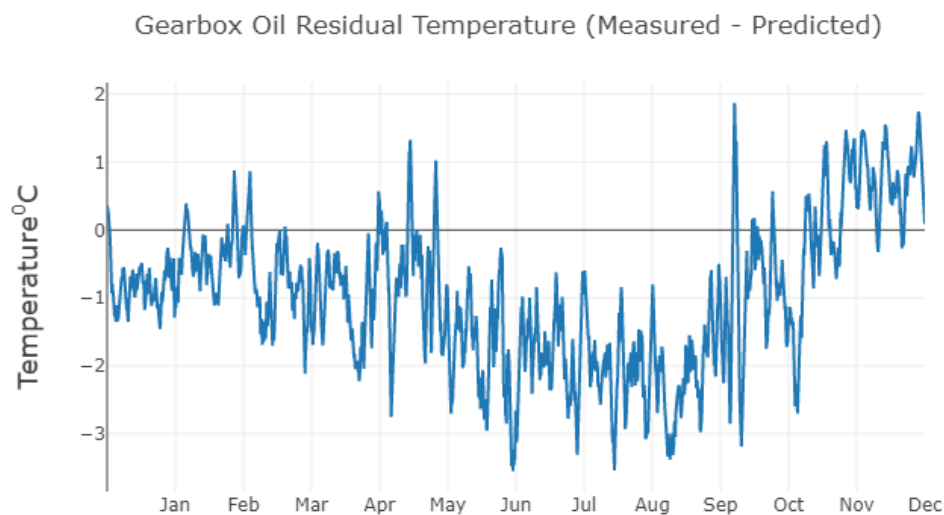


Figure 5.7: Gearbox oil residual temperature for a healthy turbine - Year 2011.

The figures 5.6 and 5.7 shows the residual temperatures for periods when the turbine was operating normally. The residual is always below $+3^{\circ}\text{C}$. Figures 5.8 and 5.9 shows the residuals for period before the failure. The first alarm in both the bearing and oil temperature model occurs on 31st October 2012, about 5 months before the replacement. There is an increase in the frequency of alarms before the eventual failure of the bearings. An alarm of severity level 3 and severity level 2 were received on 21st December 2012 and 4th February 2013 respectively from the vibration monitoring systems. The vibration analysis of the high speed shaft bearing is shown in Figure B.6. The trend of vibration analysis complements the temperature based model. Hence, it is safe to say that the performance of the ANN model is on par with the much more evolved vibration analysis techniques, in some cases even better.

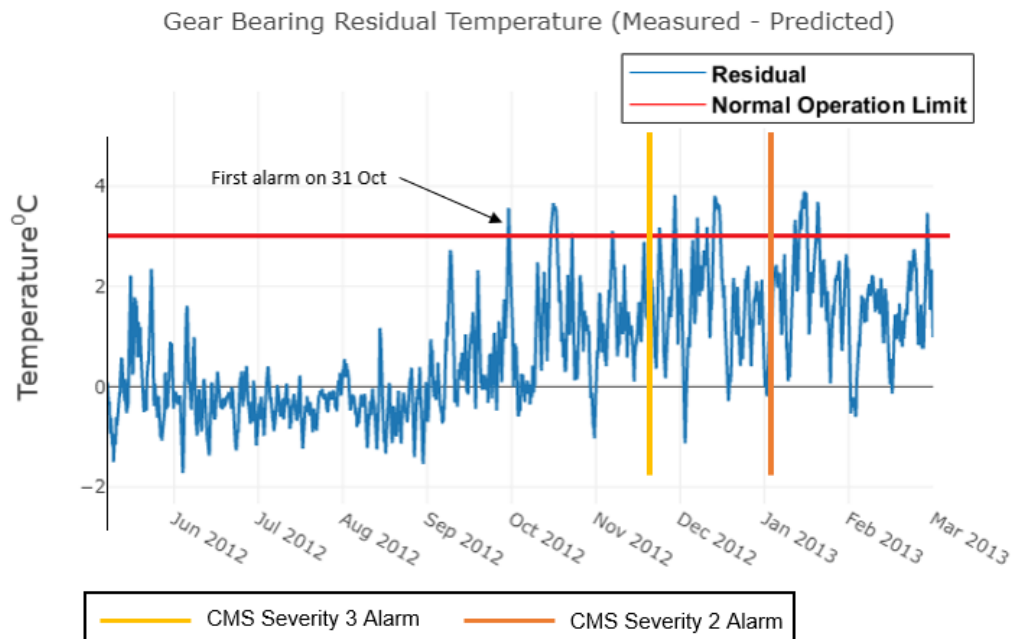


Figure 5.8: Gear bearing residual temperature for a faulty turbine.

Physics of failure model output: A boxplot representing turbine 23 against the rest of the fleet with respect to damage operating conditions as described in Section 3.10 is shown in Figure 5.10. It is observed that Turbine - 23 has experienced more braking events at high wind speed when compared to the average of the rest of the fleet. This signifies that the failed turbine was operating at a high risk of failure on a wind farm level before breakdown based on its operating conditions.

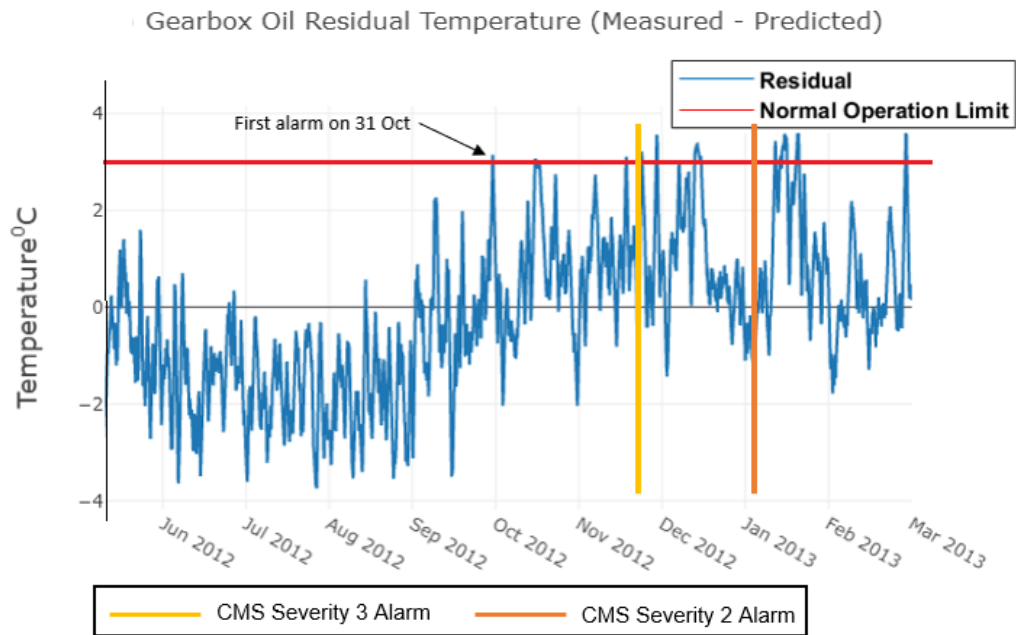


Figure 5.9: Gearbox oil residual temperature for a faulty turbine.

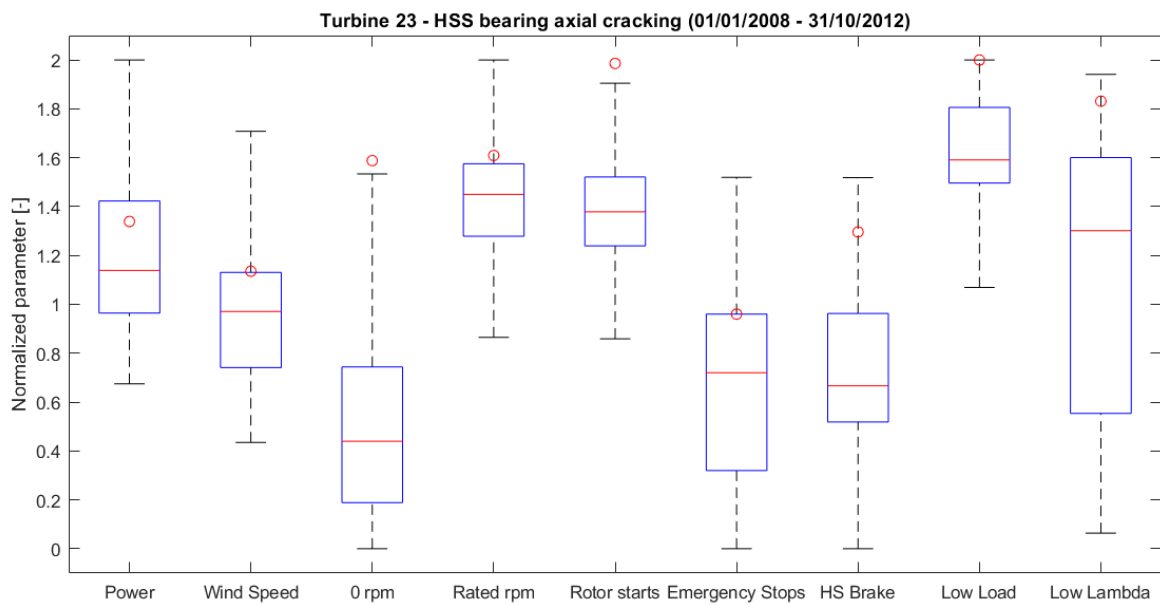


Figure 5.10: Turbine 23 - HSS bearing axial cracking

5.2.3. Turbine - 7

The high speed shaft and the bearings of Turbine - 23 failed on 5th April 2013. There was spalling observed in the inner ring of the tapered roller bearings. The ANN model was trained on data from the year of 2010 and applied for anomaly detection from August 2012 to April 2013. The fault free data of the calendar year 2011 - 2012 was used as test data to check whether the model gives false alarms. The output is presented for both the gear bearing and the gearbox lubrication oil models.

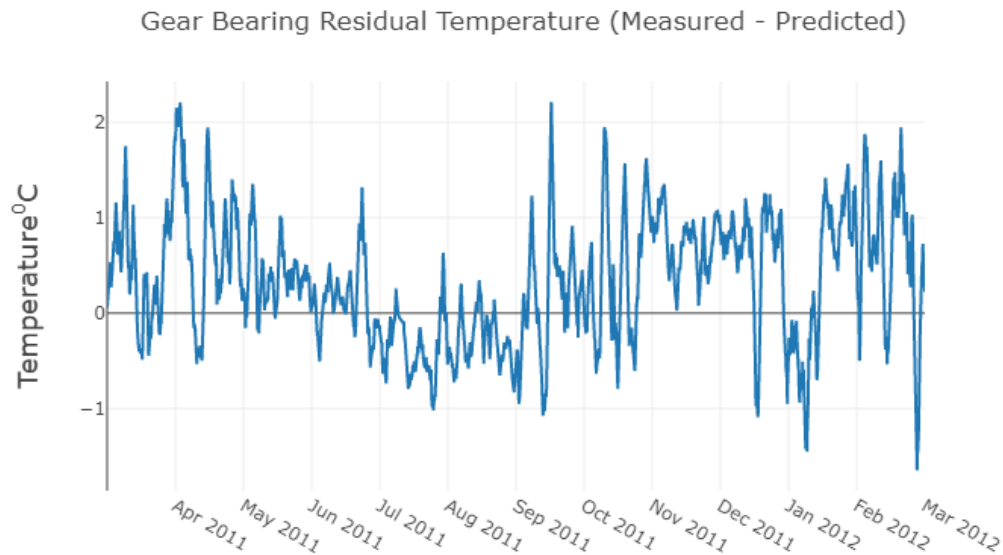


Figure 5.11: Gear bearing residual temperature for a healthy turbine.

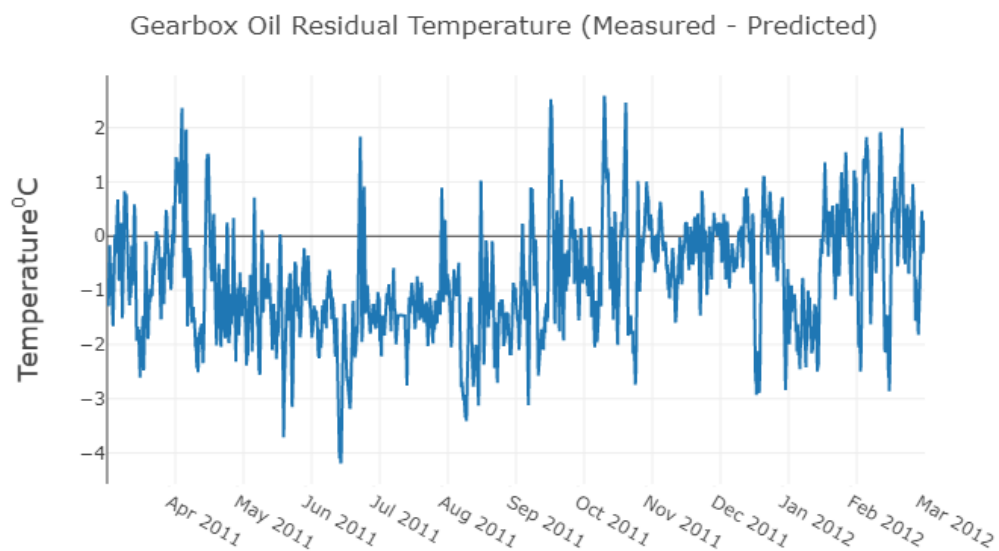


Figure 5.12: Gearbox oil residual temperature for a healthy turbine.

The figures 5.11 and 5.12 shows the residual temperatures for periods when the turbine was operating normally, and as observed the residual temperatures is always below $+3^{\circ}\text{C}$.

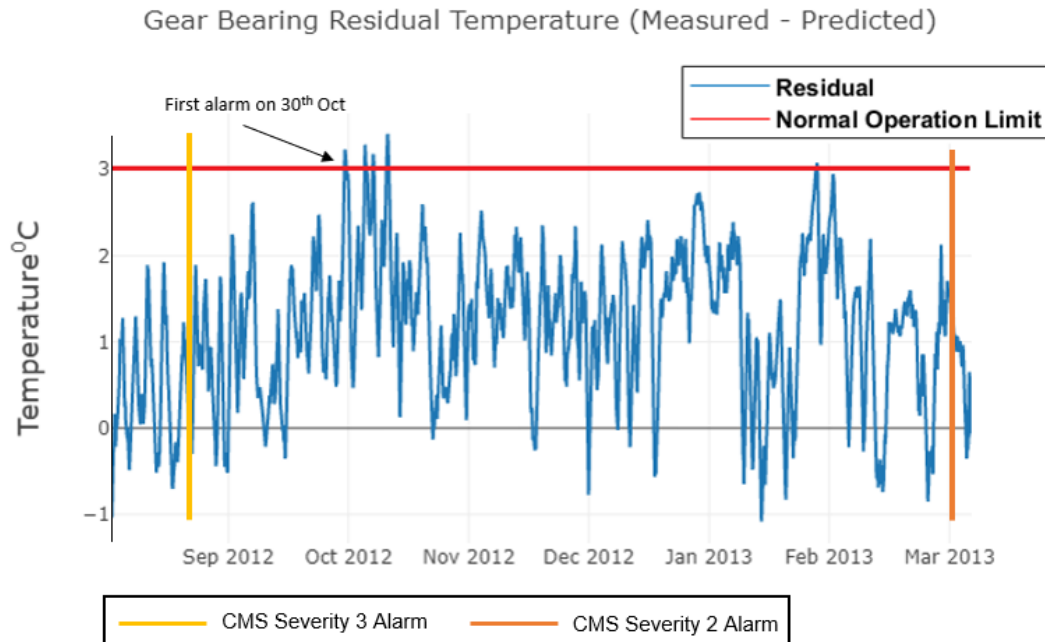


Figure 5.13: Gear bearing residual temperature for a faulty turbine.

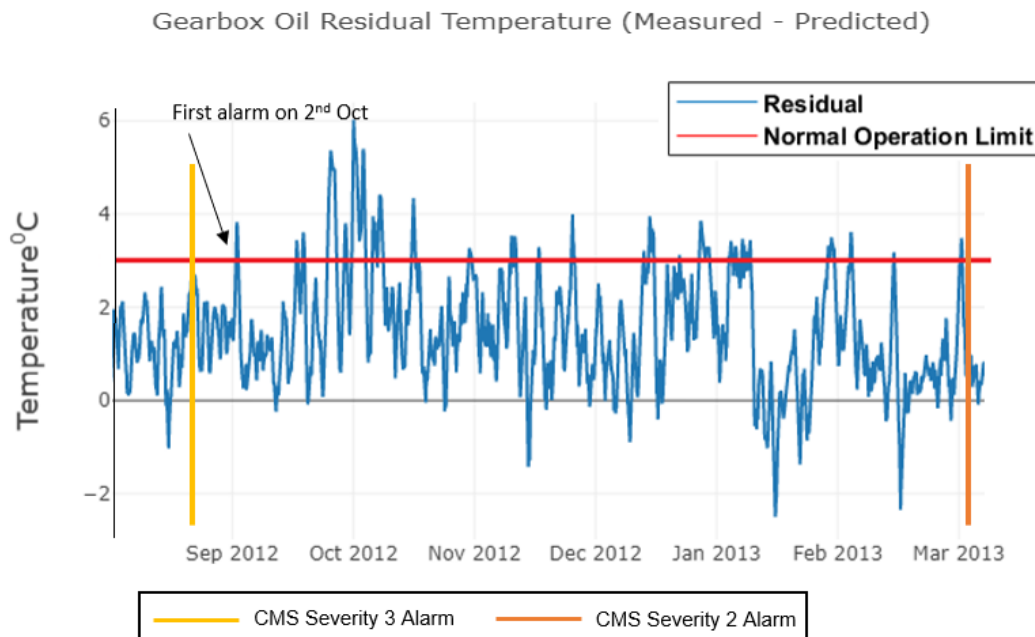


Figure 5.14: Gearbox oil residual temperature for a faulty turbine.

The first alarm from the oil temperature model occurs on 2 October 2012, while bearing temperature model detects an anomaly on 30 October 2012 as shown in Figures 5.13 and 5.14. This is followed by multiple alarms until the failure of the component. There was also an indication of low oil level in the gearbox in the months of November 2012 to January 2013

given by SCADA alarms, which could be another reason for multiple alarms given by the oil temperature model. The vibration monitoring of high-speed shaft bearing is shown in Figure B.12. There is a rising trend in vibrations from the end of September 2012. A severity 3 alarm was received on 19th September 2012 and severity 2 alarm on 3rd April 2013.

As per inspection reports, based on vibration monitoring system alarms, there was an inspection done on the gearbox on 10th October 2012, but nothing alarming was found. It was recommended to keep the turbine under close observation for any further signals from the vibration monitoring system. As there were no actions taken, the condition of the bearings continued to deteriorate as seen in both vibration monitoring system and ANN-based temperature models. Both oil and bearings temperature models send out multiple alarms before the bearing failure. The ANN models could detect anomaly 6 months before the component was replaced, almost on par with the vibration monitoring system.

Physics of failure model output: A boxplot representing turbine 7 against the rest of the fleet with respect to damage operating conditions as described in Section 3.10 is shown in Figure 5.15. It is observed that Turbine - 7 has experienced more braking events at high wind speed when compared to the average of the rest of the fleet. Also it was operating at idling speeds for longer periods as compared to the mean of the rest of the fleet. This signifies that the failed turbine was operating at a high risk of failure on a wind farm level before breakdown based on its operating conditions.

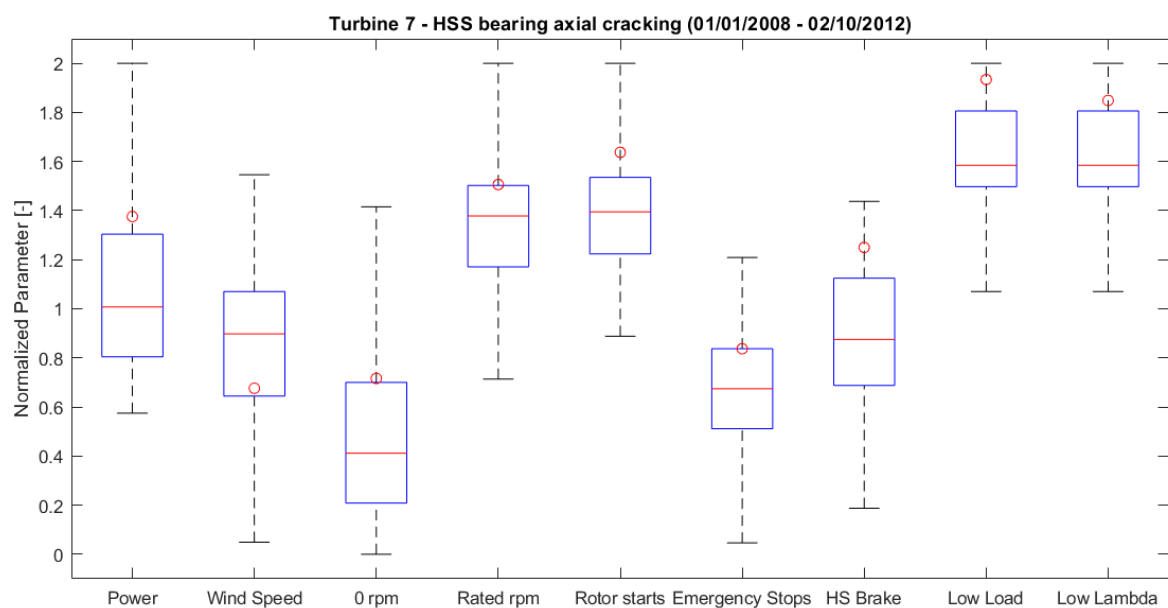


Figure 5.15: Turbine 7 - HSS bearing axial cracking

5.3. Discussion

There is a recurring pattern observed in the physics of failure plots based on operational statistics for bearings which failed due to *White Etching Cracks*. In all the 3 case studies presented in Section 5.2, it is clear that all the failed bearings had experienced *higher number of braking events at rated power, or had more emergency braking events, or was operating at low rotational speeds for longer duration* when compared to the average of the fleet.

It is important to note here that, none of the White Etching cracks occurs due to high cycle fatigue. It is also clear that the operational data of the turbines whose HSS bearings failed due to WEC correlate to the various hypothetical failure theories as mentioned in 3.9.1, especially braking events at rated power and operation at low lambda.

5.4. Damage Risk Assessment

The developed methodology uses the outputs of both physics of failure model and ANN model to flag anomalies.

Table 5.3: Signals triggered after ANN based temperature model detects anomaly.

Turbine No.	ANN Model Alarm	0 RPM	Emergency Braking	High Speed Brake	Low Load	Low Lambda	Rotor Starts
7	✓	✓	✓	✓	✓	✓	✓
23	✓	✓	✓	✓	✓	✓	✓
48	✓	✗	✗	✓	✓	✓	✗
2	N.A	✓	✗	✗	✓	✓	✓

Usually an alarm from both ANN based temperature model and PoF model can indicate a high probability of failure before 3-6 months prior to replacement of the components. It is important to note here that, damage operating conditions such as *Power, Rated rpm, Wind Speed* as defined in section 3.3 are not the root causes of failure. These parameters just accelerate the time to failure. It could therefore be argued that keeping track of damage operating conditions, especially *high speed braking events, emergency stops, operation at low load and lambda* to assess the risk of damage in high speed shaft of a wind turbine gearbox might be more helpful.

5.5. Case Study for non HSS bearing failures

In this section a case study to show the effectiveness of the temperature based anomaly detection model for detecting failure in components other than high speed shaft bearing is shown. A physics of failure model is not presented here, as only load sharing on the high speed shaft bearings was modeled for the thesis due to time constraint.

5.5.1. Turbine - 32

The complete gearbox of turbine-32 was replaced on 1st June 2014. The root cause for failure was reported as micropitting on all the gears and indentations due to those debris on few bearings. The ANN model was trained on data from June 2012 to June 2013 and applied for anomaly detection from June 2013 to May 2014. The ANN-based bearing temperature model was not able to detect any anomaly in the system, but the oil temperature model gave an alarm on 14th February 2014, approximately 3 months prior to failure as shown in Figures 5.16 and 5.17 respectively. A severity 3 alarm was raised on 11th March 2014 by the CMS vibration systems and also an increasing trend is visible from November 2013

as shown in Figure B.9. This case study provides an insight into the limits of the proposed method. In general, it can be inferred that the CMS can detect those failure modes of the gearbox that manifest as a change in the behavior of the gearbox bearing and lubrication oil temperatures.

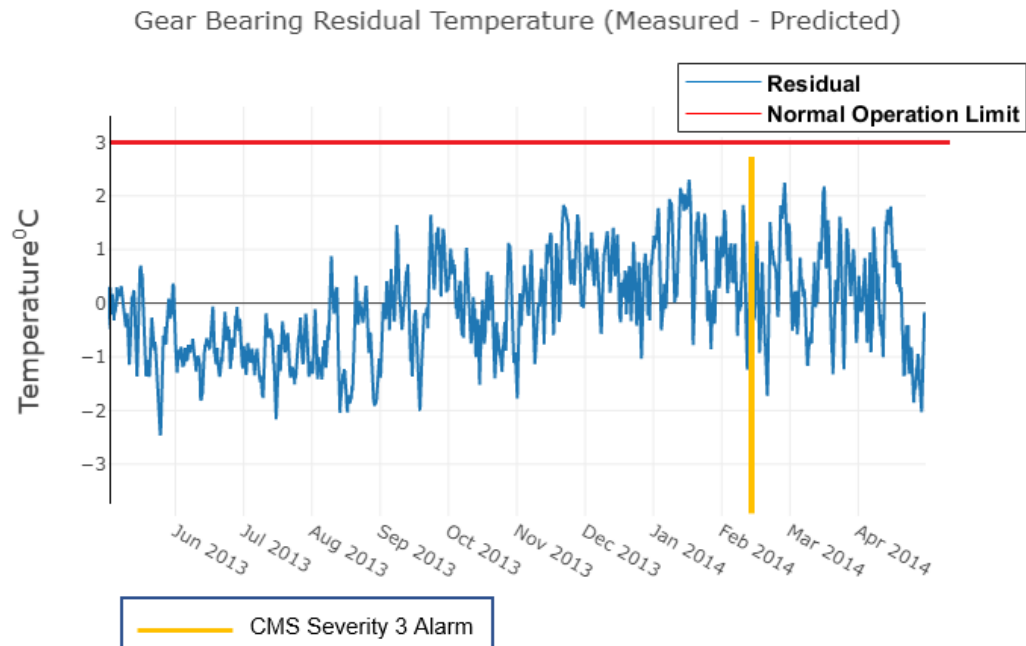


Figure 5.16: Gear bearing residual temperature for a faulty turbine.

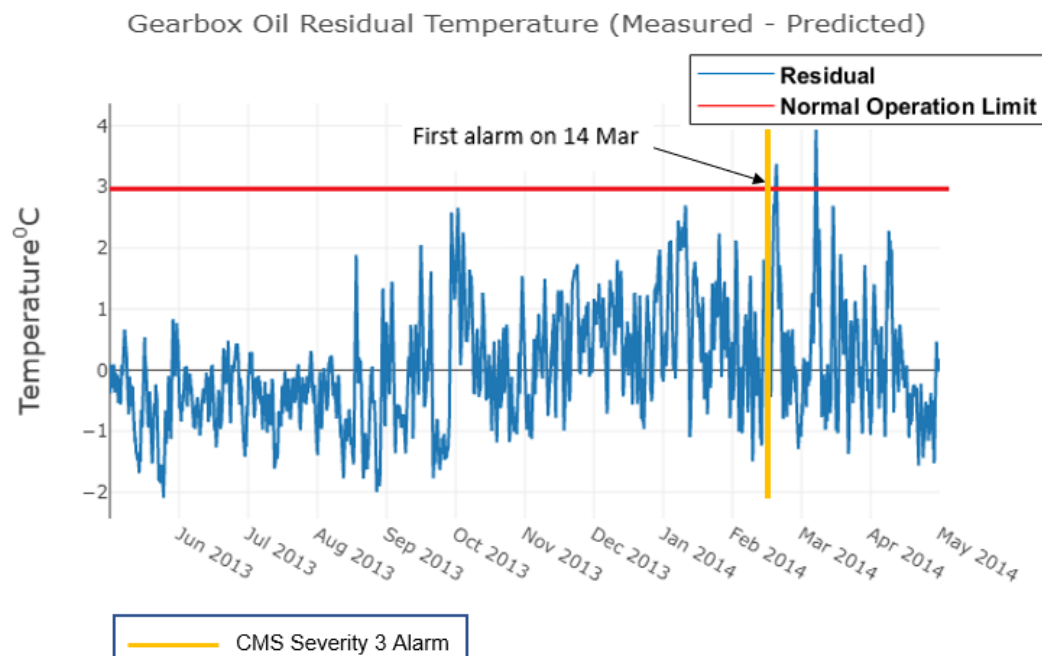


Figure 5.17: Gearbox oil residual temperature for a faulty turbine.

5.5.2. Turbine - 17

Turbine - 17 is an interesting case study where the real advantage of using a physics of failure model along with the ANN-based temperature model can be seen. There has never been any major component replacement in Turbine - 17 since the commencement of its operation. The ANN model is trained on data of the year 2010 and put for anomaly detection using data from the year 2012. The average residual output of the bearing and oil temperature models is shown in Figures 5.18 and 5.19 respectively. The ANN-based models give multiple alarms between January 2012 to April 2012 and again between October 2012 to December 2012.

During the period when the residual temperature is greater than $+3^{\circ}\text{C}$, there was multiple *low oil level* alerts recorded in the SCADA logs. Due to this alert, it is highly likely to have an increase in bearing temperature due to a lower amount of lubricant present for normal heat conduction. The turbine was serviced once in March, but the problem was not fixed. After the corrective maintenance intervention, the problem was not fixed as there were more alerts in the month of April as well. This can be seen in the residual temperature plots as well. Another service was carried out in the month of May. This intervention fixed the problem to a certain extent, and this can also be seen in the residual temperature plots as well where the residual is always below $+3^{\circ}\text{C}$. *Low oil level* alerts start to resurface again from the end of October 2012, and the residual temperature thus exceeds $+3^{\circ}\text{C}$ at the same time. Meanwhile, the physics of failure model in Figure 5.20 shows that turbine 17 is operating at low risk of failure as all the damage operating conditions lie below the 75th percentile.

This is one of the limitations of just utilizing temperature based ANN models for monitoring the condition of gearbox. As these models are just based on raw measurement of the temperature, it is highly likely to give false alarms, as in an alarm could be triggered due to a problem in the main oil system as observed in this case. However, from PoF model it is clear that the gearbox was under low risk of failure. Hence, from outputs of both the ANN and PoF model, it can be ascertained that there was a malfunction in a different component rather than in the gearbox.

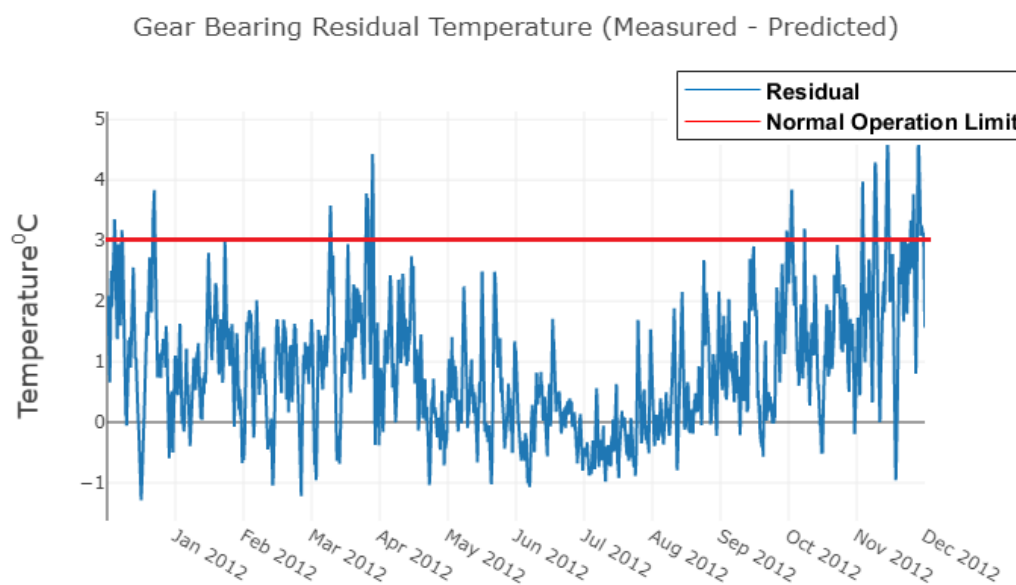


Figure 5.18: Gear bearing residual temperature for a faulty turbine.

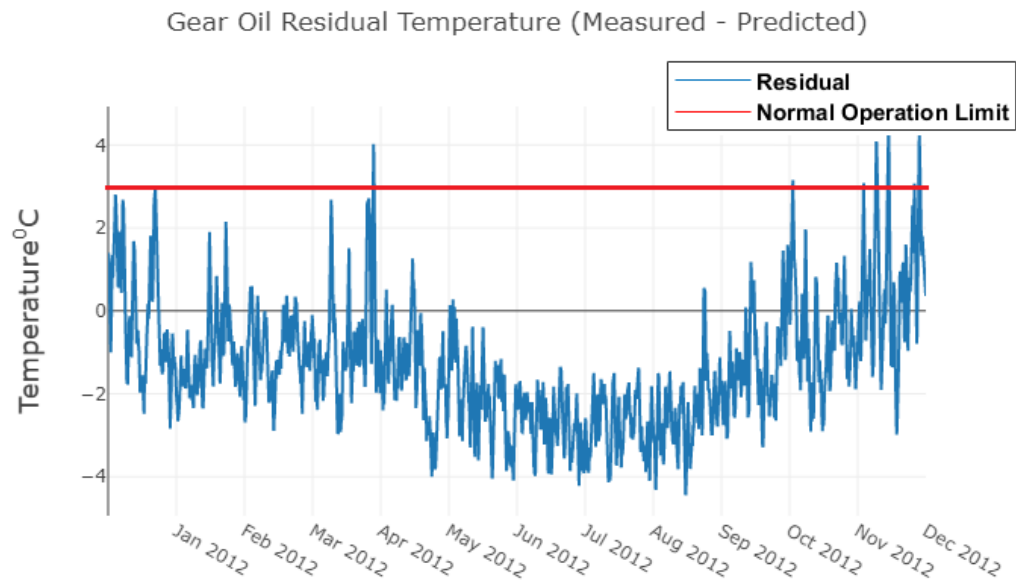


Figure 5.19: Gearbox oil residual temperature for a faulty turbine.

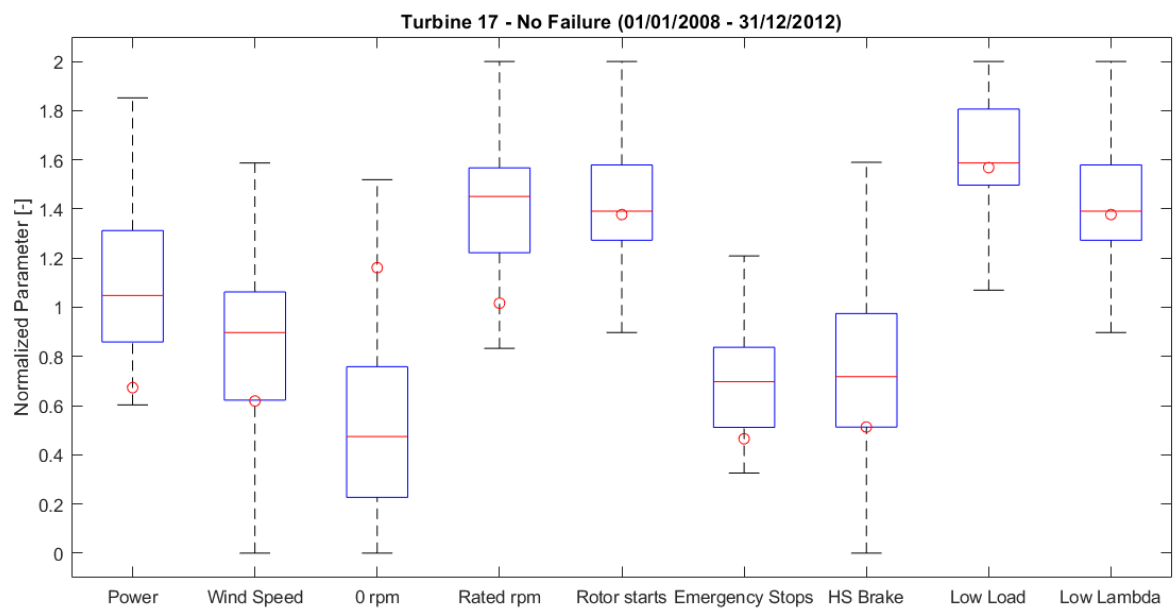


Figure 5.20: Turbine 17 - No Failure

Conclusions and Future Work

In this chapter, the main contributions and findings are discussed. Later, some recommendations for the future work is given. The main focus of the thesis was to investigate the use of operational SCADA data for effective condition monitoring of wind turbine gearbox. After literature study, two main goals were identified -

1. Investigate the correlation between operational data and observed failures on a wind farm level.
2. Develop algorithms for automatic fault detection in gearbox based on residual temperatures using recurrent neural networks and physics of failure techniques.

6.1. Using PoF on a wind farm level to improve condition monitoring and risk assessment

The first objective of the thesis was to devise a method of using historical operational data to assess the risk of failure in a turbine gearbox on a wind farm level. An extension to the *Physics of Failure* methodology was proposed to correlate the failures observed to the damage operating conditions using simple statistics. Two separate studies were conducted depending on the mode and location of failure. First a general approach to correlate failures in low speed and intermediate stage with damage operating conditions was carried out. Second, to correlate unexplained *White Etching Cracks* observed in high-speed stage bearings with newly proposed damage operating conditions taking into account load sharing of bearings in HSS, and lubrication parameters. The answer to the first research question as discussed in Section 2.9.

Can a correlation be deduced between White Etching Cracks and hypothetical failure theories using SCADA data and Physics of Failure technique?

Causes of White Etching Cracks based on SCADA Data: Based on the observations on a wind farm level, there was a strong correlation found between White Etching crack failures as observed on high speed shaft bearings and the following damage operating conditions¹ -

- Higher number of braking events at high wind speeds and rated power.
- Higher number of emergency brake applications.
- Operating at loads and lambda below minimum requisite for longer periods.
- Higher number of rotor start-stop cycles.

¹Failed turbine was compared against rest of the fleet consisting of 60 wind turbines.

It is observed that all the HSS bearings that failed due to WEC belonged in the higher risk region, however, **not** all the bearings in the high risk region failed. While statistical plots do limit the generalizability of the observations, this approach provides new insight into WEC failures based on field data.

Apart from WECs, the correlation between other modes of failure (pitting, fracture) observed in low speed & intermediate stages and damage operating conditions were also evaluated. It was found that there was no general or consistent link between failures and high risk region. Some identified damage operating conditions did have a moderate correlation and could help identify gear and bearing damage in low speed and intermediate stages - longer standstill periods which leads to micropitting, number of emergency brake events and number of rotor start-stop cycles. None of the failures observed in the gearbox were due to high cycle fatigue. Also, few damage operating conditions considered such as total yaw movement and variation of mean bearing temperature have no consistent correlation to any failures observed. It is important to note here that the study was limited to data from one wind farm. The exploration of condition monitoring approaches with comparisons within the wind farm showed that this approach might significantly improve the reliability of monitoring when used alongside with ANN-based monitoring of temperatures and vibration monitoring systems. This method can also help in root cause analysis as operational history of the failed component can be better understood.

6.2. RNN model for temperature based anomaly detection

The second objective of this thesis work was to develop a temperature-based anomaly detection model based on recurrent neural network. Firstly, the raw SCADA data was cleaned making use of 3 filters - General filter, Cluster filter, and missing data filter. A GRU based model with two hidden layers provided the least mean squared error loss (0.1464°C) compared to other models based on LSTM and NARX. The GRU model was trained on normal behavior data to predict gear bearing and gearbox oil temperature. An anomaly detection method utilizing the moving average over 1 day of the residual temperature measurement was presented. The ANN condition monitoring model was able to detect anomalies 3-6 months before the failure depending on the severity of the damage, providing ample opportunity for condition-based preventive maintenance planning. In one of the case study however, the identification of the abnormality was ambiguous. A case study (Turbine-17) is shown where the importance of utilizing both PoF and ANN based models to ascertain the condition of a gearbox. Finally, the second research question can be answered -

Is it possible to develop a robust condition monitoring technique for a wind turbine gearbox that utilizes both physics of failure model and machine learning techniques?

It was observed that making use of both PoF and RNN models for failure detection in gearboxes outperforms pure data-driven solutions. The number of true positives, i.e., the percentage of actual failures detected correctly is much higher when these two techniques are combined.

One of the limitation of this study is lack of edge cases where both the ANN and PoF fails to predict failures or outputs false positives.

6.3. Future Work

The future development of the concepts for condition monitoring presented in this thesis can be divided into two areas of application: physics of failure models and the ANN based models. A small section on future work for wind turbines operations team is also included. A few ideas for future work in each area are presented in the following sections -

6.3.1. Physics of failure model

The following items could be explored or implemented to improve the existing model.

- The load sharing and elastohydrodynamic lubricant parameters of the **low speed and intermediate stages** can be used for more robust condition monitoring of all gearbox components.
- Additional damage-inducing parameters like **turbulence intensity** and **number of grid loss events** could be included in PoF models to check if they have any correlation to gearbox failures.
- Ideally it would be preferable to count **all braking events** by classifying them under different power bins, since braking at higher power induces higher load and torque reversals on the gearbox.
- **Sensitivity analysis:** In this thesis, only a correlation between the damage operating conditions and observed gearbox failures is presented. Carrying out a sensitivity study of these damage operating conditions and observed failures would be more insightful for the operators and can be used to accurately predict failures.
- **Hybrid prognosis:** A hybrid prognosis framework can be developed where an ANN model is used for diagnosis and a physics of failure model could be used for estimating the remaining useful life of the component under study. The remaining useful lifetime of the gearbox components can be estimated with the availability of high-frequency SCADA data, which can aid in the calculation of loads during non-transient events like braking, sudden-gusts, and so on.

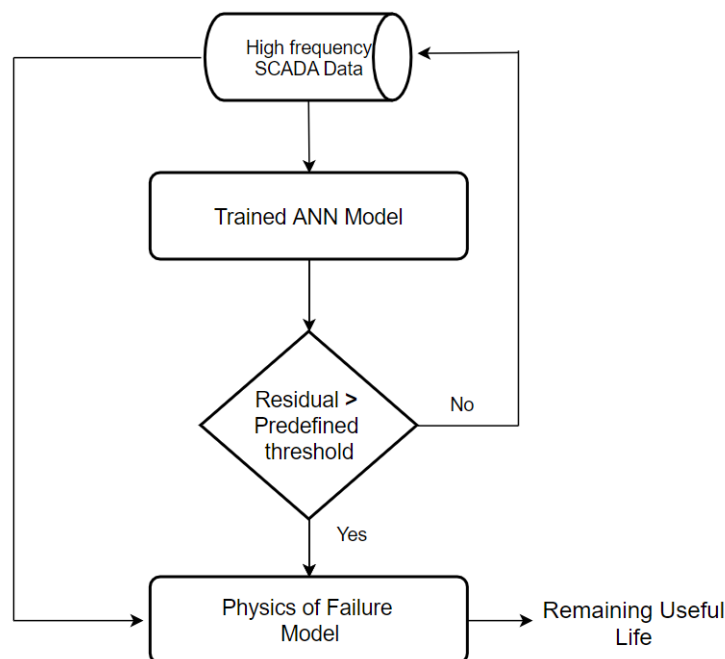


Figure 6.1: Framework for hybrid prognosis.

6.3.2. ANN based condition monitoring

- **Bayesian hyperparameter optimization** could be explored for searching more efficient and better hyperparameters to improve the performance of the neural network.

- **Distances measures** between the ANN predicted and actual temperatures, such as Euclidean, Minkowski, Mahalanobis could be further used to improve failure detection capabilities as discussed by Jannis Tautz-Weinert [44].
- **Autoencoders:** The main limitation of the present models is that different networks have to be developed to monitor different temperatures in the turbine. Autoencoders are a type of ANN that “attempts” to reconstruct its inputs. The reconstruction error can be used to flag anomalies. All the temperatures signals could be monitored with just one model.

6.3.3. Engineering recommendations

In this section, some recommendations which are valuable for wind farm owners such as Eneco is presented. A real-time operational tool can be developed for the use of wind turbine operators. For anomaly detection, a machine learning pipeline could be created with the proposed model, where the real-time SCADA data is used to ascertain the condition of wind turbines. Alongside, physics of failure models could run in the background. Most of the operators on-field might not be able to understand the complexities of the proposed model. Hence, it would be more useful to make attractive and simple dashboards, where the turbines at high risk of failure could be updated regularly based on the combined outputs of ANN and PoF based models.

Bibliography

- [1] World wind energy association, wind power capacity worldwide reaches 597 gw, 50.1 gw added in 2018. [Online]. Available: <https://wwindea.org/blog/2019/02/25/wind-power-capacity-worldwide-reaches-600-gw-539-gw-added-in-2018/>
- [2] C. Röckmann, S. Lagerveld, and J. Stavenhuet, *Operation and Maintenance Costs of Offshore Wind Farms and Potential Multi-use Platforms in the Dutch North Sea*, 04 2017, pp. 97–113.
- [3] M. D. Reder, E. Gonzalez, and J. J. Melero, “Wind turbine failures-tackling current problems in failure data analysis,” in *Journal of Physics: Conference Series*, vol. 753, no. 7. IOP Publishing, 2016, p. 072027.
- [4] S. Pfaffel, S. Faulstich, and K. Rohrig, “Performance and reliability of wind turbines: A review,” *Energies*, vol. 10, no. 11, 2017. [Online]. Available: <https://www.mdpi.com/1996-1073/10/11/1904>
- [5] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, “Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges,” *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014. [Online]. Available: <https://www.mdpi.com/1996-1073/7/4/2595>
- [6] D. An, N. H. Kim, and J. H. Choi, “Practical options for selecting data-driven or physics-based prognostics algorithms with reviews,” *Rel. Eng. & Sys. Safety*, vol. 133, pp. 223–236, 2015.
- [7] N. M. Vichare and M. G. Pecht, “Prognostics and health management of electronics,” *IEEE Transactions on Components and Packaging Technologies*, vol. 29, no. 1, pp. 222–229, 2006.
- [8] E. J. Alvarez and A. P. Ribaric, “An improved-accuracy method for fatigue load analysis of wind turbine gearbox based on scada,” *Renewable Energy*, vol. 115, p. 391–399, 2018.
- [9] T. Wang, “Trajectory similarity based prediction for remaining useful life estimation,” *PhD Thesis, University of Cincinnati*, 2010.
- [10] C. S. Gray and S. J. Watson, “Physics of failure approach to wind turbine condition based maintenance,” *Wind Energy*, vol. 13, no. 5, pp. 395–405, 2010. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.360>
- [11] S. Watson, I. Kennedy, and C. S. Gray, “The use of physics of failure modelling in wind turbine condition monitoring,” *EWEA Annual Conference*, pp. 309–312, 2011.
- [12] Y. Qiu, L. Chen, Y. Feng, and Y. Xu, “An approach of quantifying gear fatigue life for wind turbine gearboxes using supervisory control and data acquisition data,” *Energies*, vol. 10, no. 8, 2017. [Online]. Available: <http://www.mdpi.com/1996-1073/10/8/1084>
- [13] P. Cambron, A. Tahan, C. Masson, and F. Pelletier, “Bearing temperature monitoring of a wind turbine using physics-based model,” *Journal of Quality in Maintenance Engineering*, vol. 23, no. 4, pp. 479–488, 2017. [Online]. Available: <https://doi.org/10.1108/JQME-06-2016-0028>
- [14] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, “Monitoring wind turbine gearboxes,” *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1521>

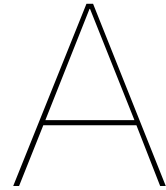
- [15] M. Wilkinson, B. Darnell, T. V. Delft, and K. Harman, "Comparison of methods for wind turbine condition monitoring with scada data," *IET Renewable Power Generation*, vol. 8, no. 4, pp. 390–397, May 2014.
- [16] D. Breteler, C. Kaidis, and R. Loendersloot, "Physics based methodology for wind turbine failure detection , diagnostics & prognostics," *EWEA Annual Conference*, 2015.
- [17] I. Al-Tubi, H. Long, P. Tavner, B. Shaw, and J. Zhang, "Probabilistic analysis of gear flank micro-pitting risk in wind turbine gearbox using supervisory control and data acquisition data," *IET Renewable Power Generation*, vol. 9, no. 6, pp. 610–617, 2015.
- [18] K. Javed, "A robust & reliable Data-driven prognostics approach based on extreme learning machine and fuzzy clustering." Theses, Université de Franche-Comté, Apr. 2014. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-01025295>
- [19] A. Karpatne, W. Watkins, J. S. Read, and V. Kumar, "Physics-guided neural networks (PGNN): an application in lake temperature modeling," *CoRR*, vol. abs/1710.11431, 2017. [Online]. Available: <http://arxiv.org/abs/1710.11431>
- [20] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1996.
- [21] H. Pokharna. For dummies — the introduction to neural networks we all need ! (part 1). Accessed: 2019-07-15. [Online]. Available: <https://medium.com/technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb>
- [22] S. Sharma. Activation functions in neural networks. Accessed: 2019-07-15. [Online]. Available: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
- [23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2017.
- [24] E. Cadenas, W. Rivera, R. Campos-Amezcuca, and C. Heard, "Wind speed prediction using a univariate arima model and a multivariate narx model," *Energies*, vol. 9, no. 2, 2016. [Online]. Available: <http://www.mdpi.com/1996-1073/9/2/109>
- [25] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, and N. Mrabet Bellaaj, "A nonlinear autoregressive exogenous (narx) neural network model for the prediction of the daily direct solar radiation," *Energies*, vol. 11, no. 3, 2018. [Online]. Available: <http://www.mdpi.com/1996-1073/11/3/620>
- [26] Y. LeCun, Y. Bengio, and G. Hinton, *Deep Learning*. New York, NY, USA: Nature, 2015, vol. 521. [Online]. Available: <http://dx.doi.org/10.1038/nature14539>
- [27] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct 1990.
- [28] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, March 1994.
- [29] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13. JMLR.org, 2013, pp. III–1310–III–1318. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3042817.3043083>
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>

- [31] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *CoRR*, vol. abs/1503.04069, 2015. [Online]. Available: <http://arxiv.org/abs/1503.04069>
- [32] C. Olah. Understanding lstm networks. Accessed: 2019-04-15. [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [33] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *CoRR*, vol. abs/1409.1259, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1259>
- [34] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, "Simap: Intelligent system for predictive maintenance: Application to the health condition monitoring of a windturbine gearbox," *Computers in Industry*, vol. 57, pp. 552–568, 2006.
- [35] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated scada data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.319>
- [36] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renewable Energy*, vol. 48, no. C, pp. 110–116, 2012. [Online]. Available: <https://ideas.repec.org/a/eee/renene/v48y2012icp110-116.html>
- [37] Z.-Y. Zhang and K.-S. Wang, "Wind turbine fault detection based on scada data analysis using ann," *Advances in Manufacturing*, vol. 2, no. 1, pp. 70–78, Mar 2014. [Online]. Available: <https://doi.org/10.1007/s40436-014-0061-6>
- [38] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, pp. 980–987, 2015.
- [39] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox," *Wind Energy*, vol. 20, no. 8, pp. 1421–1438, 2017. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2102>
- [40] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *ESANN*, 2015.
- [41] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using recurrent neural networks (rnn)," in *2016 Integrated Communications Navigation and Surveillance (ICNS)*, April 2016, pp. 5C2–1–5C2–8.
- [42] J. Fu, J. Chu, P. Guo, and C. Zhenyu, "Condition monitoring of wind turbine gearbox bearing based on deep learning model," *IEEE Access*, vol. PP, pp. 1–1, 04 2019.
- [43] G. Helbing and M. Ritter, "Deep Learning for fault detection in wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 98, no. C, pp. 189–198, 2018. [Online]. Available: <https://ideas.repec.org/a/eee/rensus/v98y2018icp189-198.html>
- [44] J. Tautz-Weinert, "Improved wind turbine monitoring using operational data," Ph.D. dissertation, Loughborough University, 2018. [Online]. Available: <https://dspace.lboro.ac.uk/2134/36199>
- [45] E. Gonzalez, J. Tautz-Weinert, J. J. Melero, and S. J. Watson, "Statistical evaluation of SCADA data for wind turbine condition monitoring and farm assessment," *Journal of Physics: Conference Series*, vol. 1037, p. 032038, June 2018. [Online]. Available: <https://doi.org/10.1088%2F1742-6596%2F1037%2F3%2F032038>
- [46] S. Shanbr, F. Elasha, M. Elforjani, and J. C. Teixeira, "Bearing fault detection within wind turbine gearbox," *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, pp. 565–570, 2017.

- [47] A. R. Nejad, Z. Gao, and T. Moan, "Fatigue reliability-based inspection and maintenance planning of gearbox components in wind turbine drivetrains," *Energy Procedia*, vol. 53, p. 248–257, 12 2014.
- [48] S. Sheng, M. McDade, and R. Errichello, "Wind turbine gearbox failure modes - a brief (presentation)," 04 2019, accessed: 2019-04-29. [Online]. Available: https://www.researchgate.net/publication/255248159_Wind_Turbine_Gearbox_Failure_Modes_-_A_Brief_Presentation
- [49] A. Ruellan Du Crehu, "Tribological analysis of White Etching Crack (WEC) failures in rolling element bearings," Theses, INSA de Lyon, Dec. 2014. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-01153235>
- [50] "ISO 15243:2017 Rolling bearings – Damage and failures – Terms, characteristics and causes," International Organization for Standardization, Standard, 2017. [Online]. Available: <https://www.iso.org/standard/59619.html>
- [51] SKFWind. Bearing the burden- the leading causes of wind turbine bearing failures. Accessed: 2019-04-30. [Online]. Available: <http://windfarmmanagement.skf.com/bearing-the-burden/>
- [52] "Bearing failure modes," NSK Bearing Ltd, Report, 2017, accessed: 2019-04-30. [Online]. Available: <https://www.nskeurope.com/en/services/troubleshooting/damage-by-type/flaking.html>
- [53] Y. Guo and J. Keller, "Investigation of high-speed shaft bearing loads in wind turbine gearboxes through dynamometer testing," *Wind Energy*, vol. 21, no. 2, pp. 139–150, 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2150>
- [54] Y. Guo, J. Keller, L. Sethuraman, and B. McNiff, "High-speed shaft bearing loads testing and modeling in the nrel gearbox reliability collaborative," 03 2015.
- [55] SKF, "Calculating the radial load acting on matched bearings," accessed: 2019-04-30. [Online]. Available: <https://www.skf.com/uk/products/bearings-units-housings/roller-bearings/tapered-roller-bearings/matched-tapered-roller-bearings/loads/calculating-radial-load-on-matched-bearings/index.html>
- [56] "ISO 281:2007 Rolling bearings – Dynamic load ratings and rating life," International Organization for Standardization, Standard, 2007. [Online]. Available: <https://www.iso.org/standard/38102.html>
- [57] "ASTM D341 - 17 Standard Practice for Viscosity-Temperature Charts for Liquid Petroleum Products," American Society for Testing and Materials, Standard, 2017. [Online]. Available: <https://www.astm.org/Standards/D341.htm>
- [58] SKF, "Life modification factor, askf," accessed: 2019-04-30. [Online]. Available: <https://www.skf.com/group/products/bearings-units-housings/principles/bearing-selection-process/bearing-size/size-selection-based-on-rating-life/life-modification-factor/index.html>
- [59] M.-H. Evans, "An updated review: white etching cracks (wecs) and axial cracks in wind turbine gearbox bearings," *Materials Science and Technology*, vol. 32, no. 11, pp. 1133–1169, 2016.
- [60] P. Lynwander, *Gear Drive Systems: Design and Application*, 01 2019.
- [61] J. R. Barber, *Hertzian Contact*. Cham: Springer International Publishing, 2018, pp. 29–41.
- [62] J. Gegner, "Tribological aspects of rolling bearing failures," in *Tribology-lubricants and lubrication*. IntechOpen, 2011.

- [63] "Rolling Bearing Lubrication," Schaeffler, FAG Kugelfischer Georg Schäfer AG, 2002, accessed: 2019-04-30.
- [64] D. Dowson, "Elastohydrodynamic and micro-elastohydrodynamic lubrication," *Wear*, vol. 190, no. 2, pp. 125 – 138, 1995.
- [65] "ISO/TR 1281-2:2008 Rolling bearings – Explanatory notes on ISO 281 – Part 2: Modified rating life calculation, based on a systems approach to fatigue stresses, url = <https://www.iso.org/standard/40622.html>," International Organization for Standardization, Standard, 2008. [Online]. Available: <https://www.iso.org/standard/40622.html>
- [66] K. Baalman, "Gleichung für die sollviskosität nach din iso 281." *Tribologie und Schmierungstechnik*, 1994.
- [67] D. Gonçalves, A. Vieira, A. Carneiro, A. V. Campos, and J. H. O. Seabra, "Film thickness and friction relationship in grease lubricated rough contacts," *Lubricants*, vol. 5, no. 3, 2017. [Online]. Available: <http://www.mdpi.com/2075-4442/5/3/34>
- [68] SKF, "Bearing friction, power loss and starting torque," accessed: 2019-05-03. [Online]. Available: <https://www.skf.com/group/products/bearings-units-housings/principles/bearing-selection-process/operating-temperature-and-speed/friction-powerloss-startingtorque/index.html>
- [69] K. Stadler and A. Stubenrauch, "Premature bearing failures in wind gear-boxes and white etching cracks," *SKF business and technology magazine (evolution.skf.com)*, 2013. [Online]. Available: <http://evolution.skf.com/premature-bearing-failures-in-wind-gearboxes-and-white-etching-cracks-wec/>
- [70] A. Ruellan, J. Cavoret, F. Ville, X. Kleber, and B. Liard, "Understanding white etching cracks in rolling element bearings: State of art and multiple driver transposition on a twin-disc machine," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 231, no. 2, pp. 203–220, 2017. [Online]. Available: <https://doi.org/10.1177/1350650116648058>
- [71] J. Lai and K. Stadler, "Investigation on the mechanisms of white etching crack (wec) formation in rolling contact fatigue and identification of a root cause for bearing premature failure," *Wear*, vol. 364–365, pp. 244–256, 10 2016.
- [72] M.-H. Evans, A. Richardson, L. Wang, and R. Wood, "Serial sectioning investigation of butterfly and white etching crack (wec) formation in wind turbine gearbox bearings," *Wear*, vol. 302, no. 1–2, pp. 1573–1582, 2013.
- [73] A. M. Diederichs, S. Barteldes, A. Schwedt, J. Mayer, and W. Holweger, "Study of sub-surface initiation mechanism for white etching crack formation," *Materials Science and Technology*, vol. 32, no. 11, pp. 1170–1178, 2016.
- [74] T. Bruce, H. Long, and R. Dwyer-Joyce, "Threshold maps for inclusion-initiated micro-cracks and white etching areas in bearing steel: the role of impact loading and surface sliding," *Tribology Letters*, vol. 66, no. 3, p. 111, 2018.
- [75] S. Ooi, A. Gola, R. Vegter, P. Yan, and K. Stadler, "Evolution of white-etching cracks and associated microstructural alterations during bearing tests," *Materials Science and Technology*, vol. 33, no. 14, pp. 1657–1666, 2017.
- [76] F. Chollet, *Deep Learning with Python*, 1st ed. Greenwich, CT, USA: Manning Publications Co., 2017.
- [77] Z. Cui, R. Ke, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," *CoRR*, vol. abs/1801.02143, 2018. [Online]. Available: <http://arxiv.org/abs/1801.02143>

- [78] F. Chollet *et al.*, “Keras,” <https://keras.io>, 2015.
- [79] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” vol. 15, pp. 315–323, 11–13 Apr 2011. [Online]. Available: <http://proceedings.mlr.press/v15/glorot11a.html>
- [80] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.
- [81] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2015.
- [82] DNV-GL, “Assess and improve wind farm profitability using windgemini,” accessed: 2019-07-15. [Online]. Available: https://www.dnvgl.com/energy/events/registration/assess-and-improve-wind-farm-profitability.html?utm_campaign=EN_ADV_NEMEA_19Q1_WEBI_WIND_WindGEMINI_webinar_drop3_thankyou&utm_medium=email&utm_source=Eloqua&elqTrackId=98dc7479aa1b455ab4ab3d4c31b6e2d1&elq=800416452e4b4cf79d4b35e3c055f187&elqaid=10678&elqat=1&elqCampaignId=6342



Bearing Design and Lubricant Properties

The high speed shaft is supported by 3 bearings. A cylindrical roller bearing and a pair of tapered roller bearings. The mechanical properties of the bearings are tabulated in Table A.1. Similarly, the typical properties of the lubricant used in the gearboxes is tabulated in Table A.2.

Table A.1: Bearing design and mechanical properties.

Bearing	Cylindrical Roller Bearing	Upwind Tapered Roller Bearing	Downwind Tapered Roller Bearing
Dynamic load rating [kN]	1060	455	782
Static load rating [kN]	1250	560	1140
Fatigue load limit [kN]	137	57	112
Inner diameter [mm]	130	150	150
Outer diameter [mm]	280	270	270
Width [mm]	93	49	77

Table A.2: Lubricant properties.

Lubricant	Mobilgear SHC XMP Series 320
ISO Viscosity Grade	320
Kinematic viscosity at 40°C	335 cSt
Kinematic viscosity at 100°C	38.3 cSt
Viscosity Index, ASTM D 2270	164
Specific Gravity @15.6°C kg/l, ASTM D 4052	0.86
Flash Point	242 °C

B

Supporting Plots

B.1. Input/Output parameters for RNN models

All the input and output parameters used for anomaly detection in Chapter 5 is included in this section. It includes all SCADA data one year prior to replacement. Also, CMS vibration analysis report prior to component failure is also presented.

B.1.1. Turbine 48

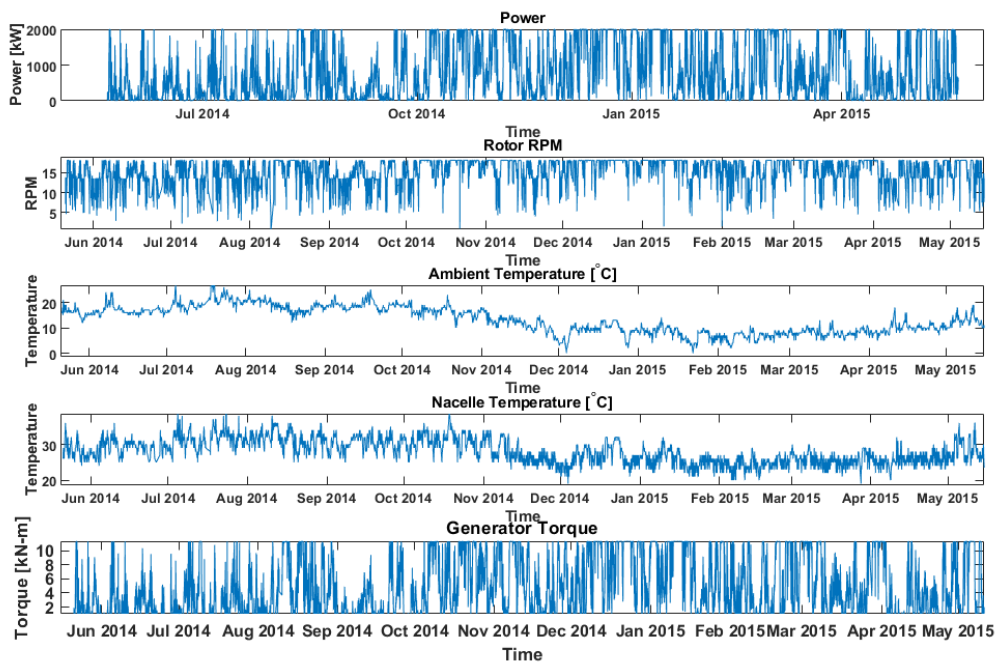


Figure B.1: Input Parameters for Turbine 48.

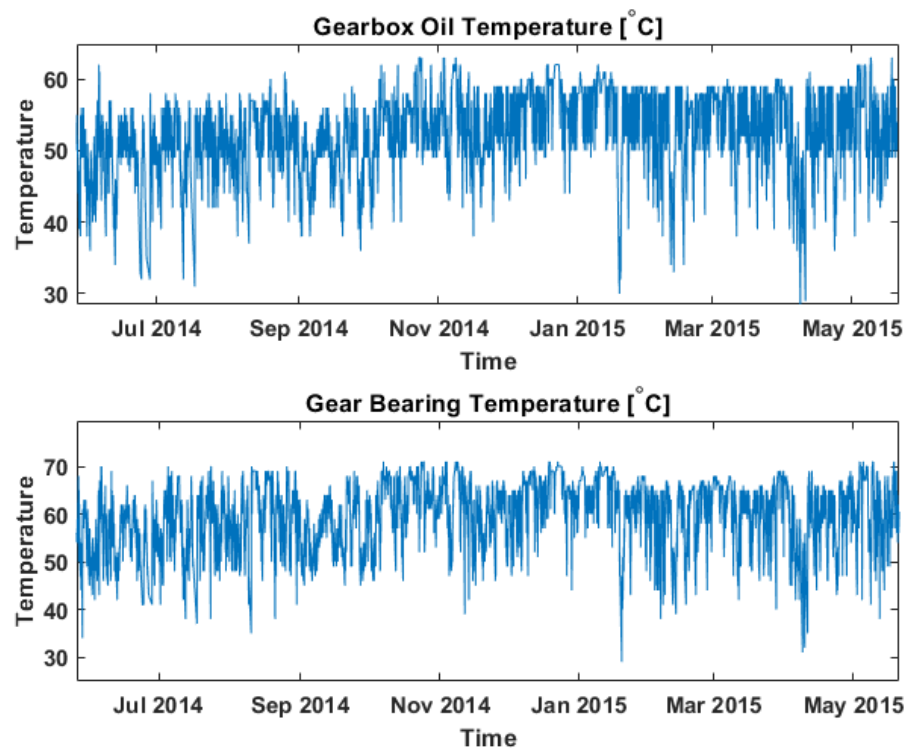


Figure B.2: Output Parameters for Turbine 48.

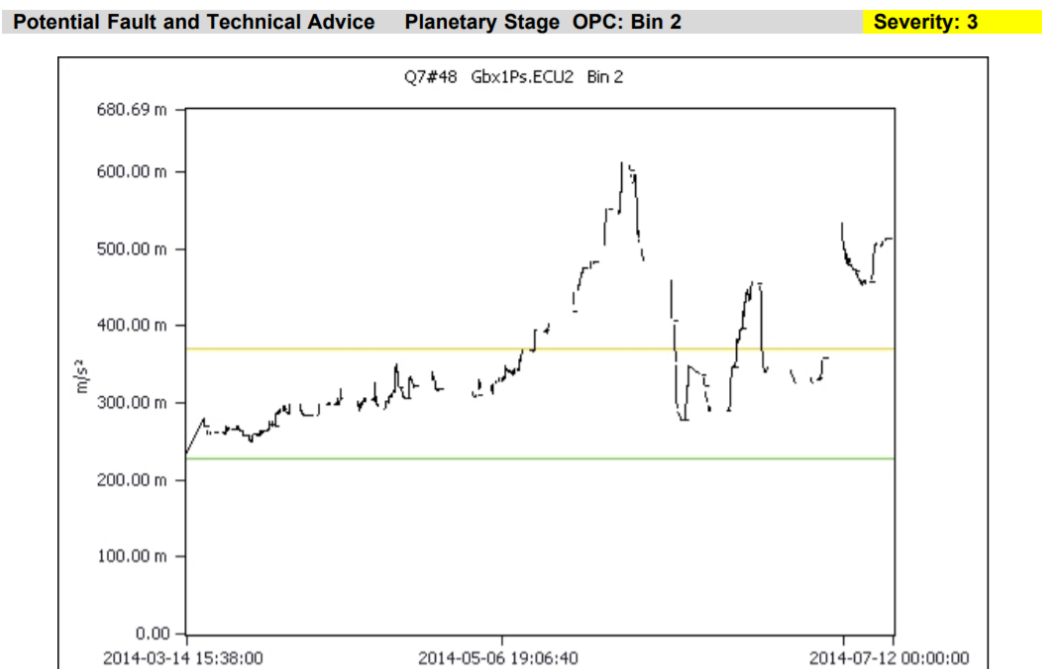


Figure B.3: Vibration analysis of planetary stage gearbox before replacement of Turbine 48.

B.1.2. Turbine 23

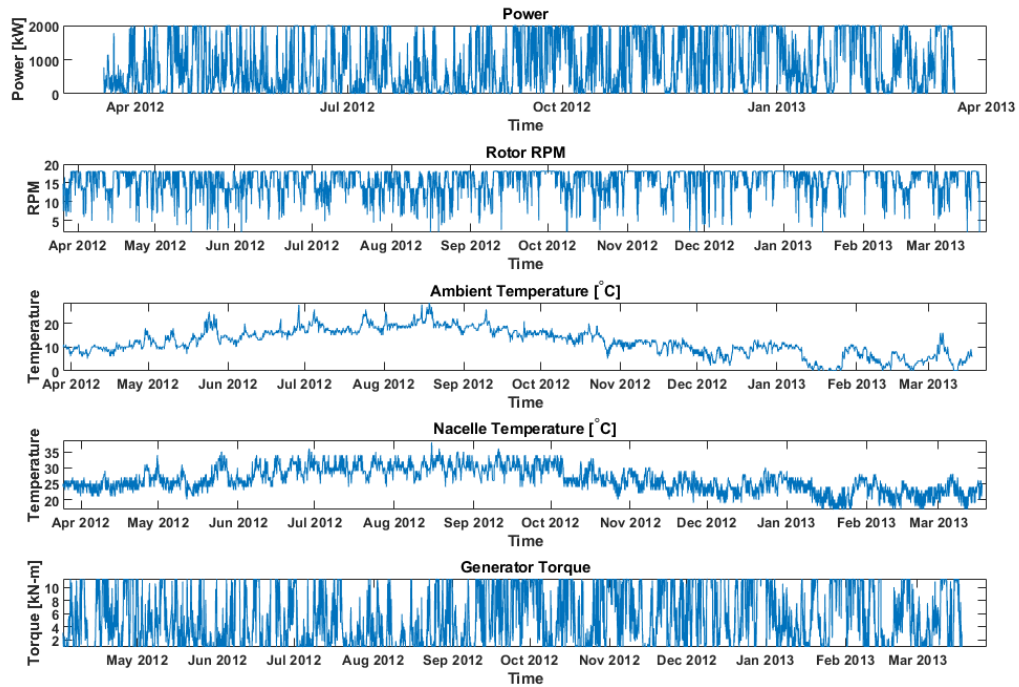


Figure B.4: Input Parameters for Turbine 23.

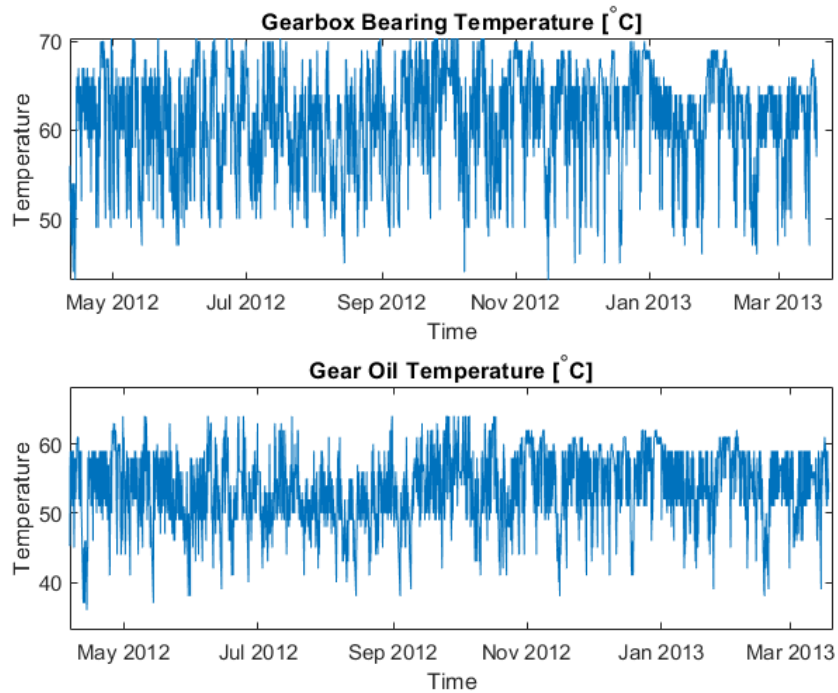


Figure B.5: Output Parameters for Turbine 23.

Potential Fault and Technical Advice High Speed Stage (HSS) Front Bearing OPC: 1/2/3/4/5 Severity: 2

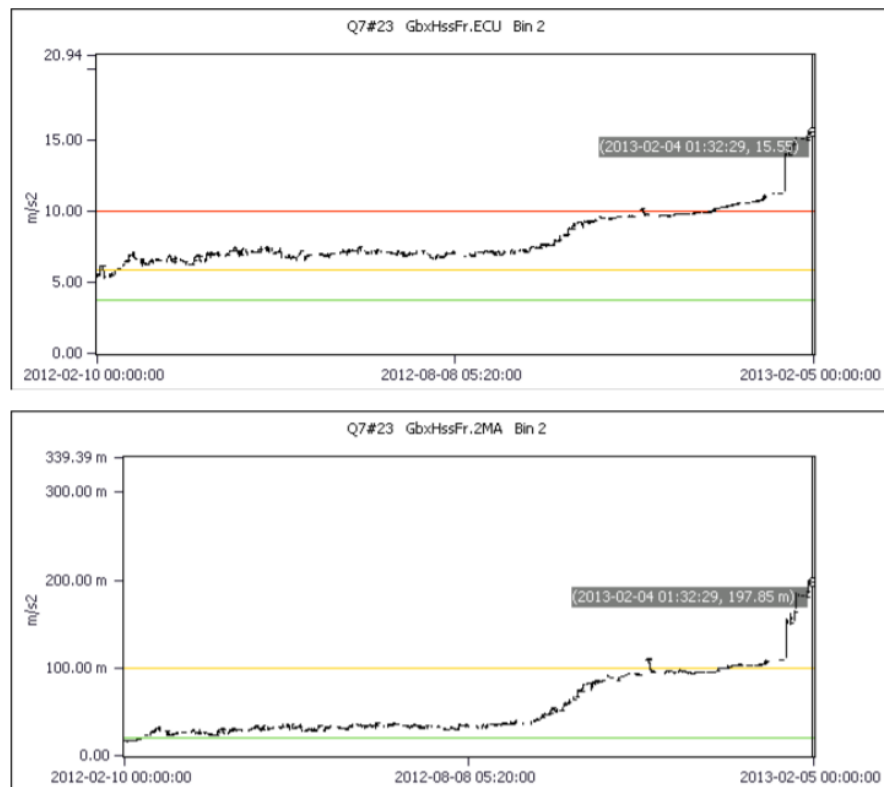


Figure B.6: Vibration analysis of high speed bearing before replacement - Turbine 23.

B.1.3. Turbine 32

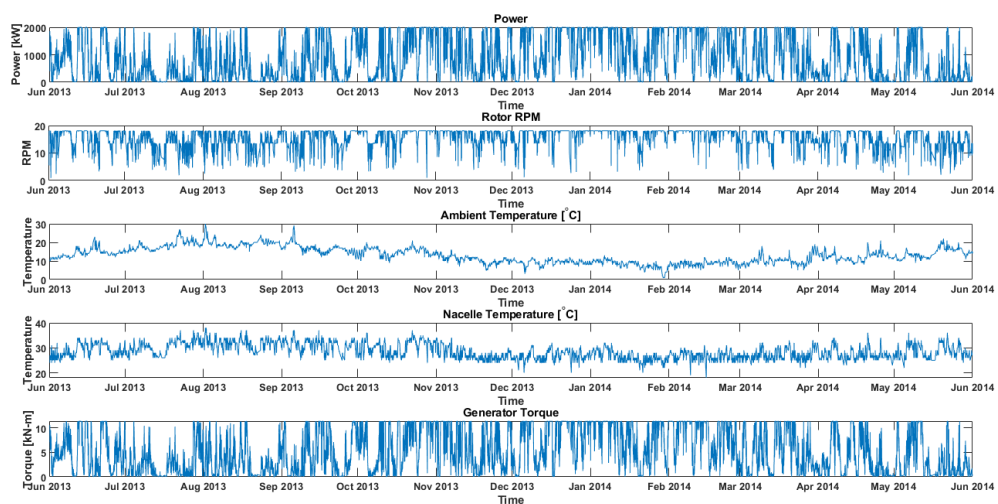


Figure B.7: Input Parameters for Turbine 32.

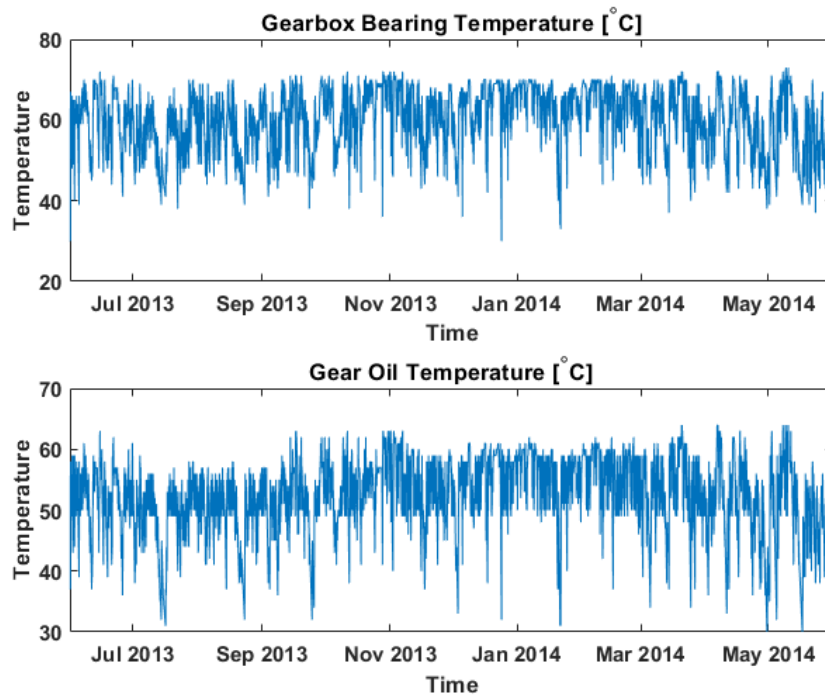


Figure B.8: Output Parameters for Turbine 32.

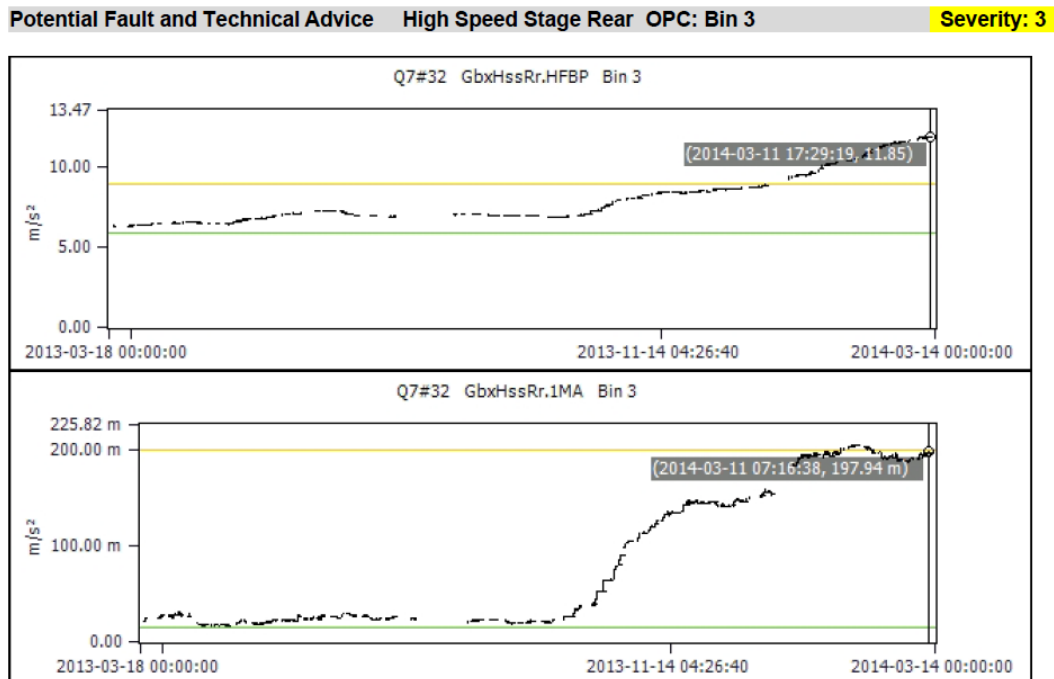


Figure B.9: Vibration analysis of high speed bearing before replacement - Turbine 32.

B.1.4. Turbine 7

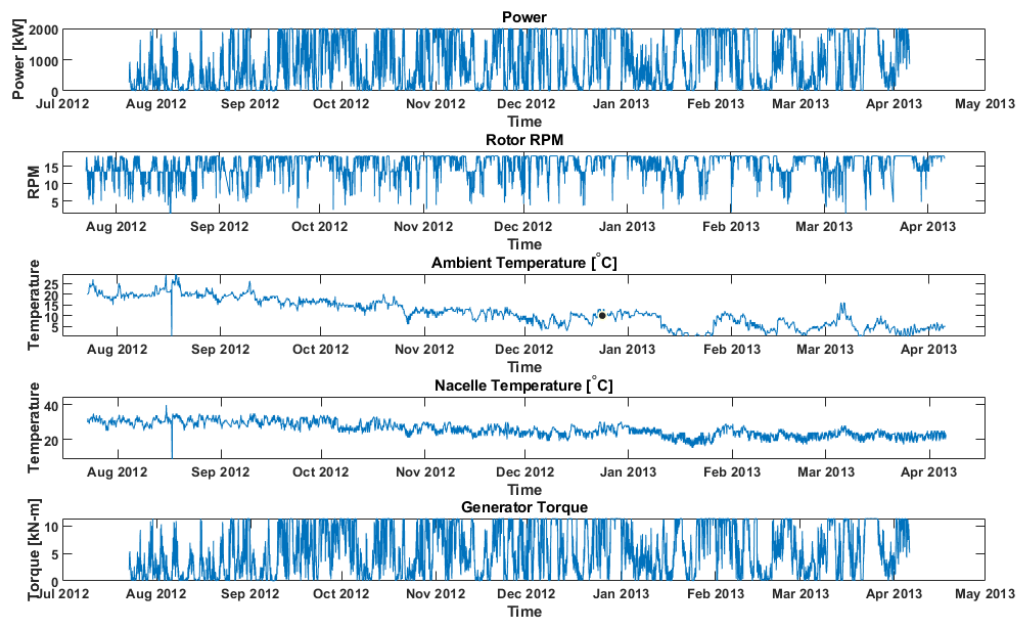


Figure B.10: Input Parameters for Turbine 7.

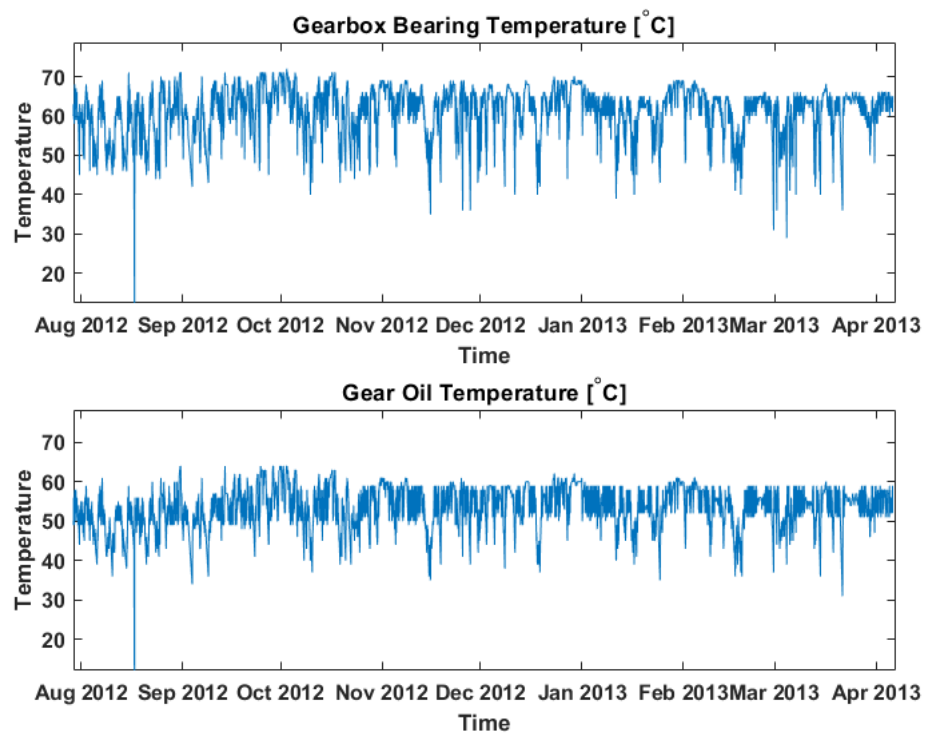


Figure B.11: Output Parameters for Turbine 7.

Potential Fault and Technical Advice High Speed Stage Rear OPC: 1/2/3/4/5 Severity: 2

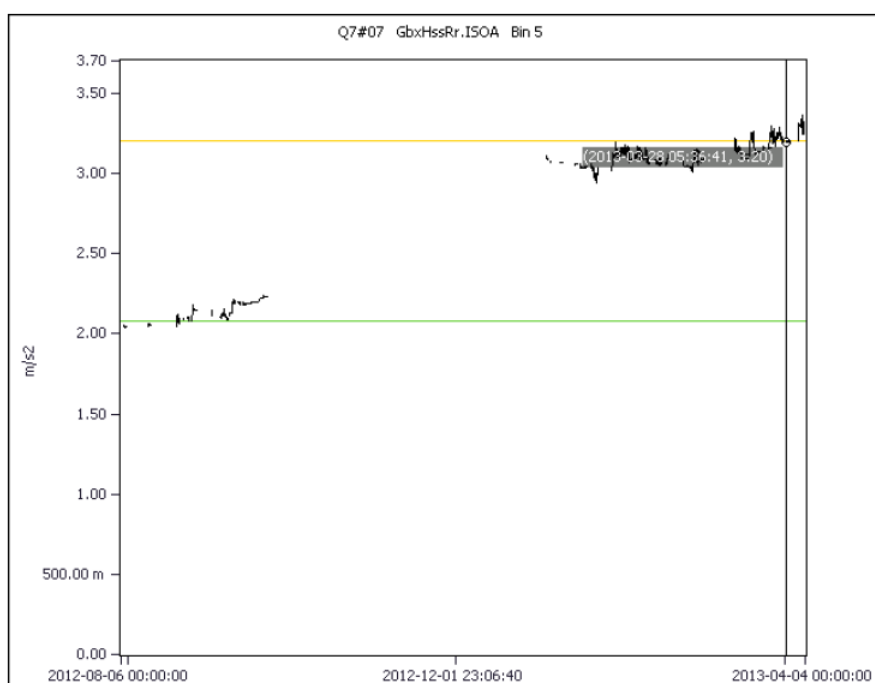


Figure B.12: Vibration analysis of high speed bearing before replacement of Turbine 7.

B.2. ANN Model Predictions

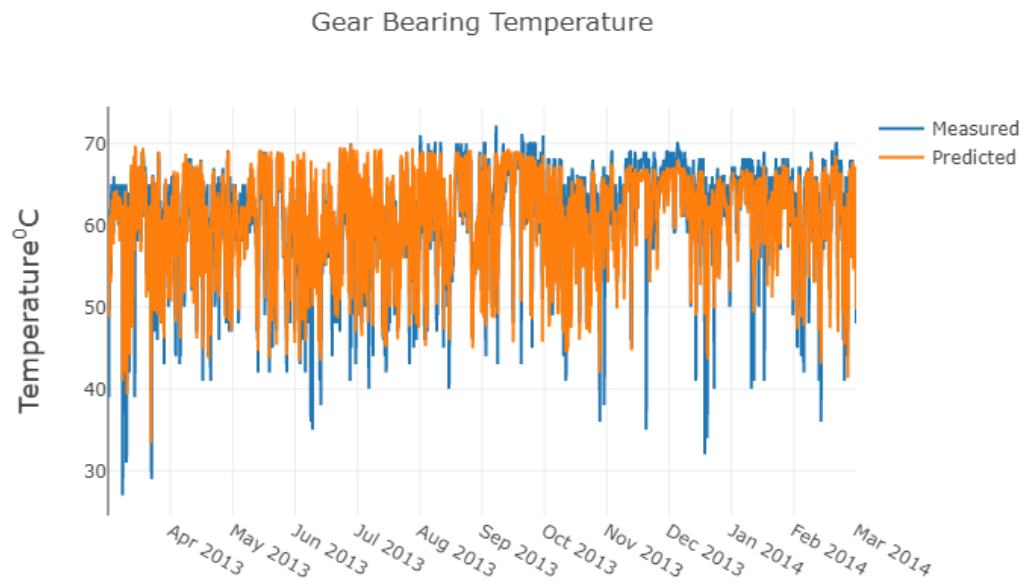


Figure B.13: Normal behaviour predictions of gear bearing temperature.

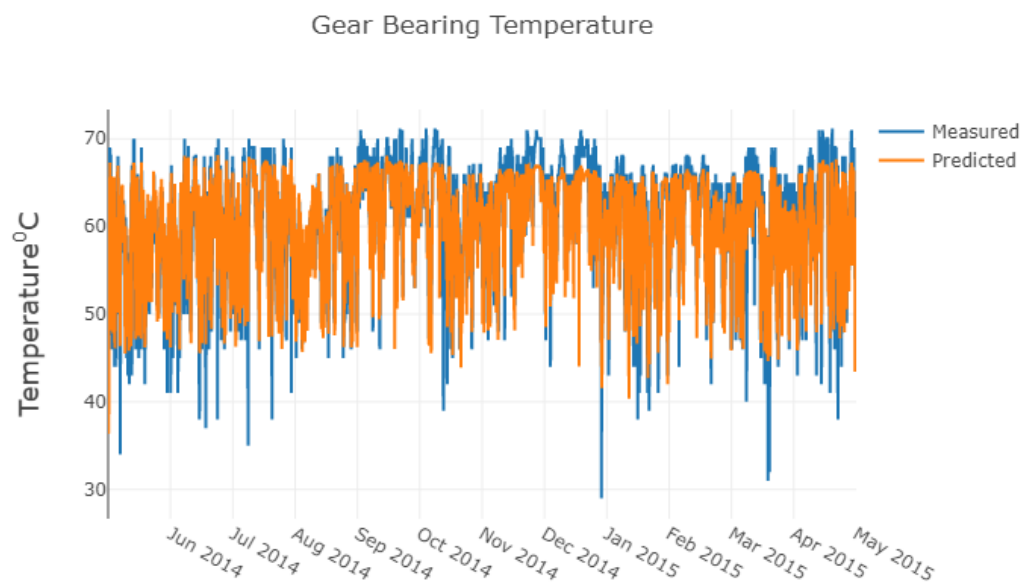


Figure B.14: Anomalous behaviour of gear bearing temperature.

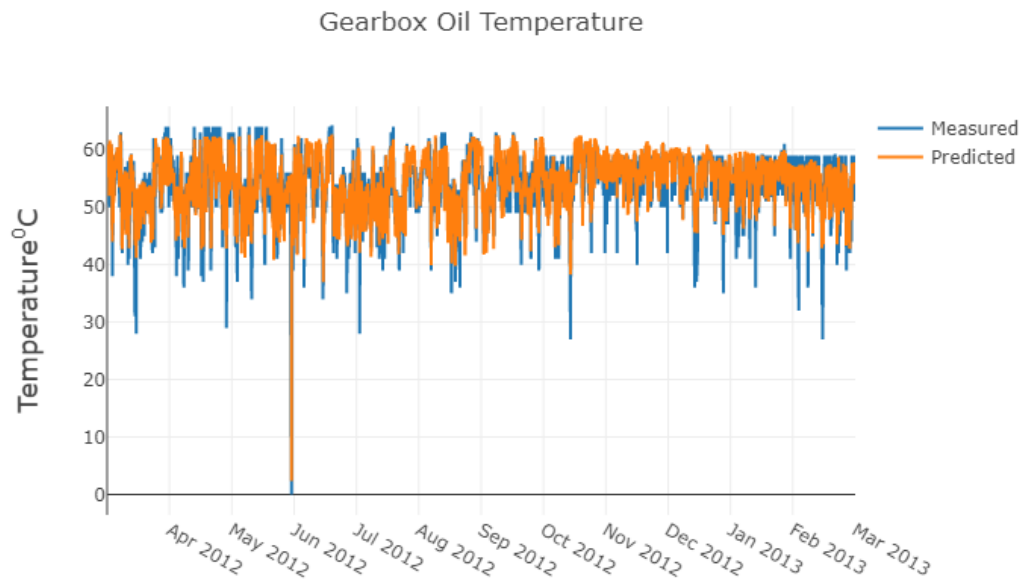


Figure B.15: Normal behaviour predictions of gearbox oil temperature.

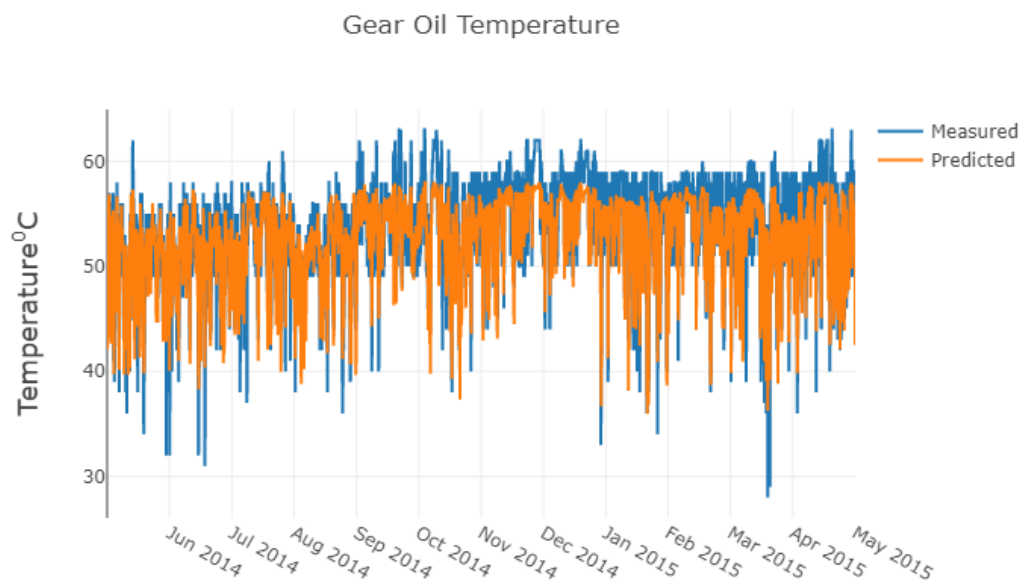


Figure B.16: Anomalous behaviour of gearbox oil temperature.