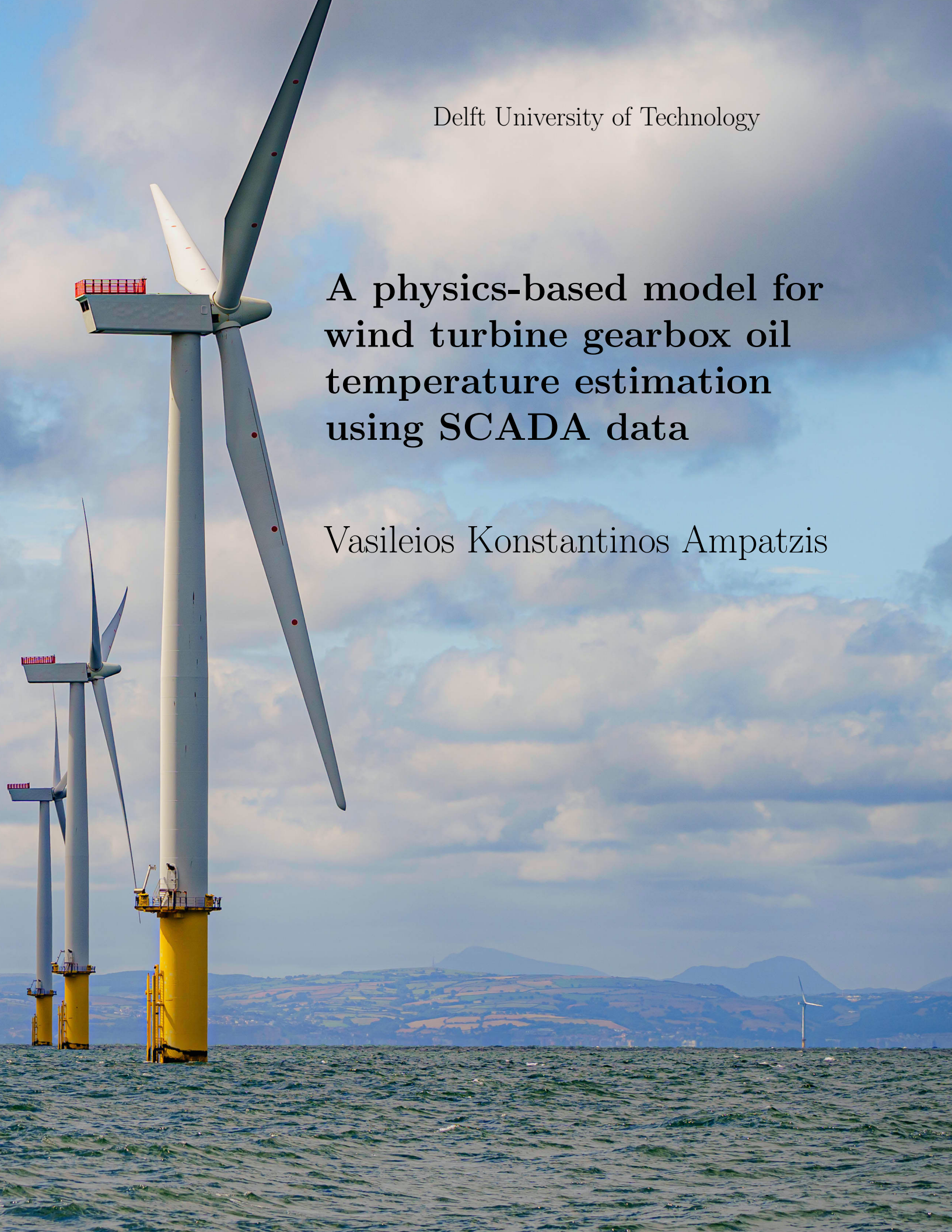


Delft University of Technology

A physics-based model for wind turbine gearbox oil temperature estimation using SCADA data

Vasileios Konstantinos Ampatzis

Delft University of Technology

A physics-based model for wind turbine gearbox oil temperature estimation using SCADA data

Author:

Vasileios Konstantinos Ampatzis

For obtaining the degree of Master of Science in Sustainable Energy Technology at the Delft University of Technology,
to be defended publicly on November 27, 2023 at 10:00 am

Student Number: 5545242
Thesis Committee: Dr. Donatella Zappalá (Supervisor)
Ali Eftekhari Milani (Supervisor)
Prof. Dr. Simon Watson
Dr. Dimitrios Zarouchas

Preface

This report was written as part of my thesis project for the MSc Sustainable Energy Technology offered by Delft University of Technology (TU Delft). As this educational chapter comes to an end, I would like to express my thoughts and gratitude to the people who have supported and helped me complete it.

First of all, I would like to express my deepest appreciation to my thesis supervisor, Dr. Donatella Zappalá for her invaluable support, guidance, and insightful feedback throughout the entire period of this thesis. Her expertise and encouragement were critical for guiding the direction of this research. Moreover, I extend my gratitude to the PhD candidate Ali Eftekhari Milani for his scientific support and helpful advice throughout this project.

I would like to thank my family for their unwavering encouragement, love, and understanding throughout my academic journey. Lastly, to my friends, especially those I met here and shared valuable moments with over these two years. Your support and motivation during difficult times and shared joy in my achievements have been a continuous wellspring of inspiration.

This achievement is an important step in my educational path, and it would not have been possible without the combined support and encouragement of these people. Thank you for being such an important part of this endeavour.

Abstract

The electricity generation from wind has seen significant growth over the past two decades with a compound annual growth rate of over 21%, and this trend is projected to continue, aligning with the goals of the European Green Deal. In the context of offshore wind turbines, operational and maintenance costs can account for up to 30% of the total costs throughout the project's lifecycle. This highlights the need for the development of efficient condition monitoring methods aimed at mitigating maintenance expenses.

The gearbox is one of the most failure-prone components in wind turbines, leading to extended downtimes and substantial financial outlays. When the components of the gearbox degrade, the heat generated within the gearbox increases. This increase leads to higher oil temperatures, making the temperature signal a suitable indicator of the gearbox health condition. The main objective of this thesis is to design a physics-based normal behaviour model (NBM) for the estimation of the wind turbine gearbox oil temperature and to use field data to validate its effectiveness. The energy conservation principle is applied to the gearbox to formulate an equation for the calculation of the gearbox oil temperature. To account for some unidentified design specification of the gearbox, such as gears and bearings dimensions, resulting in unknown parameters within the equation, the equation is fitted to available historic data from the Supervisory Control and Data Acquisition (SCADA) system to determine those parameters. The model uses as input an array of operational measured signals available in the SCADA system, including rotor speed, power output, nacelle temperature and inlet oil temperature, which is the temperature of the oil after running through the cooling system.

The model is validated by analysing its performance in estimating the oil temperature in a test dataset which has not been used for the model development according to a cross-validation scheme. Additionally, the weight of each equation term on temperature calculations is evaluated. This allowed us to understand the relative importance of different components, aiding in the refinement of the equation by excluding terms that have a negligible weight in the temperature calculation using the Akaike's Information Criterion. In order to investigate how well the physical characteristics of the gearbox are represented by the model, the coefficients of the energy balance equation derived from fitting it to the field SCADA data are compared to the parameters calculated using known characteristics of a reference gearbox. The comparison results indicate that the model's parameters are generally within the same order of magnitude as the reference values, highlighting the model's capability to capture and represent the gearbox's physical characteristics with good accuracy. To provide a benchmark, the performance of the physics-based model is also compared to that of two data-driven models using artificial neural networks (ANNs) for temperature prediction proposed in the literature. The results show that the proposed physics-based model outperforms both ANN models, achieving 14% and 59% lower root mean squared error and a reduction in time required for training of 70% and 95% when compared to the two ANN NBMs respectively. A case study demonstrates the model's effectiveness in accurately predicting the gearbox oil temperature with a mean absolute percentage error of 0.75% when the inlet oil temperature signal is available in the dataset. However, in a second case study where this critical signal is missing, the model's performance is negatively affected because the cooling system's influence on the oil temperature could not be accurately incorporated. This emphasizes the importance of including the inlet oil temperature signal for robust temperature estimation.

Contents

List of Figures

List of Tables

Nomenclature

Table 1: List of symbols

Symbol	Description	Unit
A_{ca}	Gearbox housing wall area.	mm
b	Gear teeth width and bearing width.	mm
$C_{1,2}$	Factors that depend on oil immersion depth and width of the gears.	-
C_p	Heat capacity.	J/K
C_{sp}	Oil splash factor.	-
d	Bearing bore diameter .	mm
d_m	Bearing mean diameter.	mm
d_{sh}	Shaft diameter.	mm
D	Bearing outside diameter.	mm
e	Euler number (e=2.718) .	-
f_0	Coefficient of bearing no load losses.	-
f_1	Coefficient of bearing load dependent losses.	-
F	Force.	N
F_n	Gear tooth normal force.	N
F_t	Force at pitch circle.	N
H_v	Geometrical loss factor.	-
k	Heat transmission coefficient.	$W/(m^2K)$
K_{rs}	Replenishment/starvation constant.	-
n	Rotational speed.	rpm
N	Number of data-points in regression analysis.	-
P	Power.	W
P_{losses}	Power losses in the gearbox.	W
P_{input}	Mechanical power at the input shaft of the gearbox.	W
P_V	Total power losses in the gearbox.	W
P_V	Total power losses in the gearbox.	W
P_{VD}	Losses from seals.	W
P_{VLP}	Load-dependent losses of bearings.	W
P_{VL0}	No-load losses of bearings.	W
P_{VZ0}	No-load losses of of gear mesh.	W
P_{VZP}	Load-dependent losses of gear mesh.	W
q	Heat flux.	W
Q_{diss}	Heat dissipation.	W
$\hat{r}$	Mean residual.	-
r_i	Residual at time i.	-
R	Thermal resistance.	K/W
R^2	Coefficient of determination.	-
R_a	Average roughness of pinion and gear wheel.	μm

Table 2: List of symbols

Symbol	Description	Unit
T_{gear}	Gearbox temperature.	$^{\circ}C$
T_{nac}	Nacelle temperature.	$^{\circ}C$
T_H	Hydraulic loss torque.	$N \cdot m$
T_{VLP}	Load-dependent loss torque of bearings.	$N \cdot m$
T_{VL0}	No-load loss torque of bearings.	$N \cdot m$
v_g	Sliding speed.	m/s
v_t	Tangential speed.	m/s
v_{t0}	Reference tangential speed (v_{t0}=10m/s).	m/s
v_Σ	Sum velocity.	m/s
$x_{1,2,\ldots,p}$	variables or regression analysis.	-
X_L	Lubricant factor.	-
$\bar{y}$	Mean of the observed values.	-
y_i	actual target value of regression analysis.	-
$\hat{y}_i$	actual target value of regression analysis.	-
α_t	Transverse pressure angle.	$^{\circ}$
α_{wt}	Working pressure angle.	$^{\circ}$
β_b	Helix angle at base circle.	$^{\circ}$
$\beta_{1,2,\ldots,p}$	Parameters of regression analysis.	-
η	Efficiency of the gearbox.	-
η_{oil}	Dynamic viscosity of oil.	$mPa \cdot s$
λ	Smoothing factor.	-
μ	Coefficient of friction.	-
μ_{mz}	Mean coefficient of friction.	-
ν	Kinematic viscosity of the oil.	mm^2/s
ρ	Radius of curvature at pitch point.	mm
$\rho_{C,red}$	Equivalent radius of curvature at pitch point of contact.	mm
ω	Rotational speed.	rad/sec

Table 3: List of abbreviations

Abbreviation	**Definition**
LCOE	levelized cost of electricity
O&M	operations and maintenance
SCADA	supervisory control and data acquisition
NBM	normal behaviour model
CBM	condition based maintenance
CMS	condition monitoring system
TNM	thermal network modelling
ANN	artificial neural network
FSRC	full signal reconstruction
FF	feed-forward
LR	layer-recurrent
NARX	Nonlinear AutoRegressive with eXogenous input
STIR	subspace trust-region interior reflective
MSE	mean squared error
RMSE	root mean squared error
MAPE	mean absolute percentage error
STDE	standard deviation of the error
AIC	Akaike information criterion
CV	cross-validation
TSS	time-series split
BTSS	blocking time-series split
CoV	Coefficient of Variance
EDP	Energias de Portugal

Chapter 1

Introduction

1.1 Motivation

The increasing awareness of climate change is forcing governments to focus more on sustainable energy technologies, with the power sector contributing to 39.3% of the total global CO2 emission in 2022 [1]. The global power sector is expected to transform from fossil-based to zero-carbon by the second half of this century, with renewable energy playing a major role in controlling and reducing the negative consequences of climate change. Renewable Energy Sources are a promising alternative to fossil fuels in tackling climate change. According to the European Green Deal, all members of the EU countries are obligated to reduce their greenhouse gas emissions by at least 55% by 2030 compared to the 1990 emissions. The ultimate aim is for them to achieve climate neutrality by 2050 [2]. This requires a substantial increase in both solar and wind energy generation in order to reduce the energy generated from fossil fuels, with the total renewable energy output projected to triple by 2030 [3]. This ambitious target is crucial for limiting global temperature rise to below $1.5^{\circ}C$. The energy generated by wind has been already increasing rapidly over the past 20 years with a compound annual growth rate of over 21% [4]. Figure 1.1 shows the increase of both the installed wind power global capacity and the annual additions in the past 11 years.

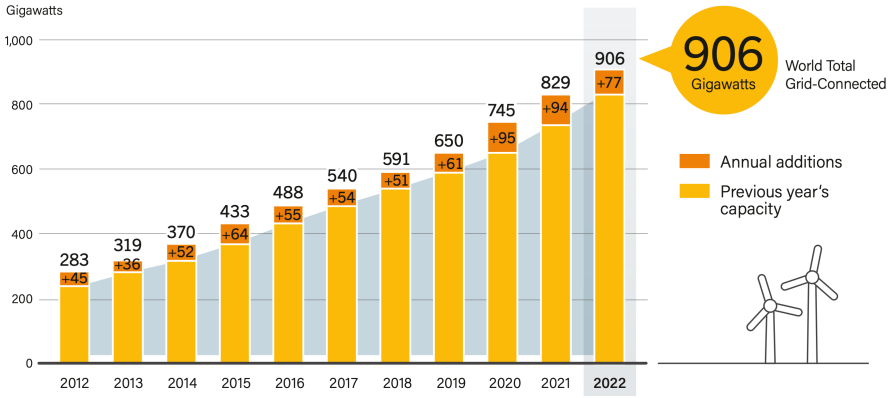


Figure 1.1: Global wind power capacity [5].

The installation costs per kW and levelised cost of electricity (LCOE) of wind power have reduced

significantly in the last decade, and the capacity factor has improved as well due to technological advancement and an increase in hub height [6]. Onshore wind energy has tripled in installed capacity in the past decade and is on the rise however, acquiring land for wind farms is a major constraint. Offshore wind has a higher capacity factor, and more favourable conditions while its LCOE decreased by 60.1% and its installation cost by 41.4% in the past 11 years as shown in the first and third plots of Figure 1.2. In addition, offshore wind farms do not face the problem of low land availability and public resistance as they are located far from shore. In fact, offshore wind energy has grown more rapidly than onshore wind energy in the last decade. Many European countries are increasing their installed wind capacity by exploiting the advantages of offshore wind, which has grown twentyfold in the past 12 years, from 3.1 GW to 63.2 GW [7]. However, offshore wind turbines operate in harsh weather and sea conditions, and this can lead to breakdowns, which reduce their reliability.

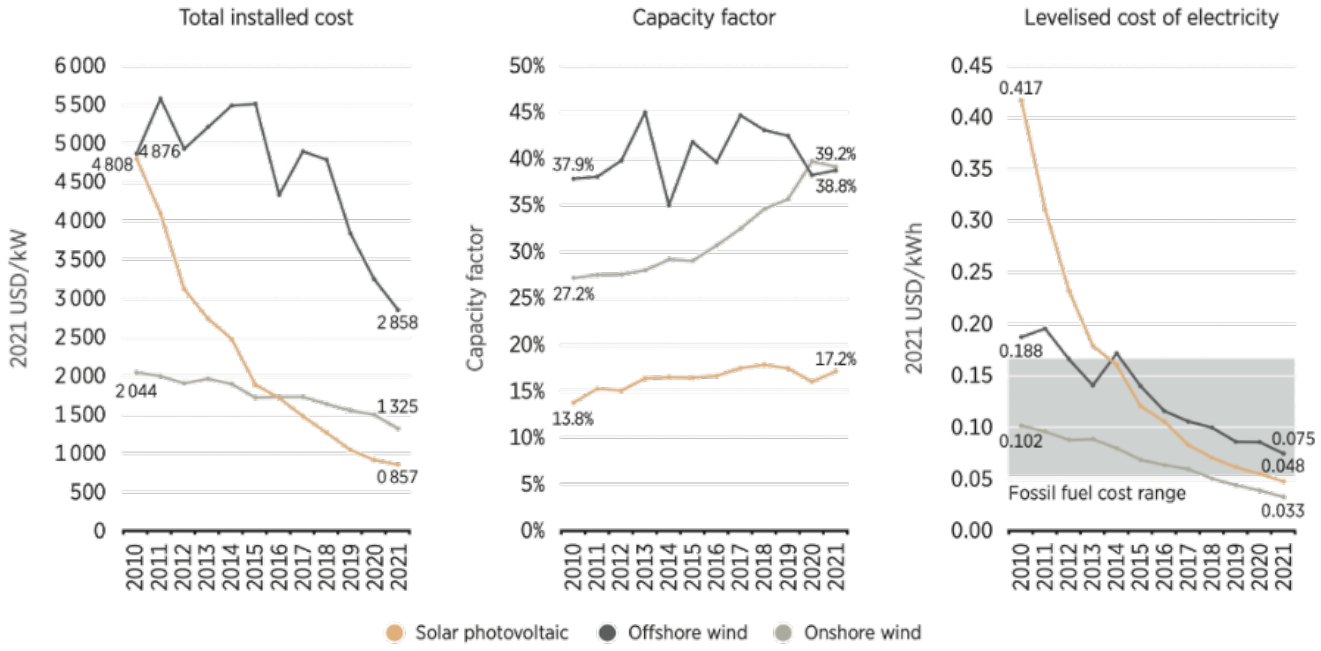
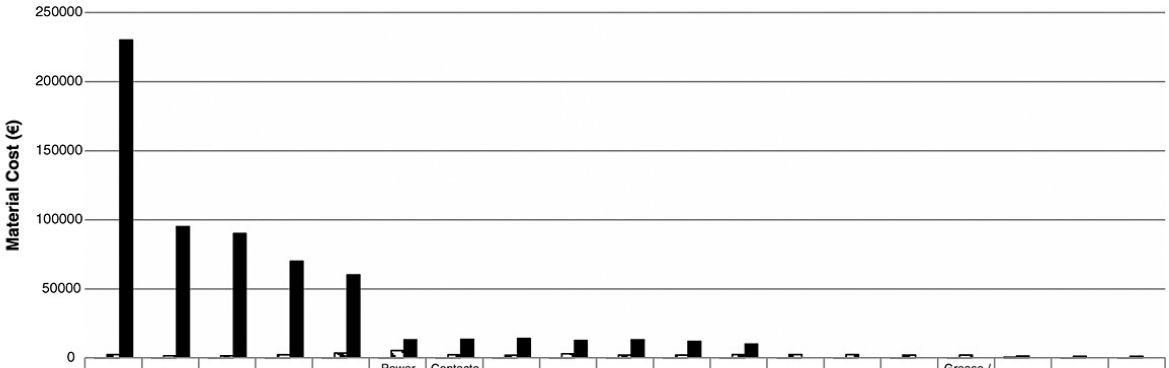


Figure 1.2: Renewable's characteristics trends [8].

As the industry trend is towards bigger and more powerful turbines, which are more prone to failure than smaller ones [9], [10], improving their reliability is crucial for the development of offshore farms but also onshore in the case wind farms are located in remote areas, such as mountains and islands. Furthermore, due to the weather conditions, there are months when it is nearly impossible to access offshore wind turbines, and this increases the need to address reliability issues. The lack of reliability, accessibility and availability leads to higher Operations and Maintenance (O&M) costs, which constitute a significant portion of wind farm project costs reaching up to 30%. By 2025, O&M costs are expected to globally reach 27.4 billion USD [11]. To keep wind energy financially viable and promote its growth, O&M costs must be reduced by enhancing the availability and reliability of wind turbines. Wind turbines typically have a designed lifetime of 20-25 years, but certain components may fail long before that, leading to significant downtime. The gearbox and generator components have the highest downtime while the gearbox is also the most costly component to repair as shown in Figures 1.3 and 1.4. Simultaneously, the gearbox is the

component with the highest failure rate, accounting for 12% of the total failures [12]. Therefore, improving the reliability of these components is crucial to minimising O&M costs and ensuring the financial viability and growth of wind energy.



	Gearbox	Hub	Blades	Transformer	Generator	Power Supply / Converter	Contactor / Circuit Breaker / Relay	Pitch / Hyd	Yaw System	Controls	Electrical Components	Other Components	Sensors	Safety	Pumps/ Motors	Grease / Oil / Cooling Liq.	Heaters / Coolers	Service Items	Tower / Foundation
Minor Repair	125	160	170	95	160	240	260	210	140	200	100	110	150	130	330	160	465	80	140
Major Repair	2500	1500	1500	2300	3500	5300	2300	1900	3000	2000	2000	2400	2500	2400	2000	2000	1300	1200	1100
Major Replacment	230000	95000	90000	70000	60000	13000	13500	14000	12500	13000	12000	10000	0	0	0	0	0	0	0

Figure 1.3: Cost of maintenance of the main wind turbine components [13].

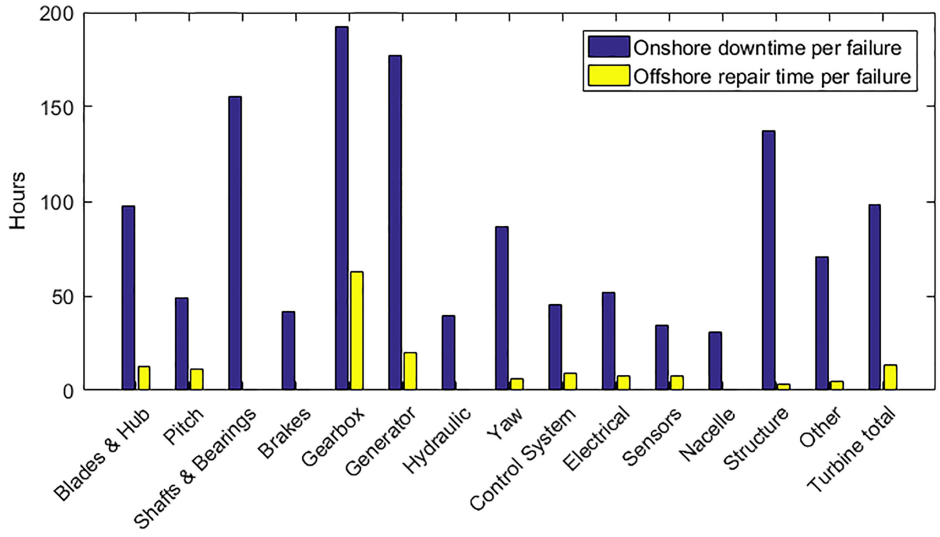


Figure 1.4: Downtime and repair time of the main wind turbine components [11].

There are several ways to improve the reliability of wind turbines such as higher manufacturing quality and proper load design [14]. One of the most widely explored methods though, is adopting effective condition monitoring approaches. Condition monitoring is the process of regularly monitoring the condition of a machine or component to detect any signs of wear, damage, or malfunction [15]. In the context of wind turbines, condition monitoring is essential to ensure their efficient and reliable operation, particularly in harsh environments such as offshore. By monitoring the condition of wind turbine components, such as the gearbox, it is possible to detect potential faults and defects before they lead to unscheduled downtime or costly repairs. This can reduce the O&M costs of the wind turbines and make them a more attractive investment in order to take a bigger share of the total electricity generation. This thesis focuses on the condition monitoring of the gearbox. This is

done using data recorded from the turbine Supervisory Control And Data Acquisition (SCADA) system. This system is included in all wind turbines and usually provides 10-minute average signals for several components of the wind turbine, including the gearbox. Further information on the different condition monitoring techniques is described in the literature review provided in Chapter 2.

1.2 Thesis scope and research questions

The main objective of this thesis is to design a physics-based normal behaviour model (NBM) for temperature calculation of wind turbine gearbox oil and use field data to validate its effectiveness. This physics-based model combines the use of SCADA signals and the application of heat balance principles to create a comprehensive approach for condition monitoring purposes. The normal behaviour can also be referred to as healthy behaviour which is the operation of the wind turbine long before a failure occurs or the components have degraded. To achieve this goal, first, a literature review of the existing methods is conducted, with a focus on the use of both physics-based and data-driven NBMs for gearbox condition monitoring. Then, a physics-based model is developed to calculate the temperature of the gearbox oil under normal behaviour conditions, using operational and environmental conditions as inputs. This model is then applied to real SCADA data from a healthy wind turbine gearbox using a training dataset. The effectiveness of the model in estimating the oil temperature under healthy conditions is then evaluated using a testing dataset. The difference between the modelled and the actual measured temperature, also known as residual, is calculated for the testing period and then used to assess the effectiveness of the model.

To achieve the goal of this thesis, the following sub-questions need to be addressed.

- Why should a physics-based model be developed for gearbox oil temperature calculation?
- How can a NBM of a wind turbine gearbox oil temperature be developed using a physics-based approach?
- How can the accuracy of a physics-based NBM be evaluated?
- What is the effect of each term of the heat balance equation in the calculation of gearbox oil temperature?
- Do the calculated parameters of the model accurately represent the physical characteristics of the gearbox?
- How does the performance of the proposed physics-based model compare to existing data-driven NBM approaches?
- Can a physics-based NBM using SCADA data accurately predict the gearbox oil temperature?
- How is the performance of the physics-based model affected when the input oil temperature signal is not available in the dataset?

The proposed research aims at the development of an effective temperature calculation model. By using a physics-based analysis, this research aims at leveraging both the physical principles governing the system and the available SCADA data to develop an accurate approach to predict the gearbox oil temperature under healthy operating conditions which can be used for detecting and diagnosing faults, ultimately leading to improved turbine performance and reduced O&M costs.

1.3 Thesis structure

The following is a breakdown of the report's structure. Chapter 2 includes a literature review of the existing research on condition monitoring of wind turbine gearboxes. Necessary information for the reader regarding the theoretical background is included in Chapter 3. Then, Chapter 4 describes the methodology proposed in this project including the description of the physics-based model. Chapter 5 presents the results of implementing and evaluating the proposed NBM, along with a comparative analysis against two data-driven NBMs. Finally, Chapter 6 provides the conclusions, discusses the key takeaways of this project and the recommendations for future work.

Chapter 2

Literature Review

This Chapter includes a literature review of the methods used for condition monitoring in wind turbine gearboxes. First, a short description of the gearbox technology used in wind turbines is provided followed by a discussion of their typical failures. Then, the different maintenance strategies typically used in the industry are discussed. The review continues with an introduction to wind turbine condition monitoring and a description of the different methods proposed in the literature. Finally, the current challenges and limitations of the existing approaches are discussed and research gaps are identified.

2.1 Wind turbine gearbox

Wind turbine gearboxes convert the low-speed blade rotations to high speeds while reducing the torque. This is necessary as the majority of the generators used in wind turbines are high speed and low torque, whereas the rotor of the wind turbine has low speed and high torque. Initially, wind turbine manufacturers tried to build wind turbines with high rotational speed in order to minimise the need for high-ratio gearboxes. This changed with the development of low-cost and high-efficiency gearboxes [16]. In Figure 2.1 the internal components of the nacelle are displayed including the position of the gearbox. Inside the gearbox, there are several components susceptible to failure. These include the bearings, the gears but also the lubricating oil [17]. A survey on the gearbox failures by the National Renewable Energy Laboratory shows that more than three-quarters of the gearbox failures are caused by bearing problems [18]. The pie chart in Figure 2.2 shows the distribution of the causes of gearbox failures.

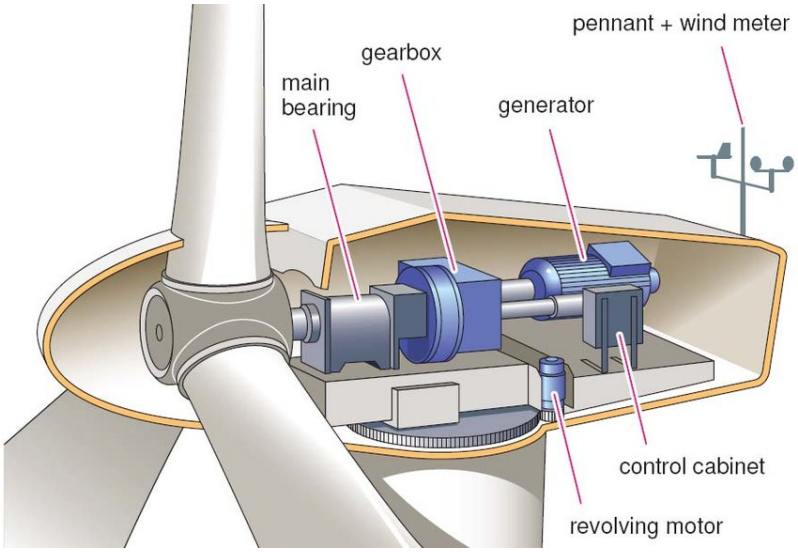


Figure 2.1: Wind turbine nacelle components [19].

The main causes of bearing failures are plastic deformation, wear, cracks and fractures, electric erosion, lubrication issues and contact fatigue. Plastic deformation can either be general surface or local surface deformation. The former usually occurs when the oil film is below a limit and the latter is due to bad assembly, misalignment or overload. Wear is the process of material removal from the bearing due to friction. Cracks are generated because of the stress in combination with the temperature exceeding the limit of the material. Fractures and cracks are divided into fatigue fractures, forced fractures and thermal cracks. Finally, electric erosion is caused by either excessive voltage or current leakage. Lubrication issues include insufficient lubrication, over-lubrication, ineffective lubrication due to the mixing of different lubricants and lubricant contamination from moisture and debris [20].

In the case of the gears, the main causes of failures are fretting corrosion, bending fatigue, scuffing and micropitting [17], [21]. Fretting corrosion starts from the lack of lubrication due to the oxidation of the material from air. It usually happens when the turbine is parked and the gears are not lubricated. Bending fatigue occurs in areas with high stresses. In gears, this is usually at the roots of the gear teeth. It starts as a crack and propagates until the tooth fractures. The main reasons for fatigue bending failure are bad design, a misalignment, assembly error or overload. The source of scuffing is overheating, which causes a small lubrication film further increasing the temperature and causing weld between the metals in contact. Finally, micropitting is also caused by a lack of sufficient thickness of the lubricant layer. However in this case the high shear stresses affect the contact areas [22].

A failure can also be caused by the lubrication due to contamination of the oil by wear debris, dust and water or due to foam generation [17]. This results in temperatures higher than those expected under normal operating conditions.

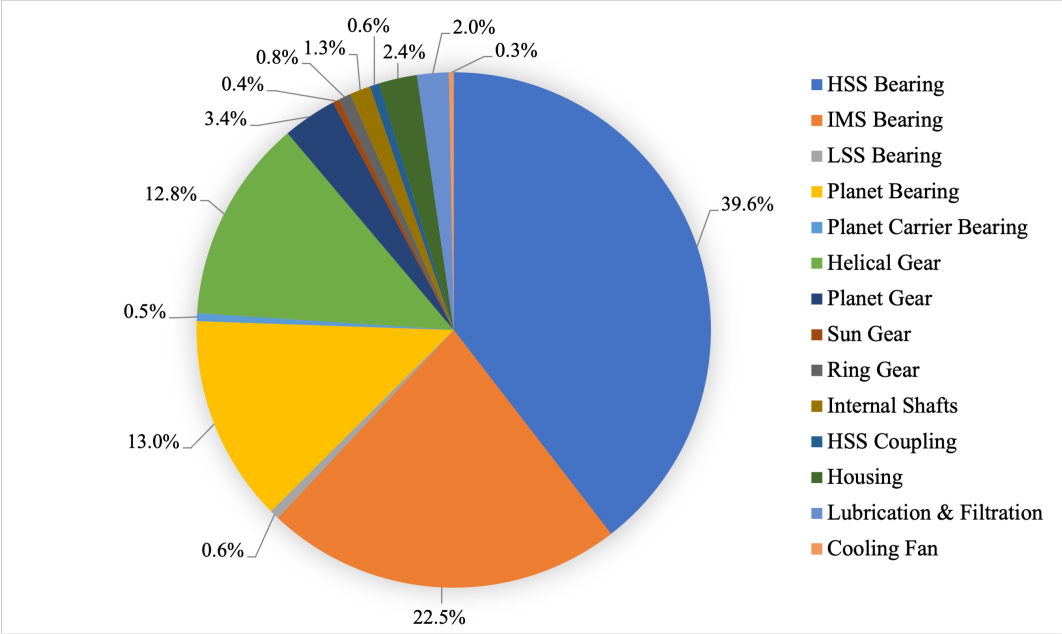


Figure 2.2: Wind turbine gearbox damage distribution statistics [18].

2.2 Maintenance Strategies

There are three main types of maintenance strategies: reactive maintenance, preventive maintenance, and predictive or maintenance [23]. Each of these strategies has its advantages and disadvantages.

- Reactive or correcting maintenance was the first form of maintenance dating back to the first machines ever made by humans. In this case, only when the part fails it is repaired or replaced. This maintenance strategy is not planned and only happens after a breakdown has occurred. It is cost-effective only when the number of failures in a system is really small. Implementing this strategy in offshore wind turbines would result in large downtimes due to the logistics required for accessing and maintaining the components.

- Preventive maintenance, also referred to as planned maintenance, involves scheduling maintenance at regular intervals to prevent failures, regardless of the actual health condition of the system. This results in an increase in the cost by replacing parts, even if some may not require replacement. Selecting an appropriate time interval for maintenance poses a challenge, particularly in the case of wind turbines, as they operate in constantly changing conditions. When this strategy is used, the interval is selected by trying to minimise the total cost of maintenance which consists of the extra cost of spare parts and the cost of not operating [24]. This strategy is cost-effective only when the frequency of failures in a system is really high and there is certainty about these failures. In this scenario, there would not be many unnecessary replacements due to the high certainty while the large number of failures would result in frequent unexpected stops in the case of reactive maintenance.

- Predictive maintenance, also known as condition based maintenance (CBM), is the most effective maintenance strategy. In this case, the maintenance occurs before the component fails but not when it still has a long remaining useful lifetime. This is achieved by monitoring

the system using sensors or on-site inspections. When a deviation from the normal behaviour is observed, an alarm for a potential failure is raised. So, it is a strategy that stands between corrective and preventive maintenance. This way, the cost of maintenance is reduced by the avoidance of catastrophic failure but also unnecessary maintenance.

Figure 2.3 shows the cost of maintenance for the three different maintenance strategies. If assisted by prognostics for the estimation of the remaining useful lifetime and maintenance scheduling, CBM can reduce the costs significantly compared to the corrective approach. [25]

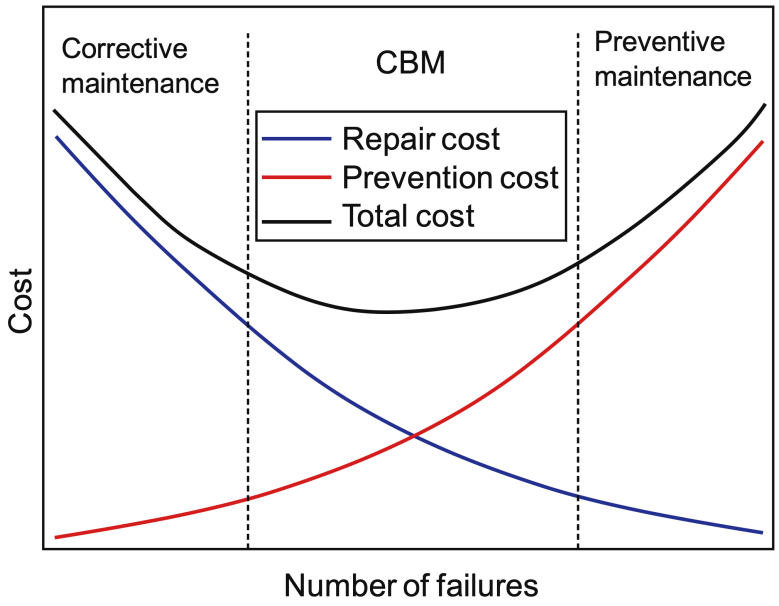


Figure 2.3: Cost for different maintenance strategies. Reproduced from [25].

2.3 Wind Turbine Condition Monitoring

Condition monitoring is an essential part of wind turbine operation, as it helps identify their health state. This can either be done online with the use of sensors or on-site with inspections happening periodically by experienced personnel and/or drones. The on-site inspections are usually more expensive for the operators as wind farms are usually installed in remote locations characterised by harsh weather conditions [26]. Several operators rely on online condition monitoring to enhance maintenance practices. It enables companies to monitor various parameters such as vibrations, oil quality, and temperatures in critical assemblies. By analysing this data, the health of the assets can be deducted, allowing for the estimation of remaining useful life or the identification of irregularities that may necessitate scheduled maintenance. This proactive approach to maintenance based on real-time monitoring helps optimise the asset performance and prevent unexpected failures [27]. The two monitoring systems that are most widely used in the industry are the SCADA system and the condition monitoring system (CMS).

- **SCADA system.** SCADA systems play a crucial role in monitoring and controlling wind turbines. These systems enable real-time data acquisition, remote control, and supervision of wind turbine operations. They collect data on various parameters like environmental

conditions, such as ambient temperature and wind speed, electrical characteristics, such as power output and current in each phase of the generator, temperatures from various components such as the gearbox sump oil and high-speed bearing, and finally control variables such as generator speed, and cooling pump status of the gearbox [28], [29]. This information is essential for efficient turbine performance. The SCADA usually provides 10-minute average values taken from 1Hz measured signals. In some cases, the minimum, the maximum and the standard deviation of each 10-minute time period are also provided. In addition, SCADA systems do not require additional costs of installing extra sensors because it is a standard feature of all wind turbines.

- **CMS.** Condition monitoring systems are installed independently in the wind turbines and usually they require the installation of several sensors in the wind turbine such as accelerometers for vibration analysis on the bearings, and oil quality sensors for debris and dust in the oil. There are several other less common types of analysis used by CMS such as acoustic emission, thermography, shock pulse, X-ray Micro-Tomography and Fiber Bragg grating sensor measurement. Usually, they are purpose-built systems providing high-frequency data of each component they monitor. Depending on the CMS the frequency can range from 1Hz to 30kHz [30].

Both SCADA systems and CMS have their respective advantages and disadvantages in the context of wind turbines. The SCADA system offers the advantage of being pre-installed in the turbines, eliminating the need for additional installation and operational costs [26]. Moreover, the storage required for the data generated by SCADA systems, typically in 10-minute averages, is relatively small. However, this limited frequency of data can also be a drawback as it restricts the system's capabilities for fault detection and prognostics [31]. Finally, condition monitoring reliant on SCADA data may not be universally applicable across all wind turbines, as the available signals within the SCADA system can vary between different turbine installations [28].

Conversely, CMSs require an initial investment that operators may be hesitant to undertake. Additionally, the storage requirements for high-frequency data in CMSs can be substantial. Transmitting 1Hz signals alone can consume around 1.8GB of storage per turbine per month [32]. Considering that wind farms often consist of numerous turbines, the storage demands further escalate. Nevertheless, CMSs benefit from the utilization of established methods like vibration analysis, which have proven effective in other industries [33].

This thesis aims at using the availability of SCADA data for monitoring. The main reason for this choice is the availability of SCADA in all wind turbines. The full potential of SCADA data needs to be investigated along with the effectiveness of analyses that do not require the extra cost of installing a CMS. Meanwhile, there are several studies suggesting the use of SCADA signals for condition monitoring of wind turbines, such as [26], [34], [35].

According to [28], [36] there are five main approaches for the use of SCADA data for condition monitoring:

- Trending
- Clustering

- Normal behaviour modelling
- Damage modelling
- Alarms assessment

These methods are described with some examples from the literature in the following subsections.

2.3.1 Trending

The trending method is a technique used in condition monitoring to analyse and interpret data patterns over time. It involves observing and tracking the changes and trends exhibited by specific parameters or variables of interest. In the context of wind turbines, trending is typically applied to SCADA data. By plotting and analysing the historical data collected from wind turbines, trending allows for the identification of patterns, deviations, and potential anomalies. These trends can provide valuable insights into the health and performance of the turbines. Trending methods can involve techniques such as regression analysis, scatter diagrams, principal component analysis, or bin averaging. Interpreting the trends requires careful analysis and domain expertise to differentiate between normal variations and abnormal behaviour. It may involve comparing current data with historical benchmarks, setting thresholds for alarm conditions, and manual interpretation of the filtered results [36].

One study using trending is [37]. It uses SCADA data in combination with a physical model to analyse the gearbox oil temperature increase for fault predictions. This model takes into account the energy flow in and out of the gearbox to show that the reduction of the efficiency due to a fault results in higher temperatures for the same power output. Also, it shows that the temperature increase is proportional to the rotational speed of the rotor assuming a fixed efficiency for the gearbox. Even though considering a constant efficiency in the gearbox is not realistic, as it varies with the output power, in this study it is sufficient to show the intended trend. By binning the gearbox oil temperature based on different power outputs and comparing the temperatures at 3, 6, and 9 months before failure, the analysis reveals a noticeable upward trend in temperature, particularly 3 months before failure occurs.

Qiu *et al.* [38] develop a thermophysical model of the gearbox using the basis of the approach in [37]. In this study, the efficiency of the gearbox is not considered constant but rather varies depending on power output according to [16] which states that half of the power loss in the gearbox is constant and the other half is linearly proportional to the power output. In this model, the power losses in the gearbox are calculated based on the efficiency and are equal to the dissipated energy. The dissipation is due to three main components:

1. Heat convection and conduction to the environment outside the gearbox, where the temperature of the nacelle is assumed to be the same as the ambient temperature.
2. Heat exchange with the oil cooling system, assuming constant inlet oil temperature and flow rate.
3. Heat conduction to adjacent components.

This paper shows the potential of using the proposed method for wind turbine fault detection as the simulated values of oil temperature are close to the real measurements from 9 and 6 months before

a failure, however, 3 months before the failure the real temperature is higher than the simulated one for all the range of power outputs.

Wilkinson *et al.* [39] compares three different fault detection methods for bearings in the gearbox and the generator. One of the methods is trending, where the temperature of a component is compared to the average temperature of the same component in all the wind turbines in the wind farm. A significant deviation from the average temperature indicates a faster degradation of a specific turbine and a potential failure. However, the authors find this approach to be inaccurate and subsequently dismissed its effectiveness due to the significant variability in operating conditions experienced by each individual wind turbine.

Corley *et al.* [40] propose a thermal network modelling (TNM) approach for the gearbox of a small wind turbine and the results are validated with an experiment setup. The TNM of the gearbox is based on [41] and is represented by an equivalent electrical circuit where each component is a node, the heat transfer is the current and the temperature difference between components is the voltage. The model requires the calculation of the losses in each component of the gearbox. Then, for constant operating conditions, the increase in temperature is calculated over time until the equilibrium temperature. The effect of degradation on the high and low-speed gearbox bearings is tested by adding heat generation on their node in order to observe how it affects the modelled temperature of the oil sump. A later study [42] utilising trending is conducted for the use of TNM in combination with SCADA data for fault detection in wind turbines. The TNM uses the signals from temperature sensors on several components of the gearbox which are indicated as nodes, to calculate the heat generated in each node. To evaluate how the model performed, the calculated heat generated in each node during healthy operation and one month before the failure are compared. This showed a significant increase in the heat generated from the faults. The authors mention that faults can be easier to detect using the heat domain instead of the temperature domain. [43] compares the effectiveness of a TNM with that of a machine learning model. Then the heat generation calculated in each node is used as input in the machine learning model and observed the importance of feature is high for the loss calculated from the TNM but varied between different wind turbines.

Using trending for condition monitoring has proven to be successful in detecting faults and especially using temperature signals of the drivetrain. Nonetheless, various studies indicate that manual interpretation is necessary while, employing a numerical representation of trends has not demonstrated clear improvement as it leads to frequent false alarms [28].

2.3.2 Clustering

Clustering is a data analysis technique used to identify groups or clusters within a dataset. In the context of SCADA data analysis for condition monitoring, clustering algorithms are applied to automate the classification of "normal" and "faulty" observations. Clustering algorithms work by grouping similar data points together based on certain characteristics or features. The goal is to maximise the similarity within clusters while maximising the dissimilarity between different clusters. The clustering process involves assigning data points to clusters iteratively until a stopping criterion is met. Various clustering algorithms exist, such as k-means clustering used in [44] and self-organising maps which is the second method introduced in the comparison by [39].

Clustering methods still require human intervention and the interpretation of results is not easy. Additionally, utilising fault data for training clustering algorithms may not always be feasible in real-world industrial settings. Therefore, further research and development are needed to refine and improve clustering methods for effective condition monitoring.

2.3.3 Normal Behaviour Modelling

Normal Behaviour Modelling (NBM) is a methodology used in condition monitoring of offshore wind turbines to model a desired variable. The basic idea is to create a model that represents the expected behaviour of the parameter under normal operating conditions. This model is built based on historical data and can be used to predict the desired target variable. The models used in the literature for the variable calculation are usually in two categories. The first one includes data-driven methods like artificial neural networks (ANNs), support vector machines, decision trees, random forests and gradient boosting [45] while the second category includes physics-based models which use physical principles of the system combined with historical data for the training to calculate the desired variable. Physics-based NBM can also be developed without the need for historical data in case all the necessary information about the components of the system is available.

The monitoring process in NBM involves comparing the measured signal with the modelled signal as shown in Figure 2.4. The difference between the two, known as the residual, acts as an indicator for potential faults or anomalies. Under normal conditions, the residual is expected to be approximately zero within a given tolerance. However, if there are changes in the conditions or the occurrence of failures, the residual will deviate from zero [46].

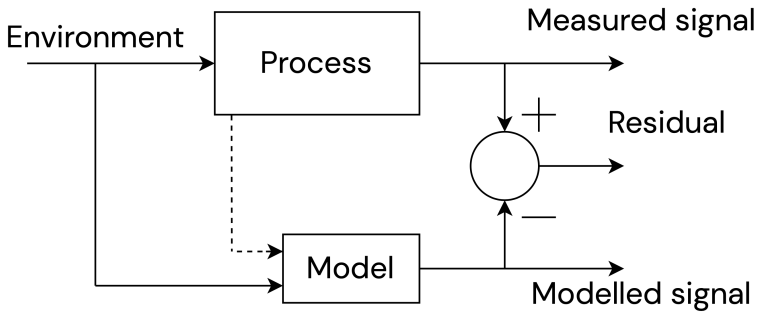


Figure 2.4: Normal Behaviour model schematic [47]

Data-driven NBM

There are several studies that use NBM and SCADA data for fault detection in wind turbines. Garcia *et al.* [48] propose an intelligent system for predictive maintenance application and test it on the wind turbine gearbox. The system is based on a NBM, which uses ANNs for the prediction of the target variables. The three models presented for the gearbox are the bearing temperature model, the thermal difference model and the cooling oil temperature. The thermal difference is described as the difference between the temperature of the oil before and after the cooling system.

Using the outputs of the three models they manage to detect the upcoming failure 26 hours before it occurs.

Garlick *et al.* [49] propose a condition monitoring technique referred as model-based that uses the generator temperature to estimate the temperature of the generator's bearing. This is done using a linear autoregressive exogenous input model with unknown parameters that are calculated with the least squares method and using SCADA data for training. The study concludes that this method needs human assistance to identify upcoming failures. It is suggested the use of a sliding window mean average of the residuals and the use of a threshold with alarms when it is exceeded many times in a certain period of time.

Schlechtingen and Santos [50] compare the use of linear regression and two different neural network models for condition monitoring. For this study, SCADA data are used to calculate the temperature of the generator and gearbox bearings in a healthy state. The two neural network methods are a full signal reconstruction (FSRC) and an autoregressive NBMs. The results of the three models are evaluated based on their success in predicting failure using the error for every 10-minute period and the daily average of the error. Overall, the linear regression model is found to be the least reliable and could only perform well in simple models. The autoregressive model performed similarly well with the FSRC model however, the latter is more likely to raise a false alarm due to the effects of seasonality. Finally, the authors suggest that only when the target variable has an autoregressive behaviour like variables with high inertia, the autoregressive NN model to be used over the FSRC model.

Tautz-Weinert and Watson [47] perform a comparative study between several data-driven NBMs which calculate the temperature of the bearing and the winding of a generator. The developed models are a simple linear NBM, a feed-forward (FF) and a layer-recurrent (LR) ANN NBM, a Gaussian process regression NBM, a Support Vector Machine NBM and an Adaptive Neuro-Fuzzy Inference System NBM. The input signals used for the models are chosen using a cross-correlation analysis between the SCADA signals and the target variables. The best performing methods are found to be the linear and the two ANNs. More than 100 turbines are tested and for some the LR-ANN performs better while for others, the FF-ANN is superior. For most turbines though, the LR-ANN has a superior performance.

McKinnon *et al.* [51] compares the use of two types of ANNs NBMs. The first one is a FSRC NBM and the second is a Nonlinear AutoRegressive with eXogenous inputs (NARX) model. In this study, the effect of the training length is investigated as well as the resolution of the data. The training lengths tested are 6 and 12 months while the resolutions are 10 and 60-minute averaged data. The best performance in predicting the target variable is achieved by using NARX model with a resolution of 10-minute averaged signals and a training period of 12 months. However, the authors suggest the use of FSRC model when only a lower resolution signal is available.

Physics-based NBM

Wilkinson *et al.* [39] propose a NBM, referred as physical model, based on a correlation analysis and the available SCADA signals. The model is applied in several components such as the main bearing, bearings and oil of the gearbox and bearings and windings of the generator. The temperature of

the component of interest is calculated as

$$T_{est} = \frac{R_1^2 (m_1 x_1 + c_1) + R_2^2 (m_2 x_2 + c_2) + \cdots + R_i^2 (m_i x_i + c_i)}{R_1^2 + R_2^2 + \cdots + R_i^2} \tag{2.1}$$

where R_i is the coefficient of determination and m_i is the gradient of the correlation of each signal used for the temperature prediction. The signals used in this prediction depend on the physics of the component. The reason this model is referred as a physical model in the paper is because the order of the polynomial used for each input signal in the equation is chosen based on the physical relation between the input and output signal. However, no further information is provided about this choice. The model parameters are calculated using data from a period of healthy operation of the wind turbine. Afterwards, the rolling average of the residual is calculated in order to identify how the actual data deviate from the estimated temperature over time. This study also uses the state of the cooling system which is a binary signal converted into a continuous signal by using a low-pass rectangular window filter. The study includes three approaches. The first uses signal trending, the second is an ANN NBM and the third is the physical method described above. Among the three proposed approaches, the physical model is the one with the highest precision to predict an upcoming failure but it is also the hardest to implement as it requires knowledge about the physical relations between the signals. The time of the failure identification ahead of the actual failure, ranges from 0 to 24 months for the gearbox with most of the detections happening 10-12 months before failure.

Cambron *et al.* [52] develop a physics-based model for the calculation of the temperature of the main bearing. This study is another example of a NBM where the predicted temperature is dynamically calculated and compared to the actual temperature taken from the SCADA data. The model uses the principle of energy conservation to formulate the equation for the calculation of the bearing temperature. The temperature of the bearing is calculated as

$$T_b(t) = \beta_1 T_b(t-1) + \beta_2 T_{nac}(t) + \beta_3 \omega^2(t) + \beta_4 P(t) \tag{2.2}$$

where T_b is the temperature of the main bearing, T_{nac} is the temperature of the nacelle, ω is the rotational speed of the rotor, P is the electrical power output and $\beta_1 - \beta_4$ are unknown parameters which are calculated by fitting the equation to the available SCADA data in a period of healthy operation of the wind turbine. The fitting is performed using the least squares algorithm. The model inputs are the temperature of the bearing in the previous timestep, the air temperature in the nacelle, the rotational speed of the rotor and the power output found in Equation 2.2. The model performance is assessed using the root mean square error, the coefficient of determination R^2 and the Durbin-Watson test which calculates the autocorrelation of the residuals. For fault detection, an exponentially weighted moving average control chart is used. The model is tested in two cases where it is able to detect the fault 30 days and 50 days ahead of the failure, respectively. Finally, the authors suggest a similar model to be developed for gearboxes which are more prone to failures compared to main bearings.

The more recent studies on NBM mainly use AI techniques. These can be shallow machine learning models like random forest [53] and gradient boosting [54] but also deep neural networks like transfer learning [55] and multi-target neural network [56]. While the latest literature primarily emphasises machine learning techniques, it remains necessary to establish their practical suitability for real-world implementation. This necessity arises not only from the substantial data requirements for

effective training but also from the time required for this training before they can be implemented in a new project. In addition, machine learning algorithms are often regarded as 'black box' methods. These models can provide accurate predictions, but they do so without offering easily interpretable insights into the underlying relationships or factors driving their results [45]. Finally, NBM has proven effective in detecting failures, but it comes with challenges such as reliance on training data and manual threshold setup, leading to potential issues like missed failures and false alarms.

2.3.4 Damage modelling

Several studies developing models for the degradation of wind turbine components have been proposed in the literature. They analyse the root causes of the failures and investigate the physics that lead to that failure. Damage modelling involves developing mathematical or physics-based models to understand and predict the behaviour and degradation of various components within a wind turbine. The development of such models is really complex as it requires a combination of engineering knowledge, field data, material fatigue expertise, and computational tools.

Gray and Watson [57] investigate the root cause of failures in various components of wind turbines. Then, the physics of failure is used in order to develop a damage model for the gearbox bearings. This model for the degradation of the bearing is based on the Lundberg-Palmgren rule. The probability of failure is then estimated based on the Kaplan-Meier estimator. This methodology is applied in a case study involving a wind farm of 160 wind turbines, six of which have a gearbox failure in a short time span. With this approach, it becomes feasible to calculate the real-time probability of failure based on the applied load and the accumulated damage.

Sepulveda *et al.* [58] use a physics-based model to calculate the accumulated damage on the bearings and the gears of the gearbox. This model uses historical data from SCADA to make batches for different power outputs of the wind turbine and calculates the frequency of occurrence for each batch. Then, by using the N-S curves of the materials and the Miner's rule the total damage is calculated. The same is done for future accumulative damage by using FAST simulation tool to generate a load spectrum. In order to perform the load calculation, this model requires the knowledge of the exact type and size of each gearbox component, which is not usually available to wind farm operators.

Jantara *et al.* [59] use damage mechanics to estimate the remaining useful lifetime of gears. A finite element model is developed to calculate the strains on the gears of wind turbine gearboxes under certain scenarios like overloading, poor lubrication, and misalignment. Then, the resulted stresses are used for the calculation of the accumulated damage and the stress-cycles curves. Finally, it is possible to calculate the remaining useful lifetime based on that information. This research shows that the angular misalignment has the biggest impact on the lifetime of the gears. Even though this study shows that it is possible to estimate the remaining useful lifetime, just like the other studies in this category it is necessary to have access to the exact dimensions of the components of interest.

2.3.5 Alarm Assessment

The assessment of alarms can either be done using the status of the wind turbine which is generated from the SCADA system or using the outputs of NBM. The former is also mentioned in the literature as status code processing and is used in [60], [61]. The latter is used in several studies

such as [48], [62].

Status code processing for wind turbine condition monitoring involves analysing codes generated by SCADA systems. These codes, which represent various turbine states and alarms, are scrutinised to assess operational health, detect faults, and inform maintenance decisions. Methods such as time-sequence analysis, Bayesian probability analysis [60], and ANNs [63] are employed to extract valuable insights from these codes. Challenges include data dependency, varying code generation practices among manufacturers, and the need for more detailed algorithms in commercial products [28].

Using expert systems and outputs of NBMs involves applying predefined rules and fuzzy logic within a computational framework. Several studies implement expert systems to assess the output of their various modelling approaches, such as ANN [48] and Adaptive Neuro-Fuzzy Interference Systems [62]. These systems use fuzzy rules and membership functions to diagnose anomalies, predict the remaining lifetime, and plan maintenance. By analysing prediction errors, integrating health degrees or risk indicators, and considering various ANN models or training data, expert systems simplify the interpretation of condition monitoring results. They offer a structured approach to integrate SCADA condition monitoring into maintenance strategies, aiding in effective decision-making and cost-efficient maintenance planning.

There are also more recent studies implementing assessment of alarms. [64] combine the use of alarms from status signals with the output of two ANN models to develop fuzzy logic-based methodology. The first ANN analyses vibration data and the second SCADA data. [65] uses alarms from the status signals of the SCADA system to compare the use of time-sequence analysis with the use of probability-based analysis. Overall, alarm assessment for condition monitoring has demonstrated advantages in the context of fully autonomous fault detection systems and is easier to implement in real-world wind farms when compared to the previous categories of condition monitoring systems [36].

2.4 Discussion

The foundation of effective wind turbine maintenance lies in selecting the most suitable strategy. The literature review suggests that predictive maintenance outperforms preventive and reactive maintenance strategies. Predictive maintenance's proactive nature, driven by real-time data and condition monitoring, allows for the early detection of faults and deviations.

Predictive maintenance can be performed either using a CMS or using the SCADA system which is installed in all turbines. Even though SCADA data include less information than a dedicated CMS as the data are stored in 10-minute intervals, it is worth exploring the full potential of the SCADA data for condition monitoring purposes before investing the extra cost and effort associated with CMS installation.

Condition monitoring in wind turbines encompasses various approaches, including trending, clustering, NBM, damage modelling, and alarm assessment, each with its unique strengths and limitations. Trending involves analysing data patterns over time, often relying on historical SCADA data to identify trends, deviations, and anomalies. Clustering, on the other hand, automates the classification of normal and faulty observations through data grouping. NBMs, whether data-driven

or physics-based, aim to model normal turbine behaviour, making them effective at detecting deviations from the healthy state. Damage modelling involves developing physics-based models to understand and predict the behaviour and degradation of various components within a wind turbine, offering a deep understanding of failure mechanisms. Alarm assessment methods, such as status code processing and expert systems, focus on analysing SCADA-generated alarms and NBMs output for fault detection and maintenance planning.

Both Trending and Clustering require human intervention for the interpretation of the models. This means that their implementation in real-world scenarios can not be automated. Damage modelling is the method used so far the most for remaining useful lifetime estimation however, it requires precise knowledge of the components included in the system which is not usually the case for wind turbine gearboxes. NBM can sometimes be challenging to develop and interpret but in combination with expert systems, they result in effective condition monitoring of wind turbines.

Physics-based NBMs in wind turbine condition monitoring are based on engineering principles, offering an understanding of the system's behaviour. These models, while interpretable and accurate under known conditions, often require specific data and parameter calibration. By using historical data this limitation can be overcome. In contrast, data-driven models rely solely on historical data, offering flexibility and adaptability to changing conditions. They can handle noisy data and can be used for various systems, though they have a "black-box" nature. The choice between these approaches hinges on the availability of data, system complexity, and the need for interpretability.

Based on the conducted literature review, it is observed that generally physics-based models have very good performance when used for the main bearing [52] and bearings of the gearbox and generator [39], [42]. This holds true for both trending and NBMs. At the same time, there is no available study on a NBM of the gearbox oil temperature that uses a physics-based approach. This has the potential to achieve similarly good results with studies on other components of the wind turbine.

Chapter 3

Theoretical Background

Before proceeding to the description of the methodology used in this study, it is necessary to provide some background information about the methods that will be used. First, the first law of thermodynamics is presented as it is the base for the development of an equation that calculates the gearbox oil temperature in the methodology. In addition, the sources of power loss in the gearbox and the heat dissipation are introduced. These are included in Section 3.1. Then, Section 3.2 introduces regression analysis which is used to calculate the unknown parameters of the equation for the calculation of the temperature of the oil. In addition, the methods used to evaluate the model performance are introduced in Section 3.3. Finally, an introduction to artificial neural networks is provided in Section 3.4. ANNs are used to develop a NBM and compare its performance with that of the physics-based NBM developed in this study.

3.1 Gearbox energy balance

Just like any system, the gearbox is governed by the first law of thermodynamics which is also known as the conservation of energy principle. This principle dictates that the difference between the total energy inflow and the total energy outflow from the system over the course of a process is equal to the overall change in the system's total energy. According to Cengel [66], for a closed system with a fixed mass, the energy conservation equation can be written as

$$E_{in} - E_{out} = \Delta U = mC\Delta T \tag{3.1}$$

where E_{in} and E_{out} are the energy inflow and outflow of the system, m is the mass of the system, C is the specific heat capacity, and ΔT is the change in temperature. Equation 3.1 can be written in the rate for as

$$\dot{E}_{in} - \dot{E}_{out} = mC\frac{dT}{dt} \tag{3.2}$$

For a wind turbine, the power flow is shown in Figure 3.1.

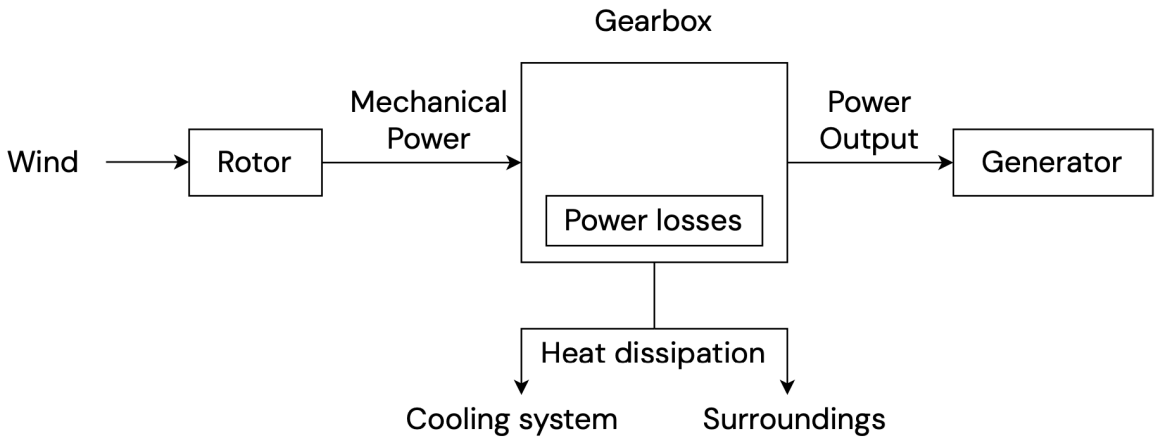


Figure 3.1: Power flow in a wind turbine drivetrain

Wind is converted into mechanical power through the rotor. Then, the power is transferred through the gearbox to the generator where it is converted to electricity. However, within the gearbox, a portion of the mechanical power is lost due to friction and it converts to heat. This heat is dissipated to the environment by the cooling system but also through convection and conduction to the surroundings. So, if we consider a control volume that includes only the gearbox, Equation 3.2 can be written as

$$P_{in} - P_{out} - \dot{Q}_{diss} = mC\frac{dT_{gear}}{dt} \tag{3.3}$$

where P_{in} and P_{out} are the power input and power output of the gearbox respectively, $\dot{Q}_{diss}$ is the heat dissipated from the gearbox, m is the mass of the gearbox and C is the specific heat capacity of the gearbox. The difference between the input and output power of the gearbox is equal to the power losses in the gearbox so Equation 3.3 can be written as

$$P_V - \dot{Q}_{diss} = mC\frac{dT_{gear}}{dt} \tag{3.4}$$

where P_V is the power loss in the gearbox. The sources of power loss in the gearbox and the heat dissipation from the gearbox are introduced in the following subsections.

3.1.1 Gearbox losses

In this section, the different sources of power loss inside the gearbox are described and the relevant equations are provided. The source of the power loss, P_V, in a gearbox is shown in Figure 3.2.

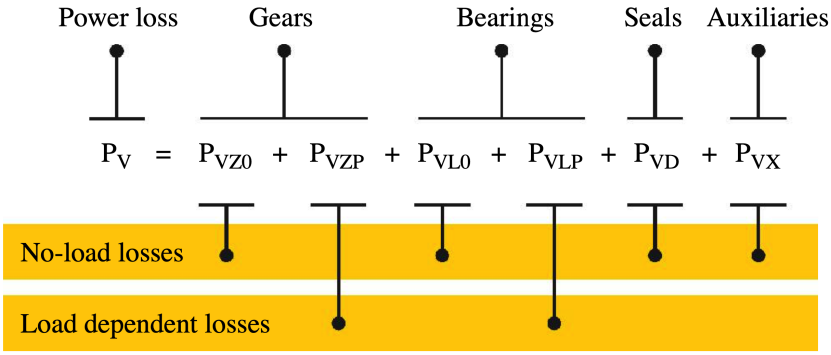


Figure 3.2: Sources of power loss in gearbox [67]

The losses consist of the load-dependent and the no-load losses of the gear mesh, P_{VZP} and P_{VZ0}, the load-dependent and the no-load losses of the bearings, P_{VL0} and P_{VLP}, as well as the load-independent losses of the seals, P_{VD}, while the auxiliary losses are usually disregarded [40], [68]–[72]. So, the total power loss is calculated as

$$P_V = P_{VZ0} + P_{VZP} + P_{VL0} + P_{VLP} + P_{VD} \tag{3.5}$$

Gear losses

The load-dependent losses of the gears are generated from friction in the meshing of the gears, whereas no-load losses are from oil churning which includes oil splash and squeeze [71].

The load-dependent losses are calculated using the Coulomb law according to ISO/TR 14179-2:2001 standard [73] as:

$$P_{VZP} = P_{\text{in}} \mu_{mz} H_v \tag{3.6}$$

where P_{in} is the input power coming from the rotor, μ_{mz} is the mean coefficient of friction, and H_v is the geometrical loss factor, which is constant. The Equation 3.6 is also used extensively in the literature for the calculation of the load-dependent losses of the gears [40], [68], [70]. P_{VZP} refers to the power loss of one gear meshing point, so the total load-dependent power loss of the gears is the sum of the losses in each meshing point. The terms of Equation 3.6 are further analysed below. First, μ_{mz} is calculated as

$$\mu_{\text{mz}} = 0.048 \left(\frac{F/b}{v_\Sigma \rho} \right)^{0.2} \eta_{\text{oil}}^{-0.05} Ra^{0.25} X_{\text{L}} \tag{3.7}$$

where b is the width of the gear teeth, η_{oil} is the dynamic viscosity of the oil, R_a is the average roughness of pinion and gear wheel, X_L is the lubricant factor, ρ the radius of the curvature at pitch point and F the force on the meshing point. The sum velocity of the meshing point of the gears, v_Σ is calculated as

$$v_\Sigma = 2 \cdot v_t \cdot sin\alpha_{wt} \tag{3.8}$$

where, v_t is the tangential speed of the gear and α_{wt} is the working pressure angle.

There are several studies on the calculation of no-load losses. In this study, the calculation is performed using the ISO/TR 14179-2:2001 standard [73] which has been used in a study developing a thermal model of the gearbox [40]. There are also more recent methods, such as [74], [75], that use multiple equations from which one is chosen based on the operating conditions. However, wind turbines operate under constantly changing conditions so the use of such methods is not considered suitable for this study. According to ISO/TR 14179-2:2001 the total no-load losses are described by

$$P_{VZ0} = T_H \cdot \omega \tag{3.9}$$

where ω is the rotational speed of the gear and T_H is the hydraulic torque loss given by

$$T_{\mathrm{H}} = C_{\mathrm{Sp}} C_1 e^{C_2\left(\frac{v_{\mathrm{t}}}{v_{\mathrm{t0}}}\right)} \tag{3.10}$$

where C_{Sp}, C_1 and C_2 are constants depending on the oil immersion depth and the width of the gears, v_t is the tangential speed which is proportional to the rotational speed of the rotor and v_{t0} is the reference tangential speed which is assumed equal to $10 m/s$.

Bearing losses

The calculation of the power loss in gearbox bearings is performed using an approach proposed in ISO/TR 14179-2:2001 [73], which is also used in [76] and [40]. The loss is divided into no-load losses (P_{VL0}) and load-dependent losses (P_{VLP}). The no-load losses are calculated as

$$\begin{aligned} P_{VL0} &= T_{VL0} \cdot \omega = 1.6 \cdot 10^{-8} \cdot f_0 \cdot d_m^3 \cdot \omega \quad \text{for } v_{oil} n < 2000 mm^2/s \cdot min \\ P_{VL0} &= T_{VL0} \cdot \omega = 10^{-10} \cdot f_0 \cdot (v_{oil} n)^{2/3} d_m^3 \cdot \omega \quad \text{for } v_{oil} n \geq 2000 mm^2/s \cdot min \end{aligned} \tag{3.11}$$

where T_{VL0} is the torque loss, f_0 is a constant that depends on the bearing type and the bearing lubrication, n is the rotational speed of the bearing [rpm], d_m is the bearing mean diameter [mm], and ω is the rotational speed [rad/sec].

As for the load-dependent losses, they are calculated as

$$P_{VLP} = T_{VLP} \cdot \omega = f_1 \cdot F^{\alpha} \cdot d_m^{\beta} \cdot 10^{-3} \cdot \omega \tag{3.12}$$

where f_1, α and β are constants that depend on the bearing type and F is the load applied on the bearing [N].

Seal losses

Seals are mechanical components designed to prevent the leakage of lubricating oil and the ingress of contaminants like dust, dirt, or moisture into the gearbox. Lip seals are placed on the input and output shafts. Due to the rotation of the shaft, there is friction between the shaft and the seals which generates losses. These seal losses are calculated as

$$P_{VD} = 7.69 \cdot 10^{-6} d_{sh}^2 n \tag{3.13}$$

where d_{sh} is the shaft diameter [mm] and n is the rotational speed [rpm]. This equation is widely used in the literature for the calculation of the seal losses [40], [72], [77] and is known as Simrit

equation.

Oil Viscosity

In some of the aforementioned equations, the dynamic and kinematic viscosity of the oil are present. The dynamic viscosity η_{oil} of Equation 3.7 can be expressed as a function of kinematic viscosity ν_{oil} with

$$\eta_{\text{oil}} = \rho_{\text{oil}} \cdot \nu_{\text{oil}} \tag{3.14}$$

where ρ_{oil} is the density of the oil [78]. The effect of temperature on the dynamic viscosity of an oil can be estimated with the Arrhenius equation [79] as

$$\eta_{\text{oil}} = A \cdot exp\left(\frac{E_a}{RT_{\text{oil}}}\right) \tag{3.15}$$

where A is a constant, E_a is the activation energy, R is the universal gas constant and T_{oil} is the temperature of the oil. By dividing both sides of the equation with ρ_{oil} and using Equation 3.14 we get

$$\frac{\eta_{\text{oil}}}{\rho_{\text{oil}}} = \frac{A}{\rho_{\text{oil}}} exp\left(\frac{E_a}{RT_{\text{oil}}}\right) \Rightarrow \nu_{\text{oil}} = A_\nu exp\left(\frac{E_a}{RT_{\text{oil}}}\right) \tag{3.16}$$

where A_ν is a constant if a constant oil density is assumed.

3.1.2 Gearbox heat dissipation

The heat generated within the gearbox due to the losses is first transferred to the lubrication oil circulating in the gearbox [73] and then dissipated to the external environment in two ways: a) through the gearbox housing and the coupling shafts, mostly through convection and conduction, to the air within the nacelle and other components in contact with the gearbox, and b) through the cooling system. The heat dissipation through radiation can be neglected as suggested in [38], because of its low relative contribution to the total heat transfer. According to [38], [73], the heat dissipated through convection and conduction to the nacelle is calculated as

$$\dot{Q}_{nac} = k_0 \cdot A_{ca}(T_{oil} - T_{nac}) \tag{3.17}$$

where k_0 is the heat transmission coefficient which includes the internal heat transfer, the heat transfer between housing and the oil, the heat conduction through the housing wall, and the external heat transfer to the environment, and A_{ca} is the gearbox housing wall area. Both k_0 and A_{ca} are assumed to be constant. T_{oil} is the oil sump temperature and T_{nac} is the temperature of the nacelle.

The heat dissipated through the cooling system [66] is calculated as

$$\dot{Q}_{cool} = \dot{m} C_p \Delta t = \dot{m} C_p (T_{oil} - T_{oil,inlet}) \tag{3.18}$$

where $\dot{m}$ is the mass flow rate of the oil in the cooling system, C_p is the specific heat capacity of the oil and $T_{oil,inlet}$ is the oil inlet temperature. The total heat dissipated from the gearbox can then be calculated as

$$\dot{Q}_{diss} = \dot{Q}_{nac} + \dot{Q}_{cool} = k_0 \cdot A_{ca}(T_{oil} - T_{nac}) + \dot{m} C_p (T_{oil} - T_{oil,inlet}) \tag{3.19}$$

3.2 Regression Analysis

Regression analysis is a statistical technique that describes a dependent variable based on a number of independent variables. In this thesis, it is used to calculate the unknown parameters in the equation for the calculation of the gearbox oil temperature. These unknown parameters represent the unknown characteristics of the gearbox which are required for the calculation of the gearbox oil temperature. In this section, the application of regression analysis in our study is discussed, with a focus on the least squares method. An example of a multiple linear regression formula is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + e \tag{3.20}$$

where y is the dependent variable, $x_1, x_2, ..., x_p$ are the independent variables, $\beta_0, \beta_1, ..., \beta_p$ the unknown parameters and e the error term.

When the relationship between the dependent and the independent variables is not linear, non-linear regression is used. Examples of non-linear relationships are exponential, logarithmic, trigonometric or power functions.

The least squares method is a regression analysis method that minimises the sum of the squared difference between the observed value y_i and the predicted value $\hat{y}_i$

$$F(\beta) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2 \tag{3.21}$$

to estimate the regression parameters $(\beta_0, \beta_1, ..., \beta_p)$. β is a vector including the regression parameters and vector x_i include the independent variables. The least squares method can be used for both linear and non-linear regression. In this study, a non-linear least squares method is used to calculate the optimal parameters. The non-linear least squares method relies on iterative algorithms like Gauss-Newton, Levenberg-Marquardt, Subspace Trust-Region interior Reflective (STIR), Conjugate Gradient or Nelder-Mead Simplex. Starting from an initial estimate of the unknown parameters, these algorithms iteratively refine the solution by modifying the parameters to minimise the sum of squared error [80]. STIR stands out as an efficient option capable of handling a substantial number of parameters and constrained problems. It strikes a balance between exploring and exploiting the solution space, which enhances the probability of converging to the global minimum and prevents becoming trapped in local minima [81].

STIR algorithm is proposed in [82], where it is described in detail and compared to the Trust-Region interior Reflective algorithm which does not introduce subspaces. In Trust-Region algorithms the function $F(\beta)$ we want to minimise is approximated by a simpler function q, which offers a reasonable representation of F in a certain region, referred as neighbourhood N, around a point β_k. The approximation q is usually defined as the first two terms of the Taylor approximation of F at β_k. Then, a subproblem is introduced which includes minimising the simplified q as

$$\min_s \{q(s), s \in N\} \tag{3.22}$$

where the step s is calculated and the the point β_k is updated to $\beta_k + s$ if

$$F(\beta + s) < F(\beta) \tag{3.23}$$

In case this condition is not met, the previous point β is kept, the size of neighbourhood N is reduced and the minimisation of q is repeated. These steps are repeated until the algorithm converges to a solution.

Solving the trust-region subproblem can be quite demanding and often necessitates multiple Hessian matrix factorisation, which can be computationally costly. To address this challenge, an approximation technique is employed in STIR algorithm. This approximation method for computing the trial step s involves confining the trust-region subproblem to a two-dimensional subspace instead of the dimension of β. These subproblems are a way of dividing the optimisation task into smaller, more manageable steps. The shape and size of the trust regions are determined based on two key factors: the distance from the defined constraints and the direction of the gradient. This way, steps directly into bounds are avoided and the whole two-dimensional space is explored. The convergence of the optimisation problem is also improved by considering search directions that are reflected from the defined constraints. Finally, to adhere to theoretical requirements and maintain the integrity of the optimisation problem, the algorithm ensures that in none of the iterations are the constraints violated.

3.3 Normal behaviour model accuracy

The evaluation of the model accuracy is a crucial step in the development of a NBM. This shows how well the model can predict the output based on the given input. This is usually done using SCADA data from a different period to the training which is called the testing period. Several metrics have been used in the literature for the evaluation of the accuracy of a NBM. Most of the metrics use the errors also referred as residuals of actual minus estimated values and are calculated as

$$r_i = y_i - \hat{y}_i \tag{3.24}$$

where r_i is the residual at time i, y_i is the actual value at time i and $\hat{y}_i$ is the NBM predicted value at time i Some of the most commonly used metrics are:

- Mean Squared Error (MSE), which is the mean of the squared difference between the predicted and the actual value calculated as

$$\text{MSE} = \frac{\sum_{i=1}^{N} r_i^2}{N} \tag{3.25}$$

  where N is the number of the data points. The MSE is sensitive to outliers due to the exponent in the error. This metric is used by [83], [84] to calculate the accuracy of the developed NBMs.

- Root Mean Square Error (RMSE), which is the square root of the mean squared difference between the predicted and the actual value and is calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} r_i^2}{N}} \tag{3.26}$$

  The RMSE has been used by several studies, such as [47], [49], [51], [52], [85], to evaluate the performance of the developed NBMs. Just like MSE, RMSE is also sensitive to outliers due

to the exponent in the residual. However, RMSE is more widely used than MSE because its unit is the same as the actual and the estimated values. This way it is easier to compare the magnitude of the error with the magnitude of those values.

- Mean absolute error (MAE), which is the average value of the absolute difference between the predicted and the actual values

$$\text{MAE} = \frac{\sum_{i=1}^{N} |r_i|}{N} \tag{3.27}$$

In [47], [86], MAE is used to compare the performance of different NBMs. MAE is less sensitive to the outliers compared to RMSE and MSE because of the lack of the exponent in the error.

- Mean absolute percentage error (MAPE), which gives an indication of the error in relation to the magnitude of the actual value and is calculated as

$$MAPE = \frac{1}{n} \sum_{i=1}^{N} \left| \frac{r_i}{y_i} \right| \tag{3.28}$$

The MAPE is used in the literature to compare different ANNs models [86].

- The standard deviation of the error (STDE) is calculated as

$$STDE = \sqrt{\frac{\sum_{i=1}^{N} (r_i - \bar{r})^2}{N-1}} \tag{3.29}$$

where $\bar{r}$ is the mean residual. The STDE is used in some studies, such as [47], [62], as an indication of the performance of the model.

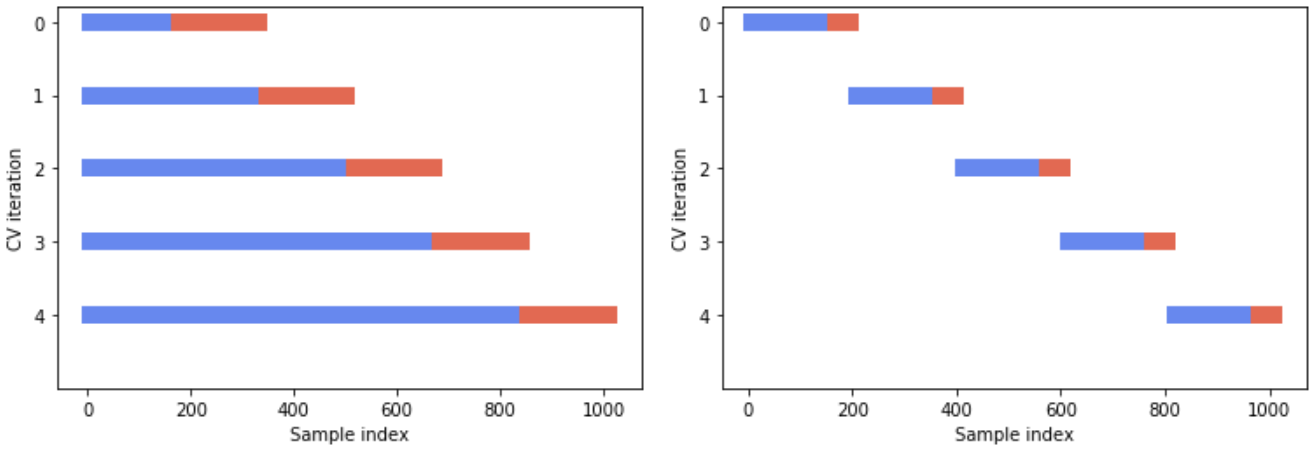
- Akaike information criterion (AIC) which is calculated as

$$AIC = N \cdot ln(STDE^2) + 2k \tag{3.30}$$

where k is the number of model parameters. AIC is designed to strike a balance between the goodness of fit of a model and its complexity, making it a valuable tool for choosing the most appropriate model among a set of candidates. It penalises complex models, discouraging overfitting while favouring models that effectively capture the underlying patterns in the data. AIC provides a quantitative way to compare models and select the one that offers the best trade-off between fit and simplicity. Lower AIC values indicate better-fitting models, and the criterion is particularly valuable when comparing models with different numbers of parameters.

A method to assess the effectiveness of a model is cross-validation (CV). It involves partitioning a dataset into multiple subsets, known as folds. Each fold includes a training and a testing dataset. This approach provides valuable insights into how well a model generalises to new, unseen data and helps in detecting potential overfitting [87]. Overfitting occurs when a model performs exceptionally well on the training data but poorly on unseen data. One popular variant of cross-validation is

time series cross-validation, which is particularly useful for evaluating models using temporal data. Time series CV makes sure that the data are always in chronological order and the testing set is after the training set which is not always the case for other types of CV [88]. The two types of time series CV are time-series split (TSS) and blocking time-series split (BTSS). A representation of how the folds in each method are generated is shown in Figure 3.3. The blue lines represent the training period and the red lines represent the testing period. In TSS CV, the training length is gradually increased, while in BTSS CV the training length remains constant. In the case of the BTSS, the training data can have overlaps between different CV folds. After the training and testing periods are set, the metric used for the model performance is calculated for each fold. Then, the mean value of the metric from the different folds is calculated as well as its standard deviation.



(a) Time-series split

(b) Blocking time-series split

Figure 3.3: Example of time-series split and blocking time-series split [89].

In this study, CV is used to investigate the effect of changing training and testing periods on both the performance of the model and the values of the calculated parameters coming from the least squares fitting of the equation for the calculation of the gearbox oil temperature to the SCADA data.

3.4 Artificial Neural Networks

ANNs are widely used in the literature for the construction of NBMs with some examples mentioned subsection 2.3.3 of the literature review. In this study, 2 ANN models, namely the NARX-ANN and the LR-ANN, are employed for comparative analysis with the physics-based NBM developed in this research. This section gives an introduction to ANNs and then provides information about the two ANN models used in this study.

ANNs are computational models inspired by the way our brains process information. They are at the heart of modern machine learning and artificial intelligence, revolutionising our ability to solve complex problems. Just as our brains consist of interconnected neurons, ANNs comprise artificial "neurons". These artificial neurons are interconnected in layers, forming the basic architecture of an ANN and work together to understand data, make decisions, and solve tasks [90].

The architecture of an ANN comprises layers of interconnected nodes, or neurons, organised into three primary types of layers: the input layer, one or more hidden layers, and the output layer. Neurons in the input layer receive and process the initial data. The connection between neurons carries a weight that determines the strength of the connection. Neurons in the hidden layers perform calculations by taking a weighted sum of their inputs and passing it through an activation function, which introduces nonlinearity and allows the network to model intricate patterns and dependencies in the data. The architecture of an ANN is determined by factors such as the number of layers, the number of neurons in each layer, the type of activation functions used, and the interconnections between neurons. An example of a ANN architecture with one hidden layer is shown in Figure 3.4.

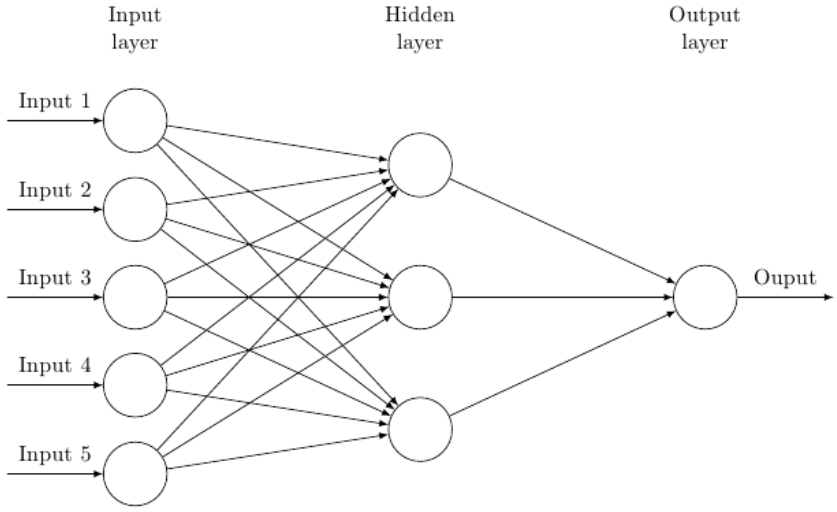


Figure 3.4: ANN architecture example

The two main categories of ANNs are the feedforward ANNs and the recurrent ANNs. Feedforward neural networks are the most basic and widely used type of neural network. The information flows in one direction, from the input layer through the hidden layers to the output layer. On the other hand, recurrent neural networks are designed to handle sequential data where the order of input is important. They include recurrent connections that allow information to cycle through the network.

In each layer of a neural network, a parameter known as the activation function plays a crucial role. This function is responsible for transforming input values into output values, which then become the input for the subsequent layer. These functions decide whether a neuron should become active, influence the extent to which signals progress through the network, and ultimately shape the final prediction. There is a variety of activation functions to choose from, each serving a specific purpose based on the layer's function. These functions include, among others, the threshold function, the sigmoid function, and the rectifier functions[91]. The two activation functions used in this study are the hyperbolic tangent used in the hidden layer and the pure linear in the output layer and are displayed in Figure 3.5.

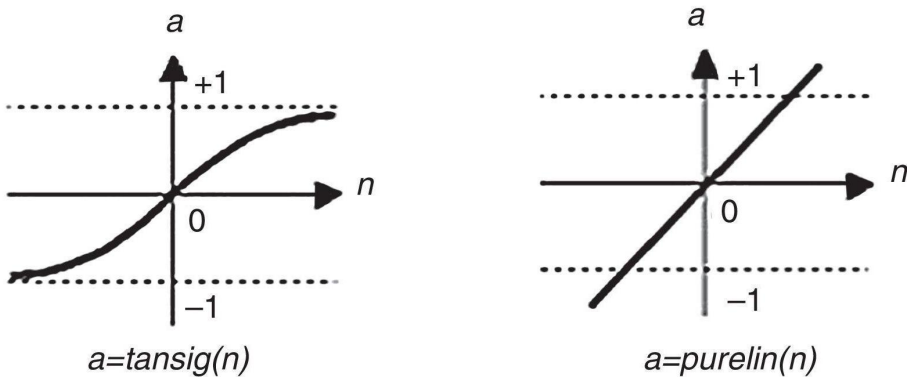


Figure 3.5: ANN architecture example

Based on a given architecture, the ANN is trained using the input and target data. Training ANNs is a fundamental process involving iterative learning from data. It encompasses presenting the network with input data, propagating it through the network, comparing predictions with actual target values, and fine-tuning weights to minimise prediction errors. This iterative training equips ANNs with the capability to make increasingly accurate predictions.

Overall, the main advantages and disadvantages of ANNs are [92]:

Advantages:

- Pattern Detection: Neural networks excel at uncovering subtle patterns within complex multivariate data.
- Flexibility: They offer versatility, suitable for both regression and classification tasks, making them adaptable to a wide range of problems.
- Robustness: Neural networks are robust when it comes to noisy training data. They can handle errors or missing values without significantly affecting their output.
- Non-Linearity: They are well-suited for handling non-linear data, and can accommodate various input dimensions and layer configurations.

Disadvantages:

- Hardware dependency: ANNs, especially the more complex architectures, require heavy computational resources for training, resulting in high costs and time requirements.
- Transparency: Due to their "black-box" nature, the decision-making process of ANNs is less transparent and their predictions are less explainable compared to other approaches.
- Data demands: ANNs demand substantial amounts of data for effective training, and tuning their parameters to achieve optimal performance can be challenging and resource-intensive.

Two models are used in this study: the LR-ANN and the NARX-ANN. The decision to use the LR-ANN was based on [47] which compares several data-driven NBMs and shows that LR-ANN has the best performance even though the performance is not consistent for all wind turbines that are tested. NARX-ANN is widely used in the literature on NBMs, such as in [51], [93], [94], and has demonstrated superior performance compared to feed-forward networks. In addition, these studies provide sufficient information to replicate the methods used, which is not always the case for data-driven NBMs found in the literature.

LR-ANNs are a specialised variant of recurrent neural networks designed to capture sequential dependencies and temporal patterns in data. Unlike feed-forward networks, where information flows only in one direction, LR-ANNs include recurrent connections within individual layers, enabling them to model dynamic sequences and time-dependent patterns [95]. They are used in tasks involving sequential data, such as natural language processing, speech recognition, and time series prediction. The recurrent connections enable them to retain information from previous time steps and use it to influence future predictions. Also, the use of this delayed feedback makes it possible to take into account the system inertia [47].

Figure 3.6 shows the architecture of a LR-ANN. In the input Layer, the network receives inputs and a delay is applied to these inputs (D). Each input is individually multiplied by its corresponding weight (W). The recurrent connection incorporates a delay and they are associated with specific weights. The output of the hidden layer at a given time step is used as input D time steps later, effectively introducing a temporal memory mechanism. The inputs and the recurrent signal are combined within the hidden layer. This combination involves summation, where the inputs and recurrent signals are added. Additionally, a bias term (b) is introduced. The resulting sum is then transformed using the activation function of the hidden layer, which is responsible for introducing nonlinearity into the network. The output of the hidden layer is further processed. It is multiplied by a weight specific to the output layer and combined with an output layer bias. Similar to the hidden layer, this sum is subjected to an activation function, gives the network's final output.

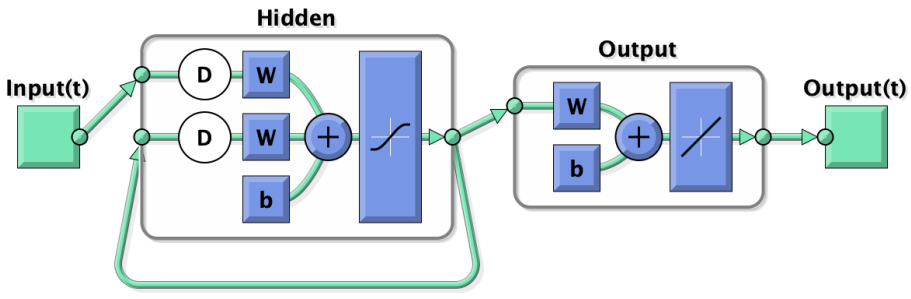


Figure 3.6: LR-ANN architecture [96]

The NARX neural network is a specialised type of recurrent neural network designed for modelling and predicting time series data. It is particularly useful for problems where past observations of the target sequences are required to make predictions. In addition to using past observations of the target variable, NARX networks can also incorporate exogenous inputs, which are external factors that influence the prediction. This makes them versatile for modelling complex systems influenced by multiple variables [51]. Figures 3.7 and 3.8 reveal two distinct types for the NARX ANN model: the series-parallel architecture, also known as the open-loop configuration, and the

parallel architecture, alternatively referred to as the closed-loop configuration.

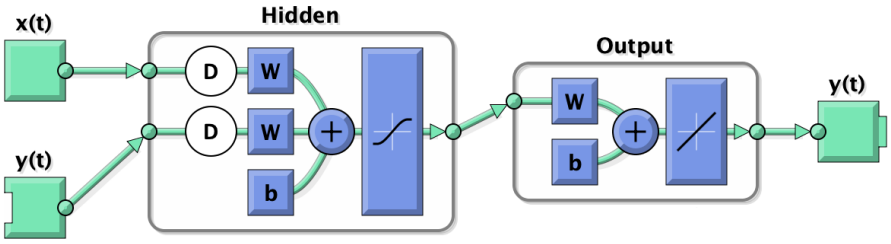


Figure 3.7: Series Parallel NARX-ANN architecture (open-loop) [97]

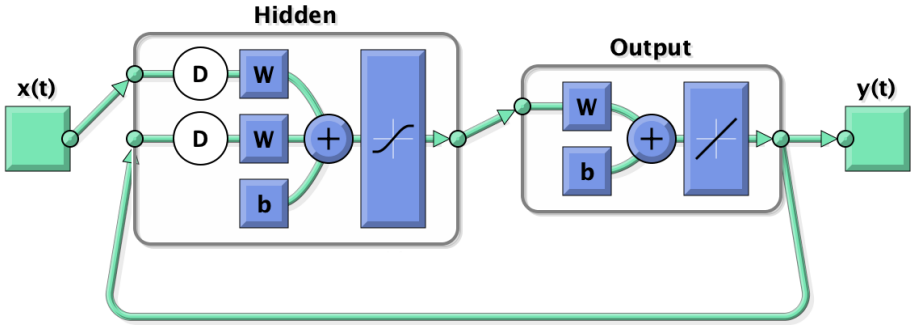


Figure 3.8: Parallel NARX-ANN architecture (closed-loop) [97]

In open loop configuration, the time history of the target variable is used as input for the ANN. The delay defines which previous values are available as inputs. On the other hand, in the closed loop configuration, the past predicted value of the output is used as input. This is the only difference between the two configurations. The inputs after the implementation of delays are multiplied with their corresponding weights and then summed. The bias of the hidden layer is also added and the sum is transformed using the activation function. The output of the hidden layer is multiplied with a weight and the bias is added. Finally, using the transfer function of the output layer, the sum is transformed and the output is calculated.

Chapter 4

Methodology

The main objective of this thesis is to design a physics-based NBM of the gearbox oil temperature. This model is based on the first law of thermodynamics and considers the heat generated in different components of the gearbox as well as the heat dissipated to the environment. As input for the model, SCADA data are used. Due to some unidentified technical specifications of the wind turbine gearboxes, the equation for the calculation of the oil temperature has some unknown parameters. These parameters are calculated by implementing a regression analysis to fit the equation to historic SCADA data. The estimated temperature describes the behaviour of the oil temperature under healthy conditions and can be used for condition monitoring purposes.

An overview of the methodological approach undertaken in this research is shown in Figure 4.1. The process starts with the pre-processing of the available SCADA data which includes data cleaning and data partitioning which are explained in Section 4.1. Then, the gearbox NBM for the calculation of the oil temperature under healthy operating conditions is developed. Section 4.2 includes the derivation of the gearbox oil temperature equation as well as the estimation of the unknown parameters of the equation. The testing dataset is then used to evaluate the effectiveness of the model in predicting the gearbox oil temperature. Section 4.3 includes the different methods used to evaluate the performance of the model. Finally, Section 4.4 includes the description of the ANN NBMs used for a comparative study with the physics-based model developed in this research.

In this research, Python is selected as the primary programming language for the development and analysis of the model. Python's versatility and extensive ecosystem of scientific libraries make it a suitable choice for the tasks involved in this study. The flexibility of Python allows us to implement, customise, and optimise our model efficiently. However, the ANN models used in the comparative study are developed in Matlab which offers a user-friendly toolbox.

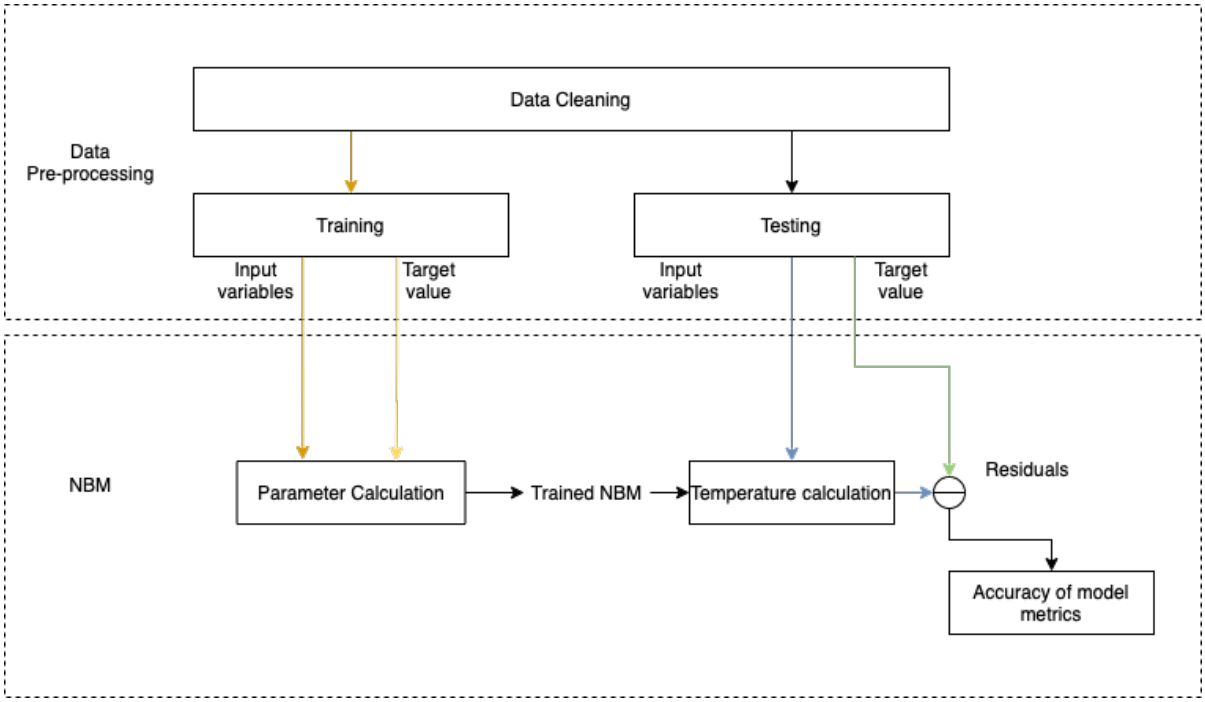


Figure 4.1: Overview of the Methodology

4.1 Pre-processing of the SCADA data

The pre-processing steps undertaken in this study are integral to ensuring the quality, reliability, and relevance of the data used for analysis and modelling. Outlier removal, guided by the expected range of each variable found in [62], is an essential measure to eliminate data points that may introduce noise and inaccuracies into the results. By removing outliers, the dataset remains a faithful representation of the real-world conditions under examination. The removal of data vectors containing NaN values is important for maintaining data integrity. Missing values can lead to inconsistencies and cause problems in calculations, justifying their exclusion. In addition, any negative values of power output are replaced with zeros. This serves to align the data with the physical constraints of the system, avoiding misleading results in calculations, particularly those involving the losses calculation in the gearbox. Lastly, data partitioning into training and testing sets is crucial to prevent information leakage and to assess model performance rigorously. It ensures that models are evaluated on previously unseen data, providing a robust measure of their effectiveness in real-world applications. Together, these pre-processing steps uphold the credibility and robustness of the data, underpinning the quality of subsequent analyses and findings.

4.2 Physics-based normal behaviour model

The NBM calculates the oil temperature, based on its input signals coming from the SCADA system, and the output of the model is then compared to the corresponding temperature measurements from the sensor placed on the wind turbine gearbox. Figure 4.2 shows a schematic of the principle of operation of a NBM.

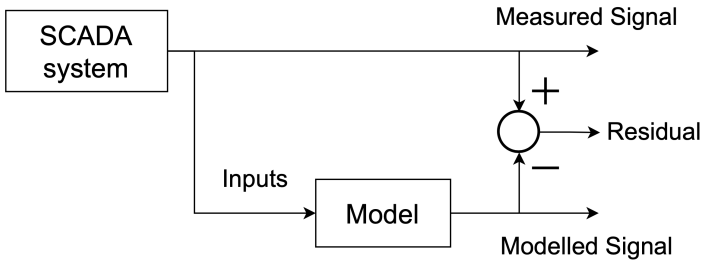


Figure 4.2: Normal Behaviour model schematic

As shown in Chapter 3, several characteristics of a gearbox are necessary to calculate the power losses and the heat dissipated from the gearbox. If all the information is available, it is possible to formulate an equation that using specific input calculates the gearbox oil temperature. However, this study focuses on a different scenario where this information is not available and only relies on information available in SCADA data. Regardless of the missing characteristics, an equation for the gearbox oil temperature can still be formulated, however, in this case there will be some unknown parameters in this equation. These parameters are calculated by fitting the equation to the available SCADA data.

The formulation of the equation for the gearbox oil temperature calculation is based on the heat balance equation

$$P_V - \dot{Q}_{diss} = mC\frac{dT_{gear}}{dt} \tag{4.1}$$

which was introduced in Section 3.1. As suggested in [38], the gearbox temperature is assumed to be the same throughout the gearbox and equal to the gearbox oil temperature, due to the high conductivity of the gearbox material so the Equation 4.1 can be written as

$$P_V - \dot{Q}_{diss} = mC\frac{dT_{oil}}{dt} \tag{4.2}$$

P_V and $\dot{Q}_{diss}$ include multiple terms each one including many unknown parameters. In order to reduce the number of those parameters, several of them are grouped together to make the regression problem simpler. In addition, the variables of the equation for the losses need to be expressed as functions of SCADA signals. This is performed in the following subsections.

4.2.1 Gearbox losses

Gear losses

According to subsection 3.1.1, using equations 3.6, 3.7 and 3.8 the load-dependent gear losses can be calculated as

$$P_{VZP} = P_{\text{in}} \cdot 0.048 \left(\frac{F/b}{2 \cdot v_t \cdot sin\alpha_{wt} \cdot \rho} \right)^{0.2} \eta_{\text{oil}}^{-0.05} Ra^{0.25} X_{\text{L}} \cdot H_v \tag{4.3}$$

where b, ρ, R_a, X_L, a_t and α_{wt} are constants as geometrical features of the gears.

It is assumed that the forces applied on the gears are proportional to the torque transferred through the gearbox which is equal to the power output divided by the rotational speed as mentioned in [57]. In addition, the tangential velocity v_t is proportional to the rotational speed of the rotor n_{rotor}. These relations are written as

$$F \propto T \propto \frac{P_{el}}{n_{rotor}} \Rightarrow F = a_F \frac{P_{el}}{n_{rotor}} \tag{4.4}$$

$$\begin{aligned} &v_t = \omega \cdot R_{gear} \propto n_{rotor} \Rightarrow \\ &v_t = a_{vt} \cdot n_{rotor} \text{ and } \omega = a_\omega \cdot n_{rotor} \end{aligned} \tag{4.5}$$

where a_F and a_{vt} are the proportionality constants.

The input power from the rotor is usually not available in SCADA data so the electrical active power output, P_{el}, is used instead. In addition, all the equations regarding the power losses used in this model are given in relation to the rotational speed of the rotor (n_{rotor}) which again is available in the SCADA data and is proportional to the rotational speed of every component in the gearbox.

By applying the proportionality constants of Equations 4.4 and 4.5 in Equation 4.3, using P_{el} instead of P_{in} and substituting Equation 3.14 for the viscosity we get

$$P_{VZP} = P_{el} \cdot 0.048 \left(\frac{a_F P_{el}/(n_{rotor} \cdot b)}{2\rho \cdot sin(\alpha_{wt}) \cdot a_{vt} \cdot n_{rotor}} \right)^{0.2} (\rho_{\text{oil}} \cdot \nu_{\text{oil}})^{-0.05} Ra^{0.25} X_L \cdot H_v \tag{4.6}$$

Now that the load-dependent gear losses are expressed as a function of P_{el} and n_{rotor}, the constants can be grouped using

$$a_1 = 0.048 \cdot \left(\frac{a_F}{2\rho \cdot sin(\alpha_{wt}) \cdot a_{vt} \cdot b} \right)^{0.2} \rho_{\text{oil}}^{-0.05} \cdot Ra^{0.25} X_L \cdot H_v \tag{4.7}$$

where all the constants are positive and α_{wt} is usually 20°, making a_1 positive. So, by substituting Equation 4.7 in Equation 4.6, it can be written as

$$P_{VZP} = a_1 P_{el} \left(\frac{P_{el}/n_{rotor}}{n_{rotor}} \right)^{0.2} \nu_{\text{oil}}^{-0.05} = a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} \tag{4.8}$$

where a_1 is the parameter that is determined by the fitting of the equation for the calculation of gearbox oil temperature to the SCADA data.

The no-load losses P_{VZ0} are calculated using Equations 3.9 and 3.10 as

$$P_{VZ0} = C_{\text{Sp}} C_1 e^{C_2(v_t/v_{t0})} \cdot \omega \tag{4.9}$$

By applying the proportionality constants for the rotational speed and the tangential speed of Equation 4.5, Equation 4.9 can be written as

$$P_{VZ0} = C_{\text{Sp}} C_1 e^{C_2(a_{vt} \cdot n_{rotor}/v_{t0})} \cdot a_\omega \cdot n_{rotor} \tag{4.10}$$

Then, the constants are grouped together using

$$a_2 = a_\omega \cdot C_{sp} \cdot C_1 \quad \text{and} \quad a_3 = \frac{C_2 \cdot a_{vt}}{v_{t0}} \tag{4.11}$$

where all the constants are positive making a_2 and a_3 also positive. Finally, the no-load gear losses are calculated as

$$P_{VZ0} = a_2 \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} \tag{4.12}$$

where a_2 and a_3 are calculated by the fitting of the equation to the SCADA data.

Bearing losses

Similarly with the gear losses, it is assumed that the rotational speed of every bearing in the gearbox is proportional to the rotational speed of the rotor and the force on the bearings is proportional to the torque applied on the shaft of the bearing. So, Equations 4.4 and 4.5 can also be used for the bearing losses. Using the proportionality constants of these equations to express the bearing load-dependent losses as a function of P_{el} and n_{rotor}, Equation 3.12 can be written as

$$P_{VLP} = f_1 \cdot d_m^\beta \cdot \left(a_F \frac{P_{el}}{n_{rotor}} \right)^\alpha \cdot 10^{-3} a_\omega \cdot n_{rotor} \tag{4.13}$$

Then, by assuming one bearing type in the gearbox, the constants can be grouped together using

$$b_1 = f_1 \cdot a_F^\alpha \cdot d_m^\beta \cdot 10^{-3} \cdot a_\omega \quad \text{and} \quad b_2 = \alpha \tag{4.14}$$

where all the constants are positive making b_1 and b_2 also positive. By substituting Equation 4.14 in Equation 4.13 we get the equation for the bearing load-dependent losses as

$$P_{VLP} = b_1 \left(\frac{P_{el}}{n_{rotor}} \right)^{b_2} n_{rotor} \tag{4.15}$$

where b_1 and b_2 are the parameters that are determined by the fitting of the equation to the SCADA data.

From the 2 parts of Equation 3.11 for the no-load losses introduced in subsection 3.1.1, only the second part is used. This can be explained based on the fact that the majority of losses come from the final stage of the gearbox [71] where the rotational speed is at least 250rpm during operation as observed in the SCADA data. Even the rotor that has the lowest rotational speed has a minimum of 10rpm during operation. Also, the viscosity that is considered is from the lubricant ISO VG 320 synthetic Polyalphaolefin which is a usual choice for wind turbines [98]. The temperature of the oil rarely exceeds 60 degrees Celsius and for this temperature the viscosity of the oil is $180 mm^2/s$, while for lower temperatures the viscosity is larger. This means that the product of n and ν_{oil} is always higher than 2000 $mm^2/s \cdot min$.

The rotational speed of the bearing $n[rpm]$ can be converted to $\omega[rad/s]$ as

$$n = \frac{30}{\pi} \omega \tag{4.16}$$

By substituting Equation 4.16 in the second part of Equation 3.11 we get

$$P_{VL0} = 10^{-10} \cdot f_0 \cdot (\nu_{oil} n)^{2/3} d_m^3 \cdot \omega = 10^{-10} \cdot f_0 \cdot d_m^3 \cdot \left(\nu_{oil} \frac{30}{\pi} \omega \right)^{2/3} \omega \tag{4.17}$$

Then, the proportionality constant of Equation 4.5 is applied on Equation 4.17 as

$$P_{VL0} = 10^{-10} \cdot f_0 \cdot d_m^3 \cdot \left(\nu_{oil} \frac{30}{\pi} a_\omega \cdot n_{rotor} \right)^{2/3} a_\omega \cdot n_{rotor} \tag{4.18}$$

and the constants are grouped together using

$$b_3 = 10^{-10} \cdot f_0 \cdot d_m^3 \cdot a_\omega \left(\frac{30}{\pi} a_\omega \right)^{2/3} \tag{4.19}$$

, where all the constants are positive, making b_3 also positive. The equation for the bearings load dependent losses is formulated as

$$P_{VL0} = b_3 \cdot n_{rotor} \cdot (\nu_{oil} \cdot n_{rotor})^{2/3} \tag{4.20}$$

where b_3 is the parameter that is determined by the fitting of the equation to the SCADA data.

Seal losses Just like gear and bearing losses, the rotational speed of the shaft where the seals are attached is considered proportional to the rotational speed of the rotor according to Equation 4.5. By replacing $n[rpm]$ with $\omega[rad/s]$ using Equation 4.16 and substituting the proportionality constant of Equation 4.5 in Equation 3.13, it is modified as

$$P_{VD} = 7.69 \cdot 10^{-6} d_{sh}^2 \cdot n = 7.69 \cdot 10^{-6} d_{sh}^2 \cdot \frac{30}{\pi} \omega = 7.69 \cdot 10^{-6} d_{sh}^2 \cdot \frac{30}{\pi} a_\omega \cdot n_{rotor} \tag{4.21}$$

Then, the constants are grouped together using

$$c = 7.69 \cdot 10^{-6} d_{sh}^2 \cdot \frac{30}{\pi} a_\omega \tag{4.22}$$

where all the constants are positive, making c also positive. Finally, the seal losses are calculated as

$$P_{VD} = c \cdot n_{rotor} \tag{4.23}$$

where c is the parameter that is calculated by the fitting of the equation to the SCADA data.

The total gearbox power loss is then calculated as the sum of the above-mentioned loss components (Equations 4.6,4.12,4.15,4.20,4.23) as

$$\begin{aligned} P_V = {} & a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + a_2 \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} + b_1 \left(\frac{P_{el}}{n_{rotor}} \right)^{b_2} n_{rotor} + b_3 \cdot n_{rotor} \cdot (\nu_{oil} \cdot n_{rotor})^{2/3} \\ & + c \cdot n_{rotor} \end{aligned} \tag{4.24}$$

4.2.2 Heat dissipation from the gearbox

As described in subsection 3.1.2, the total heat dissipation from the gearbox is calculated as

$$\dot{Q}_{diss} = \dot{Q}_{nac} + \dot{Q}_{cool} = k_0 \cdot A_{ca}(T_{oil} - T_{nac}) + \dot{m}C_p(T_{oil} - T_{oil,inlet}) \tag{4.25}$$

By grouping the constants using

$$k = k_0 \cdot A_{ca} \text{ and } k_1 = \dot{m}C_p \tag{4.26}$$

,where k_0, A_{ca}, $\dot{m}$ and C_p are all positive constants, making k and k_1 also positive, the equation for the total heat dissipation from the gearbox is written as

$$\dot{Q}_{diss} = k \cdot (T_{oil} - T_{nac}) + k_1 \cdot (T_{oil} - T_{oil,inlet}) \tag{4.27}$$

where k and $k1$ are the parameters that are determined by the fitting of the equation to the SCADA data.

4.2.3 Gearbox oil Temperature Calculation

The gearbox oil temperature can be obtained by solving the heat balance Equation 4.2. By substituting Equations 3.5 and 4.25 in Equation 4.2 we get

$$\begin{aligned} mC\frac{dT_{\text{oil}}}{dt} &= P_V - \dot{Q}_{\text{diss}} \Rightarrow \\ mC\frac{dT_{\text{oil}}}{dt} &= P_{VZ0} + P_{VZP} + P_{VL0} + P_{VLP} + P_{VD} - \dot{Q}_{nac} - \dot{Q}_{cool} \end{aligned} \tag{4.28}$$

Then, Equations 4.6, 4.12, 4.15, 4.20, 4.23 and 4.27 are substituted in each term of the right-hand side of the equation to get

$$\begin{aligned} mC\frac{dT_{\text{oil}}}{dt} &= a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + a_2 \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} + b_1 \left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor} \\ &+ b_3 \cdot n_{rotor}(\nu_{oil} \cdot n_{rotor})^{2/3} + c \cdot n_{rotor} - k \cdot (T_{\text{oil},t} - T_{nac,t}) - k_1 \cdot (T_{\text{oil},t} - T_{\text{oil,inlet},t}). \end{aligned} \tag{4.29}$$

SCADA data are sampled at 10-minutes intervals ($\Delta t = 10$ min). To adapt the original continuous-time equation with dT/dt to discrete-time modelling, we introduce the discrete-time counterpart with $\Delta T/\Delta t$ as

$$\begin{aligned} mC\frac{T_{\text{oil},t} - T_{\text{oil},t-1}}{\Delta t} &= a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + a_2 \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} + b_1 \left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor} \\ &+ b_3 \cdot n_{rotor}(\nu_{oil} \cdot n_{rotor})^{2/3} + c \cdot n_{rotor} - k \cdot (T_{\text{oil},t} - T_{nac,t}) - k_1 \cdot (T_{\text{oil},t} - T_{\text{oil,inlet},t}) \end{aligned} \tag{4.30}$$

This conversion is based on the assumption that over each 10-minute interval (Δt), the system's behaviour can be represented by a constant rate of change ($\Delta T/\Delta t$). By solving the equation for

$T_{\text{oil},t}$ we get

$$T_{\text{oil},t} = (\frac{mC}{\Delta t} T_{\text{oil},t-1} + k \cdot T_{nac,t} + k_1 \cdot T_{\text{oil,inlet},t} + a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + a_2 \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} + \\ + b_1 \left(\frac{P_{el}}{n_{rotor}} \right)^{b_2} n_{rotor} + b_3 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + c \cdot n_{rotor}) / (\frac{mC}{\Delta t} + k + k_1) \tag{4.31}$$

The heat capacity and the mass of the gearbox are considered constant while $\Delta t = 10$ min. So, they can be grouped together using

$$a = \frac{mC}{\Delta t} \tag{4.32}$$

, where m c and Δt are positive constants making a also positive. The equation for gearbox oil temperature can be written as

$$T_{\text{oil},t} = \frac{a}{a+k+k_1} \cdot T_{\text{oil},t-1} + \frac{k}{a+k+k_1} \cdot T_{nac,t} + \frac{k_1}{a+k+k_1} \cdot T_{\text{oil,inlet},t} + \\ + \frac{a_1}{a+k+k_1} \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + \frac{a_2}{a+k+k_1} \cdot n_{rotor} \cdot e^{a_3 \cdot n_{rotor}} + \frac{b_1}{a+k+k_1} \left(\frac{P_{el}}{n_{rotor}} \right)^{b_2} n_{rotor} + \\ + \frac{b_3}{a+k+k_1} \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + \frac{c}{a+k+k_1} \cdot n_{rotor} \tag{4.33}$$

To further simplify the equation, the parameters β_1-β_{10} are introduced as

$$\beta_1 = \frac{a}{a+k+k_1} \ , \ \beta_2 = \frac{k}{a+k+k_1} \ , \ \beta_3 = \frac{k_1}{a+k+k_1} \ , \ \beta_4 = \frac{a_1}{a+k+k_1} \ , \ \beta_5 = \frac{a_2}{a+k+k_1} \\ \beta_6 = a_3 \ , \ \beta_7 = \frac{b_1}{a+k+k_1} \ , \ \beta_8 = b_2 \ , \ \beta_9 = \frac{b_3}{a+k+k_1} \ , \ \beta_{10} = \frac{c}{a+k+k_1} \tag{4.34}$$

and Equation 4.33 is formulated as

$$T_{\text{oil},t} = \beta_1 \cdot T_{\text{oil},t-1} + \beta_2 \cdot T_{nac,t} + \beta_3 \cdot T_{\text{oil,inlet},t} + \beta_4 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \\ + \beta_7 \left(\frac{P_{el}}{n_{rotor}} \right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + \beta_{10} \cdot n_{rotor} \tag{4.35}$$

In order to enforce the constrain

$$\beta_1 + \beta_2 + \beta_3 = \frac{a}{a+k+k_1} + \frac{k}{a+k+k_1} + \frac{k}{a+k+k_1} = \frac{a+k+k_1}{a+k+k_1} = 1 \Rightarrow \\ \beta_1 + \beta_2 + \beta_3 = 1 \tag{4.36}$$

the final equation is written as

$$T_{\text{oil},t} = (1-\beta_2-\beta_3)T_{\text{oil},t-1} + \beta_2 \cdot T_{nac,t} + \beta_3 \cdot T_{\text{oil,inlet},t} + \beta_4 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \\ + \beta_7 \left(\frac{P_{el}}{n_{rotor}} \right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + \beta_{10} \cdot n_{rotor} \tag{4.37}$$

4.2.4 Parameter estimation

With Equation 4.37, formulated in subsection 4.2.3, the gearbox oil temperature can be calculated if all the design specification of the gearbox is available to calculate parameters β_1-β_{10}. However, this requires an in depth knowledge of the gearbox design which is rarely available. To account for the unidentified gearbox specifications which lead to the unknown parameters β_1-β_{10}, the equation is fitted to historic SCADA data. The data used for fitting the equation is the training data selected during the data pre-processing.

The independent variables of this regression problem are:

- Gearbox oil temperature at the previous timestep ($T_{oil,t-1}$)
- Nacelle temperature ($T_{nac,t}$)
- Gearbox inlet oil temperature ($T_{\text{oil,inlet},t}$)
- Electrical power output (P_{el})
- Rotor rotational speed (n_{rotor})
- Oil kinematic viscosity (ν_{oil})

From these, only the oil kinematic viscosity can not be found in the SCADA system but it can be calculated using Equation 3.16. The dependent variable of the regression problem is the temperature of the oil sump in the gearbox $T_{oil,t}$.

The least squares method is employed to perform regression and determine the best-fitting parameters for Equation 4.37. This is accomplished using the least squares function from the 'optimize' package in the SciPy library. The function is capable of solving non-linear least squares problems while considering boundaries on the parameters, and it employs the STIR algorithm for optimisation which is described in Section 3.2.

The parameters a, k, k_1, a_1, a_2, a_3, b_1, b_2, b_3 and c of Equation 4.34 are all positive as described in Sections 4.2.2, 4.2.1 and 4.2.3. This means that also the parameters β_1-β_{10} are positive as fractions of positive parameters. Therefore, the lower bound for all the unknown parameters is set to zero.

4.3 Model Evaluation

An important part of this study is to evaluate the performance of the developed physics-based normal behaviour model of the gearbox oil temperature. After the unknown parameters of Equation 4.37 are calculated using the training data, it is possible to calculate the modelled oil temperature and compare it to the actual oil temperature using the residual as shown in Equation 3.24. The model evaluation comprises four distinct approaches.

1. Computation of various metrics to assess the model's goodness of fit. The metrics used are the RMSE, the MAPE, the STDE and the MSE which are described in Section 3.3. The values of those metrics are also compared between the testing and the training period to identify potential overfitting which occurs when a model performs exceptionally well on the training data but poorly on unseen data. Finally, the average time to run the fitting is calculated to compare it with other methods.

2. Reduction of number of parameters in the equation, followed by a comparison using the AIC, described in Section 3.3, to determine the most suitable equation for the given dataset. A parsimonious model is one that accomplices the desired performance with the fewest possible parameters. Here the parameters obtained by the fitting of Equation 4.37 to different turbine datasets are analysed. The parameters that obtain negligible values are excluded. Then, AIC is used to examine how the model performance is affected by the reduction of the number of parameters.

3. Cross-validation to explore the model's performance under varying training periods and period lengths as described in Section 3.3. Using the cross-validation as the third step, it is possible to identify the robustness of the model. This is achieved by calculating the parameters of Equation 4.37 for a given turbine under different training periods and analysing their variability. Also the metrics of step one are calculated in each fold of the cross-validation.

4. Verification of the physical soundness of the parameters obtained from the regression analysis. This step is considered necessary, as this model is a physics-based NBM and the goal of the regression is to calculate the unknown parameters that represent the physical characteristics of the gearbox. To accomplish this, the parameters of Equation 4.37 are calculated for a wind turbine, for which detailed design information is provided in [71]. Some parameters are calculated using also information from [38]. Then these parameters are compared with the parameters derived from the regression analysis.

4.4 Comparison with ANN models

Another way to assess the performance of the model is to compare it with results from NBMs proposed in other studies. However, every study uses a different dataset which has an effect on the performance of each method. With this in mind, the same dataset has been used in two NBMs proposed in the literature and the results compared with those of the model proposed in this study. The two models that are replicated are the LR-ANN and the NARX-ANN model described in Section 3.4. The ANN-based NBMs use the same input signals used in the physics-based model developed in this study while the training and testing periods are also the same.

The architecture of the ANN models is chosen based on information found in literature. The architecture of LR-ANN model shown in Figure 4.3 includes a hyperbolic tangent sigmoid (tansig) activation function in the hidden layer and a linear activation function for the output layer. 6 neurons are used in the hidden layer while a delay of two steps is set for recurrence similarly with [47]. The architecture of NARX-ANN model shown in Figure 4.4 includes the same activation functions with the LR-ANN and has 10 neurons in the hidden layer similarly with [51]. In the NARX model, the autoregression element can either be the time history of the target variable using a delay or the calculated output of the model using a delay so, both of these were tested. This delay is set as one timestep while no delay is used for the exogenous inputs.

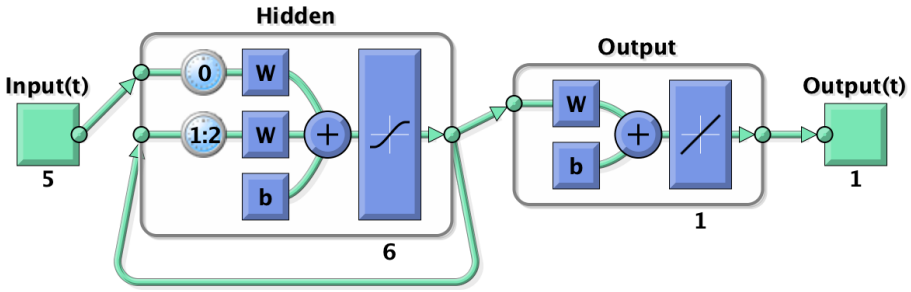


Figure 4.3: LR-ANN architecture in this study

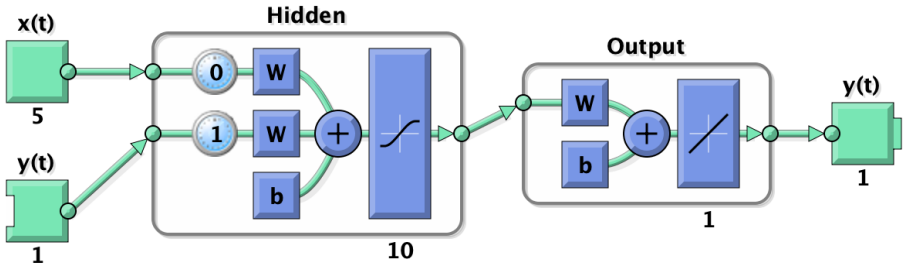


Figure 4.4: NARX-ANN architecture

The ANN models are developed in Matlab using the functions layrecnet and narxnet for LR-ANN and NARX-ANN, respectively. The pre-processing of the data for the ANNs is similar to that of the physics-based model mentioned in Section 4.1. In addition for the ANN models, it is necessary to normalise the input signals. The training period is kept the same with the physics-based model however, it is randomly split between 80% train and 20% validation data, respectively as it is necessary for the ANNs.

The metrics used for the comparison of the models are RMSE, MAPE, STDE and MSE obtained for the testing period, while the time required for training the ANN models is compared to the time required for fitting the equation for the calculation of the gearbox oil temperature (Equation 4.37) to the training data of the physics-based model.

Chapter 5

Results

In this chapter, the results of this study are described and analysed. They are divided into two case studies that refer to two different datasets used in this project. For each case study the dataset is described and the normal behaviour model described in Chapter 4 is applied.

5.1 Case study 1

This case study refers to the data from the Penmanshiel Wind Farm released by Cubico Sustainable Investments Ltd [99]. This is a publicly available dataset which includes SCADA data from June 2016 to June 2021 from 14 wind turbines.

5.1.1 Data-set description

The 14 wind turbines placed in Penmanshiel (UK) are 2.05MW MM82 machines from Senvion with a wind class of IEC Ia. Also, it is known that their gearbox has three stages [100]. The initial plan was to install 15 wind turbines however turbine T03 was removed after an appeal [101]. The available dataset includes 299 10-minute averaged SCADA signals in total. For some signals, the minimum, the maximum, and the standard deviation for each 10-minute period is included as well. The signals used in this study are:

- Average temperature of oil in gearbox (T_{oil})
- Average temperature in nacelle ($T_{nac,t}$)
- Average rotor rpm (n_{rotor})
- Average temperature of oil inlet in gearbox ($T_{\text{oil,inlet},t}$)
- Total active power (P_{el})

These are the signals required in Equation 4.37 for the calculation of the gearbox oil temperature.

Even though the dataset includes data from 2016 to 2021, some of the signals required for the model are available only from April 2018 to June 2021. This means that only this period is usable for this study. Figure 5.1 shows the periods when all the required signals for this study are available and when they are missing. By removing the data with missing values, the number of data points reduces from 266435 to 156995-160684 depending on the turbine. In addition to the SCADA signals,

status files from each turbine are also included in the dataset. These files have messages generated automatically from the wind turbine which can be warnings and errors that occur. However, no information about the maintenance or failures of the wind turbines is available.

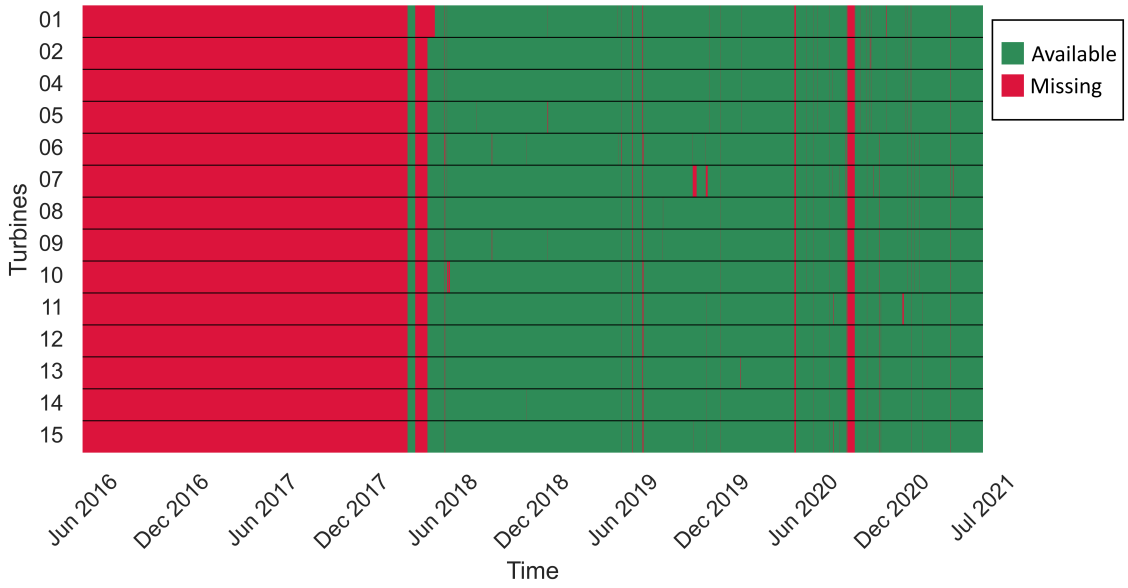


Figure 5.1: Periods of missing and available signals for Penmanshiel dataset

For two of the turbines (T09 and T10), a change in the behaviour of the cooling system has been observed. As shown in Figure 5.2 at the end of November 2019, the range of temperatures in which the cooling system operates changes suddenly with the vertical line indicating the time of the change. Before this incident, there are several manual stops for both turbines recorded in the log data which is an indication that something was changed in the cooling system. Because of this change, those 2 turbines will not be included in this study.

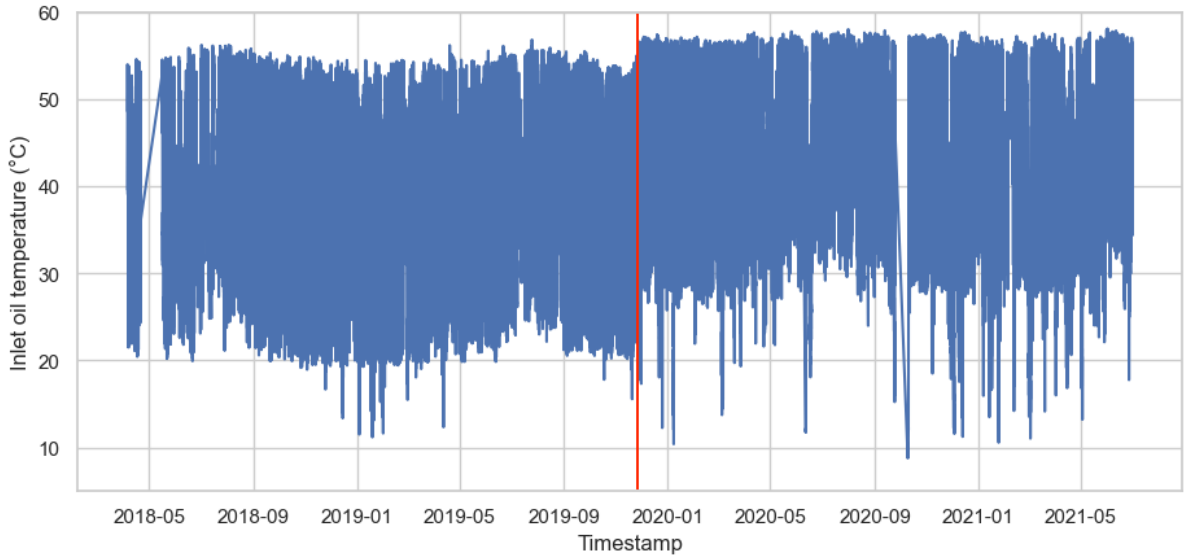


Figure 5.2: Oil inlet temperature for turbine T09 over time

5.1.2 Normal behaviour model application and evaluation

In this dataset, there is no information about the turbine maintenance and failure occurrence. In the log data there is no recorded warning regarding the gearbox so the wind turbines are considered healthy. The first year of available data is used as the training period to fit the Equation 4.37 for the calculation of the oil temperature to the data, while one year of data is also used for the testing period. The evaluation of the model is performed in four ways as described in Section 4.3. First, some metrics that determine the goodness of fit are calculated. Furthermore, the number of parameters of the equation are reduced to identify the most parsimonious model using the AIC. Then, a cross-validation is performed to evaluate the model performance at different periods and for different period lengths. Finally, the calculated parameters from the model are compared to the parameters from a gearbox with known characteristics.

Once the equation is fitted to the data of the training period, the unknown parameters are computed using the least squares technique. Subsequently, the predicted temperature of the gearbox can be determined. The relationship between the actual and modelled temperatures of the gearbox oil for turbine T01 within the testing dataset is illustrated in Figure 5.3 indicating a strong correlation between them.

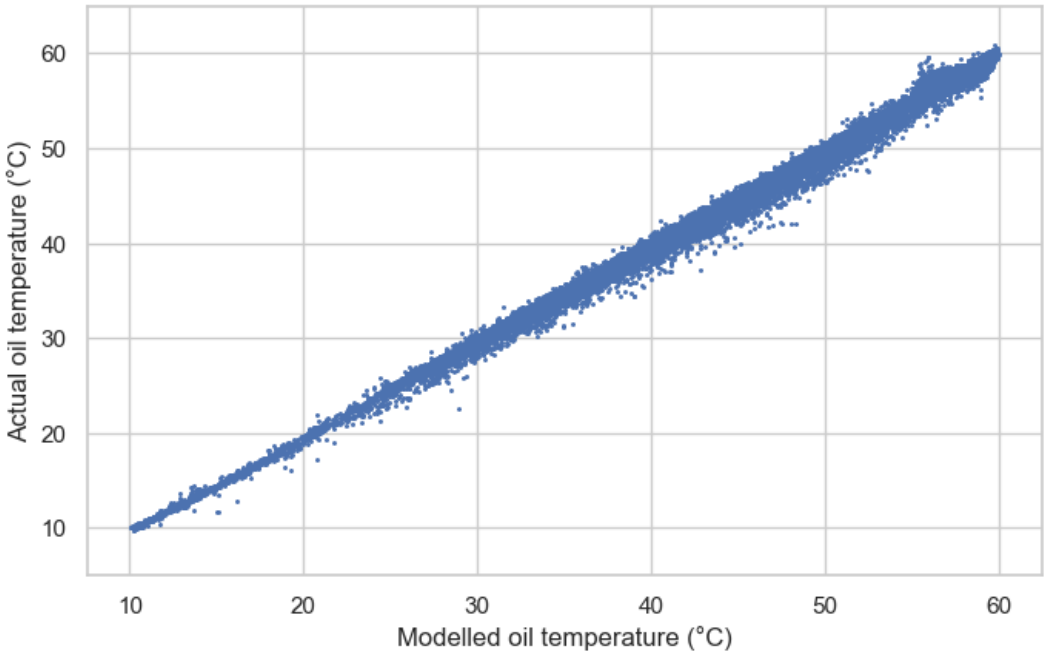


Figure 5.3: Comparison of the modelled and actual temperature of gearbox oil for Penmanshiel dataset

The mean RMSE of all turbines is 0.554 and 0.562 for the training and testing periods, respectively. The small difference (1.33%) between those two values indicates that the model is not overfitting. In addition, for the testing period the MAPE is 0.75% and the STDE is 0.554 and this suggests a good performance for the model. For comparison, previous studies using a NBM for gearbox oil temperature have achieved lower performance as shown in Table 5.1. Even though this comparison gives an indication of the good performance of the proposed model, the other studies use different datasets. Hence, it is also worth comparing the performance with other NBMs while using the same

dataset. This is implemented in subsection 5.1.3. The mean value of the time to fit the equation to the data and calculate the unknown parameters is 9.93 seconds when training on 1 year of data. In Figure 5.4, the RMSE values of the testing period as well as the time required for fitting to training SCADA data are displayed for the different turbines. The training is performed using a laptop with a 2.3GHz quad-core Intel Core i7. The variability in the time to fit can be attributed to the non-linearity of the equation that is fitted. Also, the equation is fitted to the different turbines one after the other and the limited computational power and temperature of the processor might affect the time required for training.

Table 5.1: Comparison of performance metrics of the NBMs of different studies

Study	Performance
[102]	RMSE=2.21
[84]	STDE=0.66
[62]	STDE=2.09
[53]	STDE=2.25

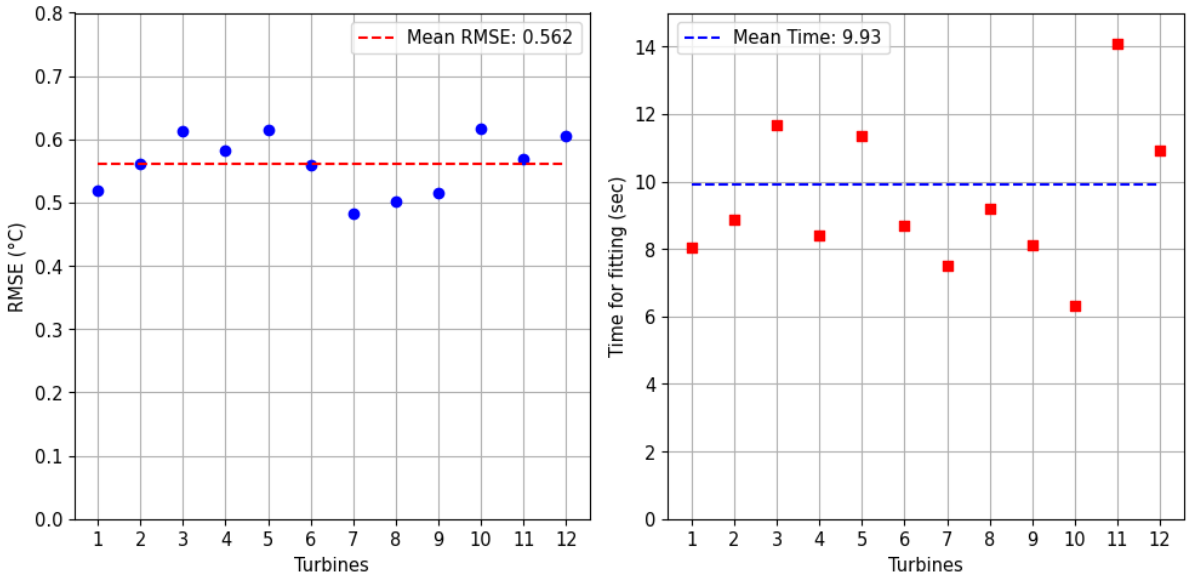


Figure 5.4: Comparison of RMSE and fitting time for different turbines

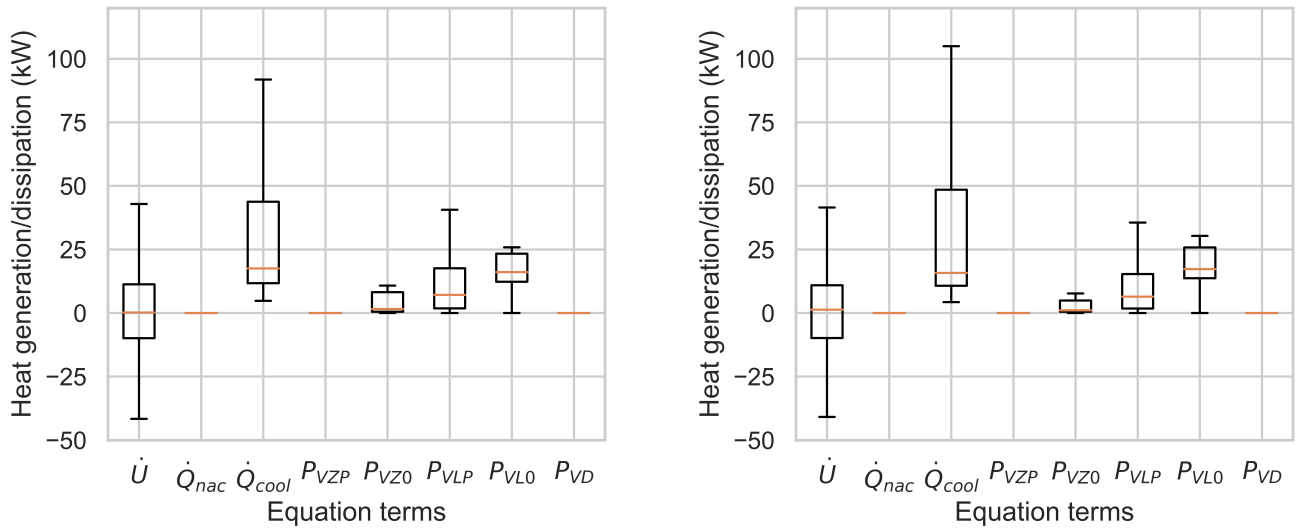
The second step of the model evaluation is to develop a parsimonious model, which is a model that has good performance while using the smallest number of parameters. The values of the parameters of Equation 4.37 obtained from the fitting, are investigated to identify which ones to exclude. The parameters are calculated for all turbines, with the same training period length of one year. The mean values, the standard deviation (σ) and the coefficient of variance (CoV) of each parameter of Equation 4.37 are included in Table 5.2. CoV is the ratio of the standard deviation to the mean. From the table, it is visible that 3 of the parameters have negligible values. These parameters are β_2, β_4 and β_{10}. As for the rest of the parameters, the CoV gets small values which indicates that the calculated parameters are similar for the different turbines. The only exceptions are parameters β_5 and β_7 and this is why further analysis of the robustness of the model is performed for each

turbine individually using cross-validation in the next step of the evaluation. A heatmap with the calculated parameters for all the turbines is shown in Section A.2 in Figure A.1.

Table 5.2: Mean, standard deviation (σ), and coefficient of variation (CoV) of the NBM equation parameters referring to the 12 wind turbines

Parameter	mean	σ	CoV
β_1	$8.2{\cdot}10^{-1}$	$1.0{\cdot}10^{-2}$	$1.2{\cdot}10^{-2}$
β_2	$4.4{\cdot}10^{-11}$	$1.3{\cdot}10^{-10}$	$3.0{\cdot}10^{+0}$
β_3	$1.8{\cdot}10^{-1}$	$1.0{\cdot}10^{-2}$	$5.6{\cdot}10^{-2}$
β_4	$8.3{\cdot}10^{-10}$	$1.6{\cdot}10^{-9}$	$1.9{\cdot}10^{+0}$
β_5	$1.3{\cdot}10^{-5}$	$1.0{\cdot}10^{-5}$	$8.4{\cdot}10^{-1}$
β_6	$4.7{\cdot}10^{-1}$	$5.7{\cdot}10^{-2}$	$1.2{\cdot}10^{-1}$
β_7	$2.8{\cdot}10^{-4}$	$1.8{\cdot}10^{-4}$	$6.5{\cdot}10^{-1}$
β_8	$1.2{\cdot}10^{+0}$	$1.3{\cdot}10^{-1}$	$1.1{\cdot}10^{-1}$
β_9	$3.5{\cdot}10^{-4}$	$5.5{\cdot}10^{-5}$	$1.5{\cdot}10^{-1}$
β_{10}	$1.5{\cdot}10^{-9}$	$2.1{\cdot}10^{-9}$	$1.4{\cdot}10^{+0}$

The values of the parameters do not give a clear indication of the weight of each term of the heat balance in Equation 4.28. So, the heat generation from different sources within the gearbox are calculated along with the heat dissipated from the gearbox. Both the generation and the dissipation depend on the conditions at which the wind turbine operates. So, the values of each term are calculated over a week with various conditions including high and low power generation and a wide range of gearbox oil temperatures. Figure 5.5 shows the distribution of heat generation (P), heat dissipation ($\dot{Q}$) and the rate of change of internal energy ($\dot{U}$) of the gearbox. It can be observed that the two turbines presented here show similar behaviour. The terms $\dot{Q}_{nac}$, P_{VZP} and P_{VD} with negligible values are those corresponding to parameters β_2, β_4 and β_{10} which also had negligible values. Figure 5.5, in Appendix A.2 shows the behaviour of all 12 turbines which have similar behaviour with turbines T01 and T02. $\dot{U}$ represents the left hand side of Equation 4.28.



(a) Turbine 01

(b) Turbine 02

Figure 5.5: Heat generation and dissipation terms of heat balance Equation 4.28

After this observation, the decision is made to exclude the terms associated to the negligible parameter values. Before modifying the equation for the calculation of the oil temperature, first, the reason behind these low values is investigated. The physical meaning of these parameters is the following:

- β_2: associated to the heat dissipation to the nacelle
- β_4: associated to the load-dependent gear losses
- β_{10}: associated to the losses from seals

While, the corresponding equations are:

- $\dot{Q}_{nac} = k{\cdot}(T_{oil} - T_{nac})$
- $P_{VZP} = a_1 P_{el} \left(\frac{P_{el}/n_{rotor}}{n_{rotor}}\right)^{0.2} \nu_{\text{oil}}^{-0.05} = a_1 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05}$
- $P_{VD} = c{\cdot}n_{rotor}$

The first way to investigate the dependence of the temperature on the different signals is by calculating the cross-correlation between the different inputs and the target signal using the data from all wind turbines. This is shown in Figure 5.6 as a heatmap. The cross-correlation between the oil temperature and the nacelle temperature is very low with a value of $7.6{\cdot}10^{-4}$. This is probably the reason why the value of β_2 obtained by fitting to the SCADA data is negligible.

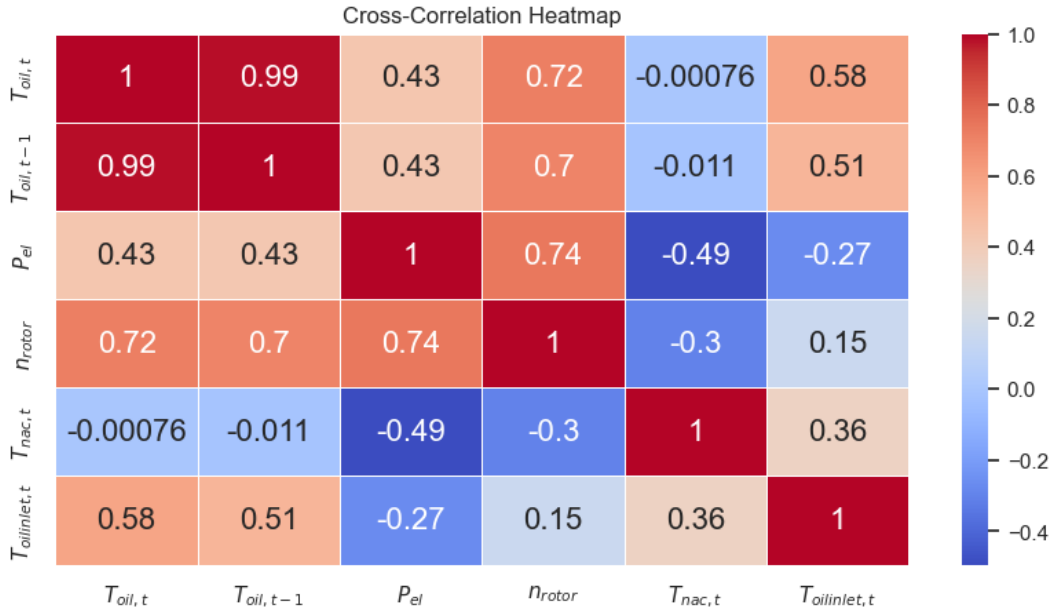


Figure 5.6: Cross-correlation heatmap for the input and target signals with zero lag and the first year of data

A second cross-correlation analysis is conducted, but this time instead of using the input and output signals, the terms of the heat balance Equation 4.28 are used. The terms are:

- Load-dependent gear losses $P_{VZP} = \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{oil}^{-0.05}$
- No-load gear losses $P_{VZ0} = n_{rotor} e^{0.47 \cdot n_{rotor}}$
- Load-dependent bearing losses $P_{VLP} = \left(\frac{P_{el}}{n_{rotor}}\right)^{1.2} n_{rotor}$
- No-load bearing losses $P_{VL0} = n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3}$
- Seal losses $P_{VD} = n_{rotor}$

For the terms of no-load gear losses and load-dependent bearing losses, the mean parameters from Table 5.2 are used in the exponents. The heatmap generated from the cross-correlation matrix is shown in Figure 5.7.

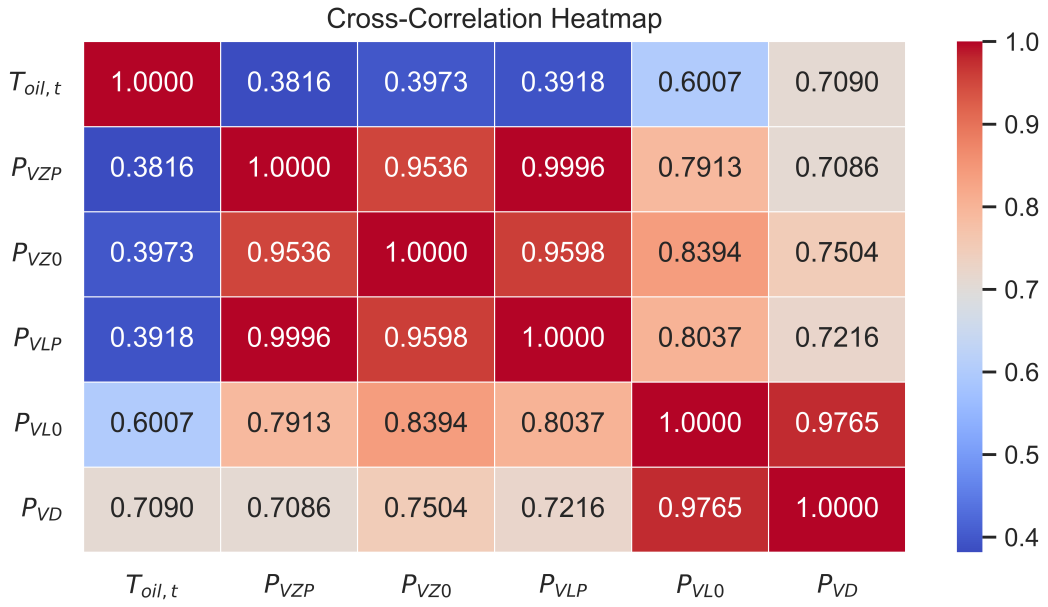


Figure 5.7: Cross-correlation heatmap for terms of the balance equation with zero lag and the first year of data

Figure 5.7 shows that the cross-correlation between the load-dependent gear losses P_{VZP} and the load-dependent bearing losses P_{VLP} is really close to 1 with a value of 0.9996. At the same time, the cross-correlation between $T_{oil,t}$ and load-dependent bearing losses P_{VLP} is slightly higher compared to the cross-correlation between the $T_{oil,t}$ and load-dependent gear losses P_{VZP}. This is an indication why during the fitting only one of the two terms gets value for the parameter which contributes to the calculation of the temperature and it is the one with the higher correlation while the other gets a negligible value.

The parameter β_{10} that describes the seal losses gets a value close to 0 in the fitting. To understand why this occurs, it is important to remember one of the initial assumptions made in this study. It is assumed that all the losses are converted to heat which is first transferred to the oil and then dissipated to the environment through the oil. However, in the case of the seals, they are located at the input and output shaft of the gearbox. This means that the heat generated from the friction can be dissipated directly to the rotating shaft and the nacelle without first being absorbed by the oil. Therefore, it does not result in a significant increase in the oil temperature.

The model is simplified by fitting 5 different equations to the training data. The first one is the initial equation with all the parameters included formulated in subsection 4.2.3 as

$$T_{\mathrm{oil},t} = (1-\beta_2-\beta_3)T_{\mathrm{oil},t-1} + \beta_2 \cdot T_{nac,t} + \beta_3 \cdot T_{\mathrm{oil,inlet},t} + \beta_4 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\mathrm{oil}}^{-0.05} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \tag{5.1}$$
$$+ \beta_7 \left(\frac{P_{el}}{n_{rotor}}\right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + \beta_{10} \cdot n_{rotor}$$

The second one does not include the parameter β_4

$$T_{\text{oil},t} = (1-\beta_2-\beta_3)T_{\text{oil},t-1} + \beta_2 \cdot T_{nac} + \beta_3 \cdot T_{\text{oil,inlet}} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \\ + \beta_7 \left(\frac{P_{el}}{n_{rotor}}\right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} + \beta_{10} \cdot n_{rotor} \tag{5.2}$$

the third one does not include the parameter β_{10}

$$T_{\text{oil},t} = (1-\beta_2-\beta_3)T_{\text{oil},t-1} + \beta_2 \cdot T_{nac} + \beta_3 \cdot T_{\text{oil,inlet}} + \beta_4 \frac{P_{el}^{1.2}}{n_{rotor}^{0.4}} \nu_{\text{oil}}^{-0.05} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \\ + \beta_7 \left(\frac{P_{el}}{n_{rotor}}\right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} \tag{5.3}$$

the fourth equation does not include the parameters β_4 and β_{10}

$$T_{\text{oil},t} = (1-\beta_2-\beta_3)T_{\text{oil},t-1} + \beta_2 \cdot T_{nac} + \beta_3 \cdot T_{\text{oil,inlet}} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \\ + \beta_7 \left(\frac{P_{el}}{n_{rotor}}\right)^{\beta_8} n_{rotor} + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} \tag{5.4}$$

Finally, the fifth one does not include the parameters β_2, β_4 and β_{10}.

$$T_{\text{oil},t} = (1-\beta_3)T_{\text{oil},t-1} + \beta_3 \cdot T_{\text{oil,inlet}} + \beta_5 \cdot n_{rotor} \cdot e^{\beta_6 \cdot n_{rotor}} + \beta_7 \left(\frac{P_{el}}{n_{rotor}}\right)^{\beta_8} n_{rotor} + \\ + \beta_9 \cdot n_{rotor} (\nu_{oil} \cdot n_{rotor})^{2/3} \tag{5.5}$$

Figure 5.8 compares the model performance when using the 5 different equations. In addition, Table 5.3 shows the percentage change of the metrics used to evaluate the performance of the model and the time to run the fitting between the Equations 5.2-5.5 and the initial Equation 5.1 as well as the AIC for the five equations. It is clear that there is almost no difference in the values of the metrics but when using Equation 5.5 which includes the least number of parameters, a 63.5% reduction in time compared to the original equation is achieved. In addition, the equation that provides the largest negative value of AIC is Equation 5.5 which is the equation with the smallest number of parameters. This was expected as AIC penalises based on the number of parameters and in this comparison, the performance remains almost the same but the number of parameters decreases.

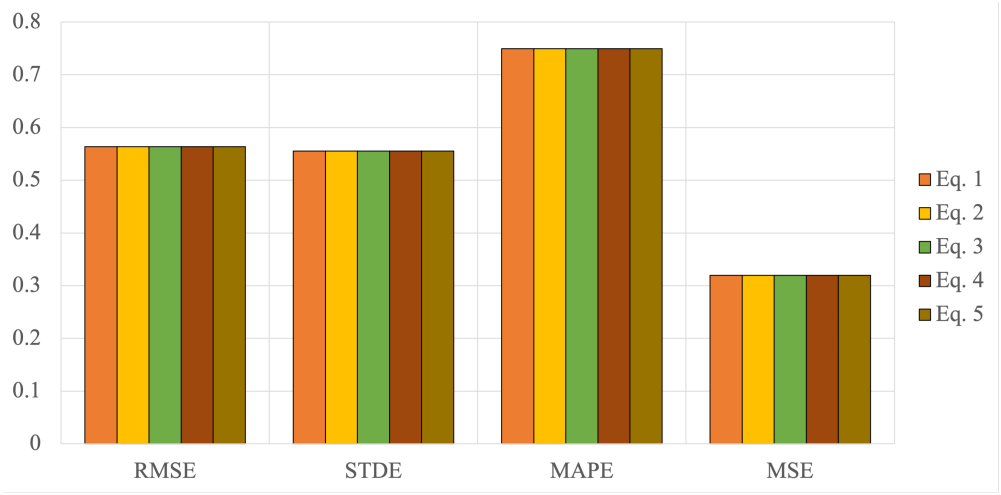


Figure 5.8: Comparison of metrics for different fitting equations

Table 5.3: Comparison of metrics and time to run for equations 5.1-5.5

	RMSE(%)	STDE(%)	MAPE(%)	MSE(%)	Time(%)	AIC
Eq. 5.1	-	-	-	-	-	-74212
Eq. 5.2	-0.0073	-0.0018	0.0167	-0.0156	-54.4	-74223
Eq. 5.3	-0.0022	-0.0006	0.0020	-0.0063	-2.3	-74217
Eq. 5.4	-0.0073	-0.0017	0.0167	-0.0156	-58.9	-74227
Eq. 5.5	-0.0073	-0.0017	0.0167	-0.0156	-63.5	-74225

Taking into account the reduction in time using the equations with the smallest number of parameters and also the fact that there is no difference in the performance of the model it was decided to adopt the Equation 5.5 which does not include the parameters β_2, β_4 and β_{10}. This equation also has the lowest AIC value indicating a more parsimonious model. So, the following steps of this study are utilising this equation.

The third method used to identify the robustness of the model is the time series cross-validation which is explained in Section 3.3. This method partitions the data in subsets keeping them in chronological order. The two CV methods used are the TSS and the BTSS. In the former method, the duration of training is gradually increased starting from 10 months and adding 2 months until the training gets to 24 months. In the latter method, the training length is kept 12 months but it is shifted in time. The testing length is kept equal to one year for both BTSS and TSS. Figure 5.9 illustrates the training and testing periods for BTSS and TSS cross-validation described above.

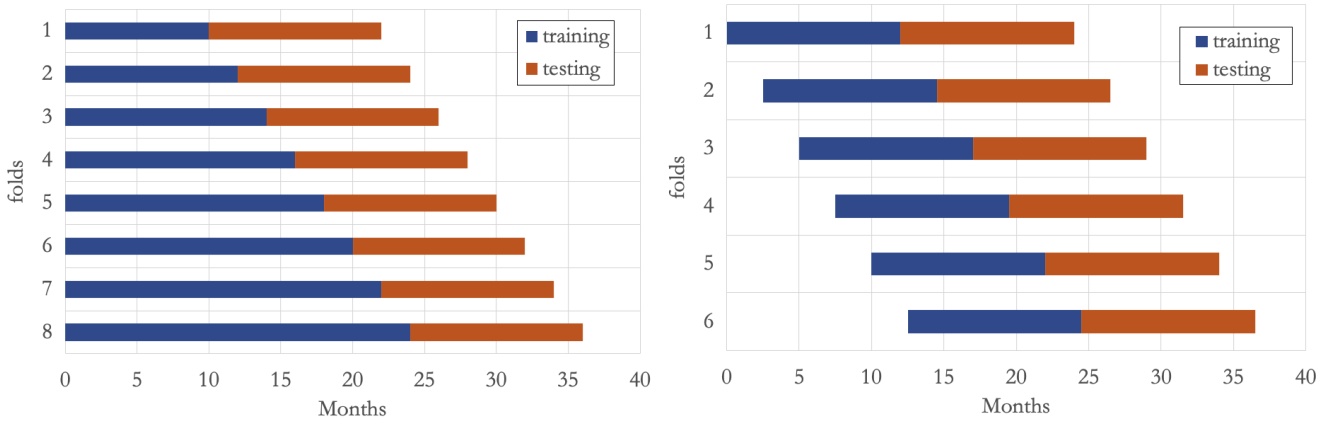


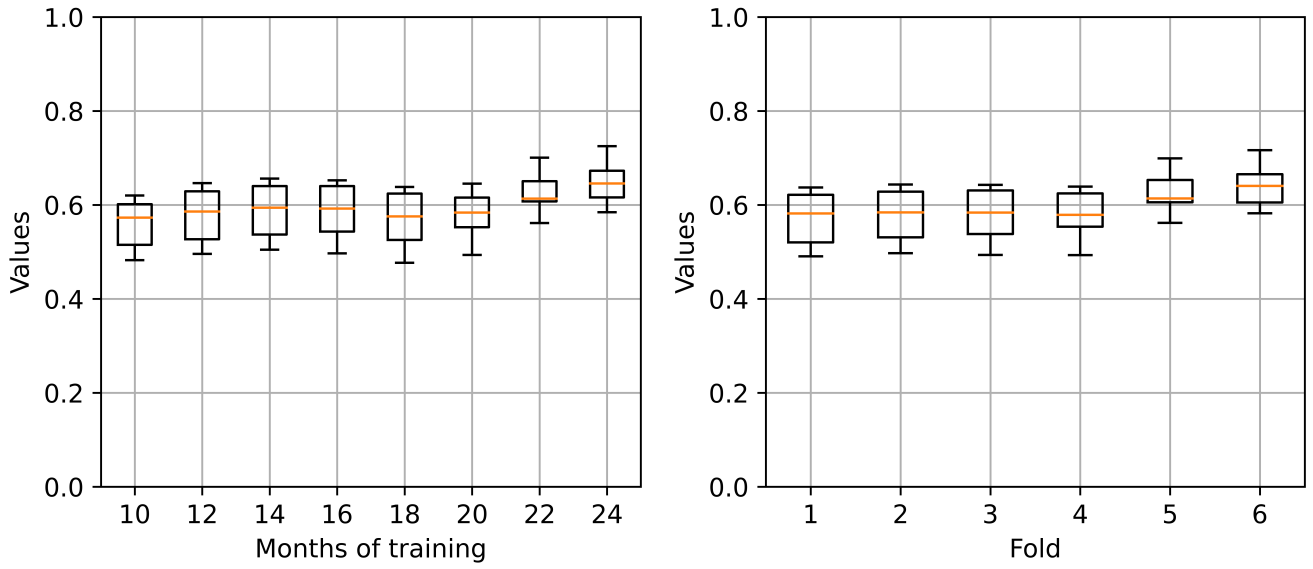
Figure 5.9: Training and testing periods for TSS and BTSS cross-validation

The results of these 2 methods are shown in Table 5.4 for turbine T01 which is the turbine in the dataset with the least number of warnings in the log data. The table shows the mean values from the different periods for each parameter. In addition, the standard deviation and the CoV are also included in the table. It can be observed that in both cases the CoV is low for all the parameters. This indicates that for a given wind turbine, the parameters do not vary much when changing training periods. This is a positive sign for the robustness of the model as it is not sensitive to different data of the same turbine. A similar behaviour is observed in all 12 turbines of the dataset. All the calculated parameters are available in Appendix A.2 in Figures A.2 and A.3.

Table 5.4: Parameters of Equation 5.5 calculated for different training periods using two methods: TSS and BTSS

Parameter	TSS			BTSS		
	Mean	σ	CoV	Mean	σ	CoV
β_1	$8.2{\cdot}10^{-01}$	$1.2{\cdot}10^{-03}$	$1.5{\cdot}10^{-03}$	$8.1{\cdot}10^{-01}$	$2.6{\cdot}10^{-03}$	$3.1{\cdot}10^{-03}$
β_3	$1.8{\cdot}10^{-01}$	$1.2{\cdot}10^{-03}$	$6.7{\cdot}10^{-03}$	$1.9{\cdot}10^{-01}$	$2.6{\cdot}10^{-03}$	$1.4{\cdot}10^{-02}$
β_5	$2.0{\cdot}10^{-05}$	$2.0{\cdot}10^{-06}$	$1.0{\cdot}10^{-01}$	$2.1{\cdot}10^{-05}$	$1.4{\cdot}10^{-06}$	$6.9{\cdot}10^{-02}$
β_6	$4.2{\cdot}10^{-01}$	$3.7{\cdot}10^{-03}$	$8.9{\cdot}10^{-03}$	$4.2{\cdot}10^{-01}$	$3.3{\cdot}10^{-03}$	$7.8{\cdot}10^{-03}$
β_7	$5.7{\cdot}10^{-04}$	$3.2{\cdot}10^{-05}$	$5.6{\cdot}10^{-02}$	$5.6{\cdot}10^{-04}$	$7.0{\cdot}10^{-05}$	$1.3{\cdot}10^{-01}$
β_8	$1.1{\cdot}10^{00}$	$1.2{\cdot}10^{-02}$	$1.1{\cdot}10^{-02}$	$1.1{\cdot}10^{00}$	$2.4{\cdot}10^{-02}$	$2.2{\cdot}10^{-02}$
β_9	$2.8{\cdot}10^{-04}$	$6.0{\cdot}10^{-06}$	$2.2{\cdot}10^{-02}$	$2.9{\cdot}10^{-04}$	$8.7{\cdot}10^{-06}$	$3.0{\cdot}10^{-02}$

Figure 5.10 shows how the RMSE changes with changing the training data using both TSS and BTSS. The RMSE is calculated for the testing period of each fold. The box plot include the RMSE for all 12 turbines. The variation appears low for the different training periods for all turbines. Also, Table 5.5 shows the mean values, standard deviations, and CoV of the performance metrics for different training periods. It can be observed that the CoV is low for all three metrics which means that the model is able to perform well independently of the changing datasets. The same is observed in the figures where only in the last two folds of both TSS and BTSS there is a slight increase in the values of RMSE which could also be due to the degradation of gearbox components. However, this is just an assumption as there is no information available about the health state of the turbines.



(a) Time-series split

(b) Blocking time-series split

Figure 5.10: RMSE for different training periods using TSS and BTSS

Table 5.5: Mean values, Standard deviations, and Coefficients of Variation of the performance metrics for different training periods using two methods: TSS and BTSS

Parameter	TSS			BTSS		
	Mean	σ	CoV	Mean	σ	CoV
RMSE	0.593	0.026	0.044	0.597	0.025	0.043
MAPE	0.784	0.027	0.035	0.789	0.024	0.031
MSE	0.354	0.032	0.089	0.359	0.030	0.085

The model used in this study is a physics-based normal behaviour model, so the parameters calculated from the regression analysis had to be checked in order to understand whether they accurately represent the physical characteristics of the gearbox. For that reason, the parameters β_5, β_6 and β_9 of Equation 5.5 are calculated for a reference gearbox with known characteristics as described in [71]. The nominal power of the gearbox used in this study is 2.5MW. The gearbox has 3 stages of which 2 are planetary and one is a spur gear stage, similarly to the 2MW MM82 wind turbines of the Penmanshiel wind farm considered in this study. The parameter β_3 regarding the cooling system is calculated based on [38]. The parameters calculated for this gearbox can then be compared to the parameters calculated from the regression analysis. The calculation of the parameters is described in Appendix A.1 and the results are shown in Table 5.6. The table also shows the MAPE of the parameters obtained from the NBM when compared to the calculated values of the rence gearbox.

Table 5.6: Parameters calculated for reference gearbox

Parameter	Value	MAPE(%)
β_1	$8.01 \cdot 10^{-1}$	1.8
β_3	$1.79 \cdot 10^{-1}$	5.6
β_5	$3.84 \cdot 10^{-5}$	70.7
β_6	$4.10 \cdot 10^{-2}$	19.6
β_8	[1,1.5]	8.2
β_9	$3.20 \cdot 10^{-4}$	17.1

These values are plotted in Figure 5.11 together with the parameters calculated by the fitting of Equation 4.37 to the data of each wind turbine. It can be observed that the calculated parameters are of the same order of magnitude as those of the reference turbine. Of course, it was expected to have some deviation as the reference turbine is different from the turbines in the dataset. Parameters such as β_5 and β_6, corresponding to no-load losses from gears, are those with the largest deviation from the reference turbine with MAPE of 70.8 and 19.6, respectively.

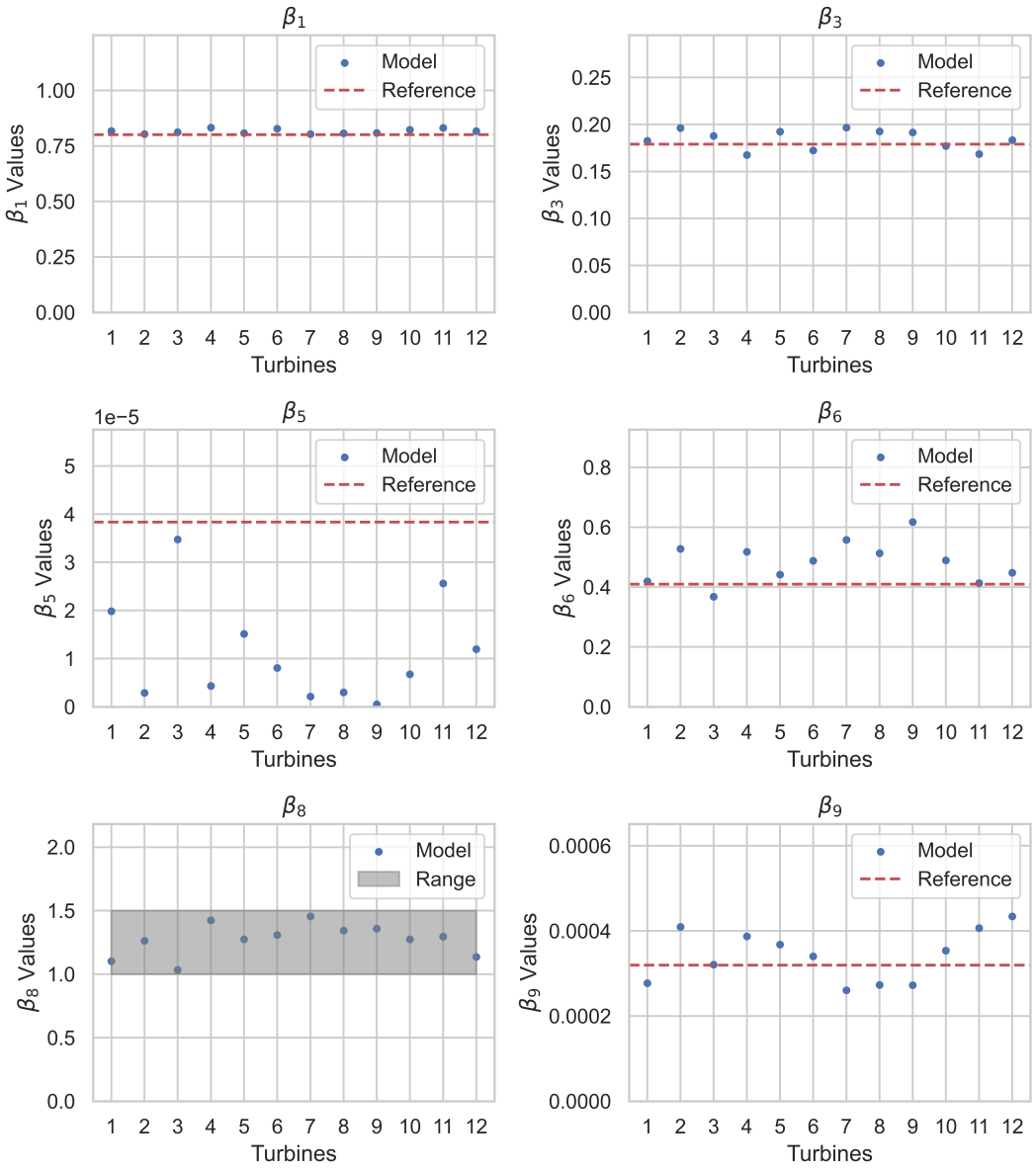


Figure 5.11: Reference parameter values and parameter values from fitting Equation 5.5 to the data of each wind turbine

In addition, the terms of the heat balance Equation 4.28 are compared between the reference turbine and the values obtained from the model for the 12 turbines of the dataset. Figure 5.12 shows the rate of change of the internal energy of the gearbox ($\dot{U}$) for different temperature changes, the heat dissipation through the cooling system ($\dot{Q}_{cool}$) for different temperature difference between the oil sump and the oil inlet temperatures and the no-load losses of the gears (P_{VZ0}) and the bearings (P_{VL0}) for different rotational speeds. It can be observed that even for P_{VZ0} , calculated using β_5 and β_6 which have the largest MAPE, the curve of the reference turbine does not significantly deviate from the curves of the 12 turbines.

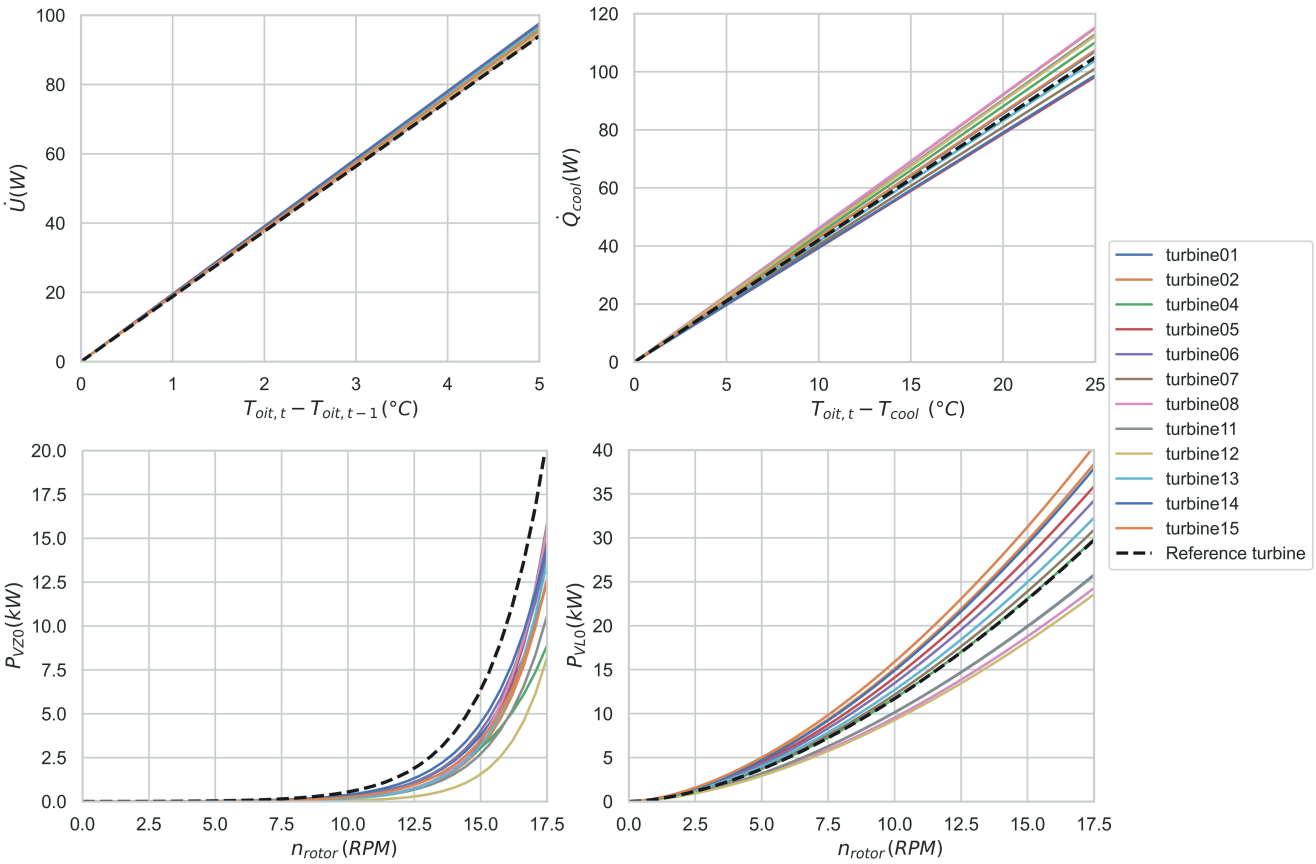


Figure 5.12: Reference parameter values and parameter values from fitting Equation 4.37 in the data of each wind turbine

5.1.3 Comparison with ANN models

To provide a benchmark, the performance of the physics-based model is also compared to that of two data-driven models using ANN for temperature prediction. These models are a NARX-ANN and a LR-ANN the architecture of which is provided in Section 4.4. As mentioned in Section 3.4, the autoregression element in NARX-ANN can either be the time history of the target variable or the calculated output of the model. Both were tested and the model using the time history had significantly better performance. So, only this case is presented here.

The metrics used to compare the performance are RMSE, MAPE, STDE and MSE which are presented in Table 5.7 along with time required for training the models. The metrics are calculated for the testing period which in this case is one year of data. The table shows that the model with the lowest values for all metrics is the physics-based NBM developed in this study and this indicates its superior performance. For example, the physics-based model exhibits a 14.1% lower RMSE compared to the NARX-ANN and a 58.9% lower RMSE than the LR-ANN. At the same time, the the physics-based model takes the shortest time to train with around 70% less time than the NARX-ANN and 95% less time than the LR-ANN. In Figure 5.13, the distribution of the residuals of the testing period is displayed. It can be seen that the distribution with the sharpest peak is that of the physics-based model. The sharpest peak indicates that a higher proportion of residuals are closer to zero, suggesting that the model's predictions align more closely with the actual values.

The broader distribution of the ANN models indicate a higher variability in prediction errors, implying a less accurate model. This again indicates a better performance of the model developed in this study compared to the two ANN models.

Table 5.7: Comparison of metrics between different NBMs

	RMSE	STDE	MAPE	MSE	Time(sec)
Physics-based	0.562	0.554	0.742	0.314	4.3
LR-ANN	1.368	1.364	1.545	2.180	92.5
NARX-ANN	0.654	0.651	0.876	0.431	14.5

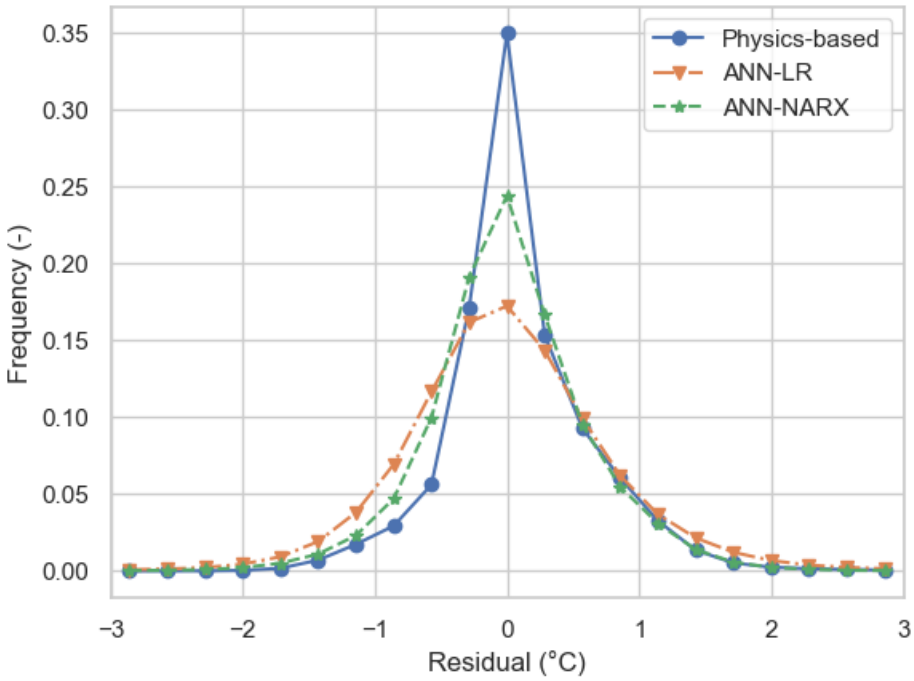


Figure 5.13: Residual distribution for the 12 wind turbines

It is important to mention that when using an ANN for the NBM, retraining the model from the beginning may yield different results. This variation arises due to the random initialisation of weights and biases. On the other hand, the physics-based model will always give the same results for a given dataset. Finally, the LR-ANN method poorly performed for three turbines however, when the ANN was retrained with adjusted initialisation of weights and biases, they achieved similar performance with the rest. This was not the case for NARX-ANN which managed to achieve good performance from the first training of the ANN for all turbines.

5.2 Case study 2

The second case study refers to the dataset made available by Energias de Portugal (EDP) a Portuguese electric utilities company. This is a publicly available dataset introduced during the Hack the Wind 2018: Wind Turbine Failure Detection competition [103].

5.2.1 Data-set description

The EDP dataset includes 2 years of SCADA data from 5 wind turbines (T01, T06, T07, T09 and T11). These are 2MW wind turbines with a wind class of IEC II. Similarly to the wind turbines in the Penmanshiel dataset, the gearbox has 3 stages as shown in the datasheet found in Appendix A.4. The dataset includes 81 SCADA signals, given for a 10-minute average period. For some signals in addition to the 10-minute average value, there is also the minimum, the maximum and the standard deviation available.

The main difference between this and the Penmanshiel dataset is that this one does not include the signal for the gearbox oil inlet temperature. This does not allow for Equation 4.37 to be used. However, within this dataset a log is also available for each turbine, including remarks about its operation, such as the operation of the gearbox cooling system and of the generator cooling fan, and the turbine manual stops.

Finally, the dataset includes 1 year of previously detected failures, as shown in Table A.2 in Appendix A.3. There are 4 gearbox failures. The first one is on turbine T01 where the gearbox pump was damaged. The second one is on turbine T09 and the available information mentions that gearbox was repaired without any further information on its problem. The third one is on turbine T06 and it is mentioned that it is about damaged bearings. The forth annotated failure refers to gearbox noise in the T09 turbine, without providing any further information.

5.2.2 Model modification

To compensate for the missing signal of this dataset, the operating state of the cooling system of the gearbox found in the log file is used in the model. The cooling systems have two operational states. When the temperature of the gearbox oil is low, the oil circulates through the cooling system and the filters, but the cooling system fan is off, whereas, when the oil temperature exceeds a certain value, the fan starts to operate [48], [104].

The log file does not store the signals for the state of the cooling system in 10-minute intervals like the SCADA data. To make it compatible with the model, the log file is merged with the SCADA dataset. This merging procedure aligns the irregular timestamps in the log file with the nearest available timestamp in the SCADA dataset as shown in Figure 5.14. As a result, a unified dataset was generated, maintaining a regular 10-minute timestep while incorporating relevant information from both the SCADA data and the log file.

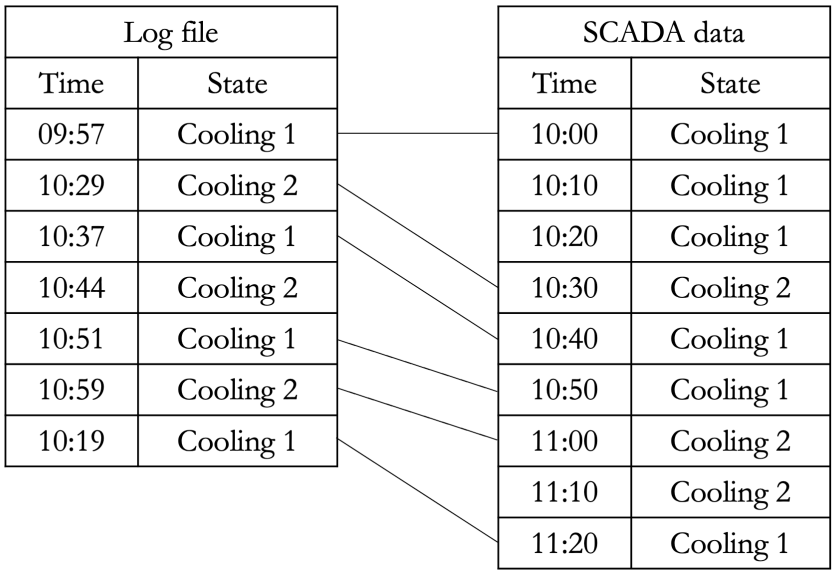


Figure 5.14: Merging of log file with SCADA data

It is assumed that the heat exchange of the cooling system happens between the oil and the ambient air and instead of using Equation 3.18 can be calculated according to

$$\begin{aligned} \dot{Q}_{cool,1} &= (1-C)\cdot k_1\cdot(T_{oil}-T_{amb}) \\ \dot{Q}_{cool,2} &= C\cdot k_2\cdot(T_{oil}-T_{amb}) \end{aligned} \tag{5.6}$$

where T_{amb} is the ambient temperature and C is a binary term which is equal to 0 and 1 when the fan is off and on respectively. $\dot{Q}_{cool,1}$ is the heat dissipated through the cooling system when the fan is on and $\dot{Q}_{cool,2}$ is the heat dissipated through the cooling system when the fan is off. So the total heat dissipation is calculated as

$$\dot{Q}_{diss} = \dot{Q}_{nac}+\dot{Q}_{cool,1}+\dot{Q}_{cool,2} = k_0\cdot A_{ca}(T_{oil}-T_{nac})+(1-C)\cdot k_1\cdot(T_{oil}-T_{amb})+C\cdot k_2\cdot(T_{oil}-T_{amb}) \tag{5.7}$$

The two parameters k_1 and k_2 are assumed different for the 2 cooling system states and are calculated by fitting the model to the data. These parameters depend on the characteristics of the cooling system.

A new equation for the calculation of the gearbox oil temperature needs to be formulated for this case study based on the heat balance Equation 4.28. By substituting 4.6, 4.12, 4.15, 4.20, 4.23 and

Equation 5.7 we get

$$mC\frac{dT_{\text{oil}}}{dt} = P_V - \dot{Q}_{\text{diss}} \Rightarrow \tag{5.8}$$
$$mC\frac{dT_{\text{oil}}}{dt} = P_{VZP} + P_{VZ0} + P_{VLP} + P_{VL0} + P_{VD} - \dot{Q}_{nac} - \dot{Q}_{cool,1} - \dot{Q}_{cool,2} \Rightarrow$$
$$mC\frac{dT_{\text{oil}}}{dt} = a_1\frac{P_{el}^{1.2}}{n_{rotor}^{0.4}}\nu_{\text{oil}}^{-0.05} + a_2{\cdot}n_{rotor}{\cdot}e^{a_3{\cdot}n_{rotor}} + b_1\left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor}$$
$$+ b_3{\cdot}n_{rotor}(\nu_{oil}{\cdot}n_{rotor})^{2/3} + c{\cdot}n_{rotor} - k{\cdot}(T_{\text{oil},t} - T_{nac,t}) - (1-C){\cdot}k_1{\cdot}(T_{\text{oil},t} - T_{\text{amb},t})$$
$$- C{\cdot}k_2{\cdot}(T_{\text{oil},t} - T_{\text{amb},t})$$

Similarly with subsection 4.2.3, a constant rate of change ($\Delta T/\Delta t$) over each 10-minute interval (Δt) is assumed so the equation is written as

$$mC\frac{T_{\text{oil},t} - T_{\text{oil},t-1}}{\Delta t} = a_1\frac{P_{el}^{1.2}}{n_{rotor}^{0.4}}\nu_{\text{oil}}^{-0.05} + a_2{\cdot}n_{rotor}{\cdot}e^{a_3{\cdot}n_{rotor}} + b_1\left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor} \tag{5.9}$$
$$+ b_3{\cdot}n_{rotor}(\nu_{oil}{\cdot}n_{rotor})^{2/3} + c{\cdot}n_{rotor} - k{\cdot}(T_{\text{oil},t} - T_{nac,t}) - (1-C){\cdot}k_1{\cdot}(T_{\text{oil},t} - T_{\text{amb},t})$$
$$- C{\cdot}k_2{\cdot}(T_{\text{oil},t} - T_{\text{amb},t})$$

By solving the equation for $T_{\text{oil},t}$ we get

$$T_{\text{oil},t} = (\frac{mC}{\Delta t}T_{\text{oil},t-1} + k{\cdot}T_{nac,t} + (1-C){\cdot}k_1{\cdot}T_{\text{amb},t} + C{\cdot}k_2{\cdot}T_{\text{amb},t} \tag{5.10}$$
$$+ a_1\frac{P_{el}^{1.2}}{n_{rotor}^{0.4}}\nu_{\text{oil}}^{-0.05} + a_2{\cdot}n_{rotor}{\cdot}e^{a_3{\cdot}n_{rotor}} + b_1\left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor} + b_3{\cdot}n_{rotor}(\nu_{oil}{\cdot}n_{rotor})^{2/3}$$
$$+ c{\cdot}n_{rotor})/(\frac{mC}{\Delta t} + k + k_1(1-C) + k_2{\cdot}C)$$

The heat capacity and the mass of the gearbox are considered constant while $\Delta t = 10$ min. So, they can be grouped together using

$$a = \frac{mC}{\Delta t} \tag{5.11}$$

similarly with subsection 4.2.3. The equation for gearbox oil temperature can be written as

$$T_{\text{oil},t} = (a{\cdot}T_{\text{oil},t-1} + k{\cdot}T_{nac,t} + (1-C){\cdot}k_1{\cdot}T_{\text{amb},t} + C{\cdot}k_2{\cdot}T_{\text{amb},t} \tag{5.12}$$
$$+ a_1\frac{P_{el}^{1.2}}{n_{rotor}^{0.4}}\nu_{\text{oil}}^{-0.05} + a_2{\cdot}n_{rotor}{\cdot}e^{a_3{\cdot}n_{rotor}} + b_1\left(\frac{P_{el}}{n_{rotor}}\right)^{b_2} n_{rotor} + b_3{\cdot}n_{rotor}(\nu_{oil}{\cdot}n_{rotor})^{2/3}$$
$$+ c{\cdot}n_{rotor})/\left(a + k + k_1(1-C) + k_2{\cdot}C\right)$$

The signals from the dataset used in this study are:

- Average temperature of oil in gearbox (T_{oil})
- Average temperature in nacelle (T_{nac})
- Average rotor rpm (n_{rotor})

- Average ambient temperature (T_{amb})
- Total active power (P_{el})

5.2.3 Normal behaviour model application and evaluation

The training period of the NBM is selected to be at least 6 months before any failure occurs as suggested by [53] while its length is selected to be 1 year in order to include the weather pattern of a whole year. The performance of the model is evaluated using a testing period after the training period which is still 6 months ahead of any failure.

The evaluation of the model performance starts with the calculation of some metrics as described in Section 4.3. Figure 5.15 shows the RMSE for all the turbines of the dataset when calculated using testing dataset. The mean RMSE of all turbines is 0.749 and 0.772 for the training and the testing period, respectively. There is only a 3% percentage difference between the two values which indicates no over-fitting. In addition, the MAPE is 1.1% for the testing period suggesting a good performance of the model.

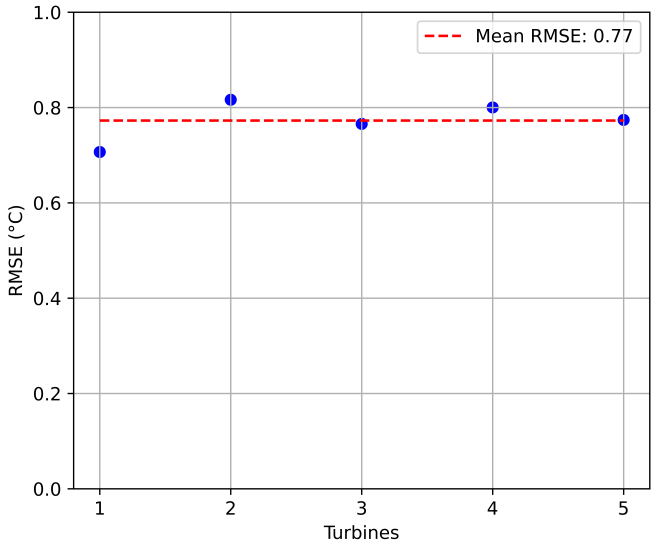


Figure 5.15: RMSE values of different turbines of the dataset

A closer inspection of the modelled temperature with respect to the actual temperature measurements shows that the model does not perform that well. In fact, the model has the tendency to estimate the temperature in close proximity to the value of the previous timestep which is used as input in the model. This is shown in Figure 5.16 where the modelled oil temperature is plotted against both the actual oil temperature and the actual oil temperature of the previous timestep. It can be observed that the correlation in plot a) is much lower than the correlation in plot b).

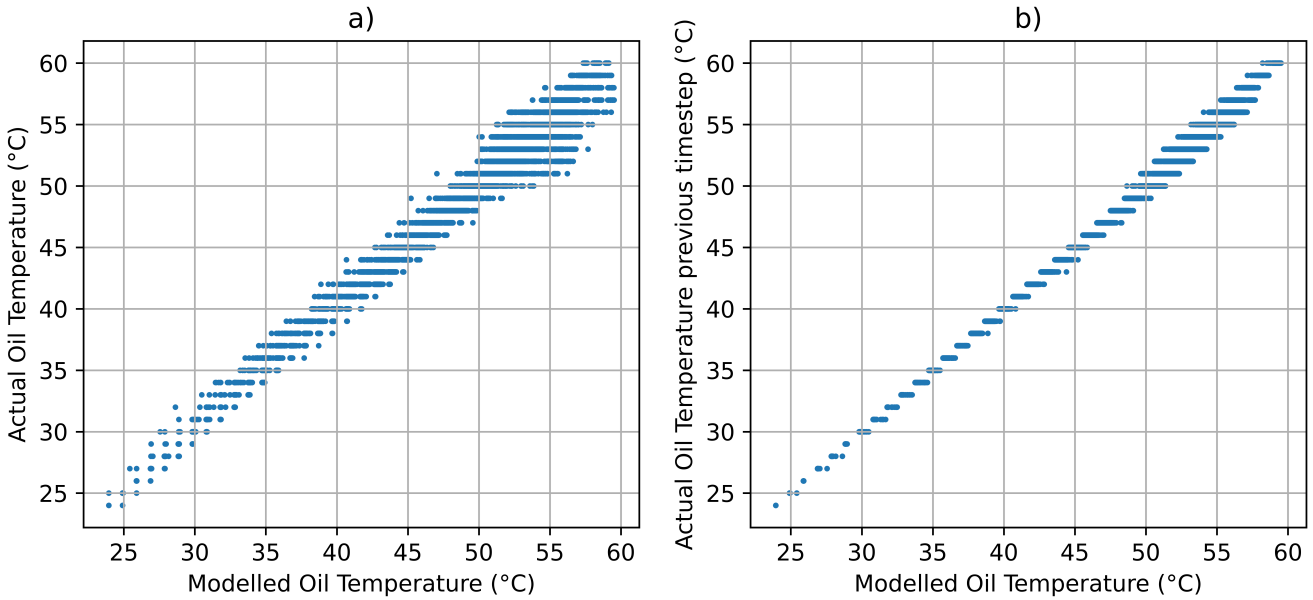


Figure 5.16: Comparison of modelled and actual temperature of gearbox oil for T07

This can be explained by the inability of the model to take into account the effect of the different states of the cooling system regardless of the two terms in the equation for the calculation of the temperature which are supposed to take the cooling system into account. This can also be observed in Figure 5.17.

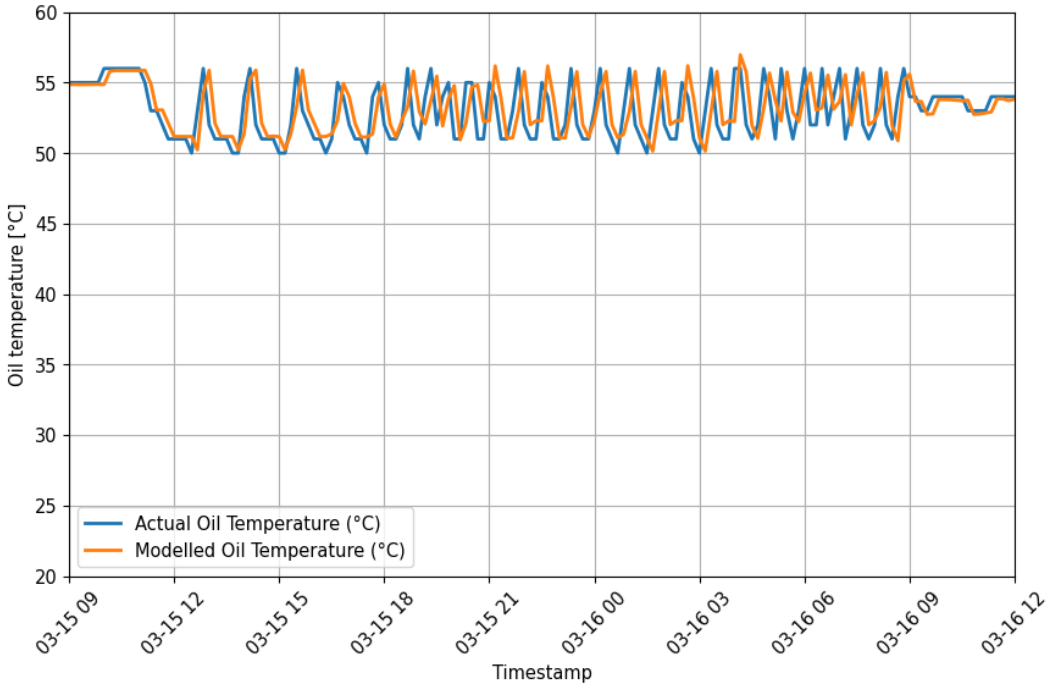


Figure 5.17: Comparison of actual and modelled temperatures over time for T07 from EDP dataset

The most likely reason for the inability of the equation to properly fit the data is the way the log file

is merged to the SCADA data. Each time the cooling system switches state, the nearest available value from the 10-minute timestamp of the SCADA data is selected and given the corresponding value. This modifies the time of the change between the states as shown in Figure 5.14 and it appears not to be a valid assumption that the time in the log will be close to the 10-minute signals. For that reason, it is concluded that this model is unable to perform well without the signal from the inlet temperature of the oil in the gearbox.

For comparison, the actual and modelled temperatures of turbine T01 from the Penmanshiel dataset are shown in Figure 5.18. When comparing the graphs for the two datasets, it can be observed that in the case of the EDP data the modelled temperature is similar to the actual temperature of the previous timestep. On the other hand, in the graph from Penmanshiel, where the inlet oil temperature is available, the modelled temperature is much closer to the actual for the same timestamp.

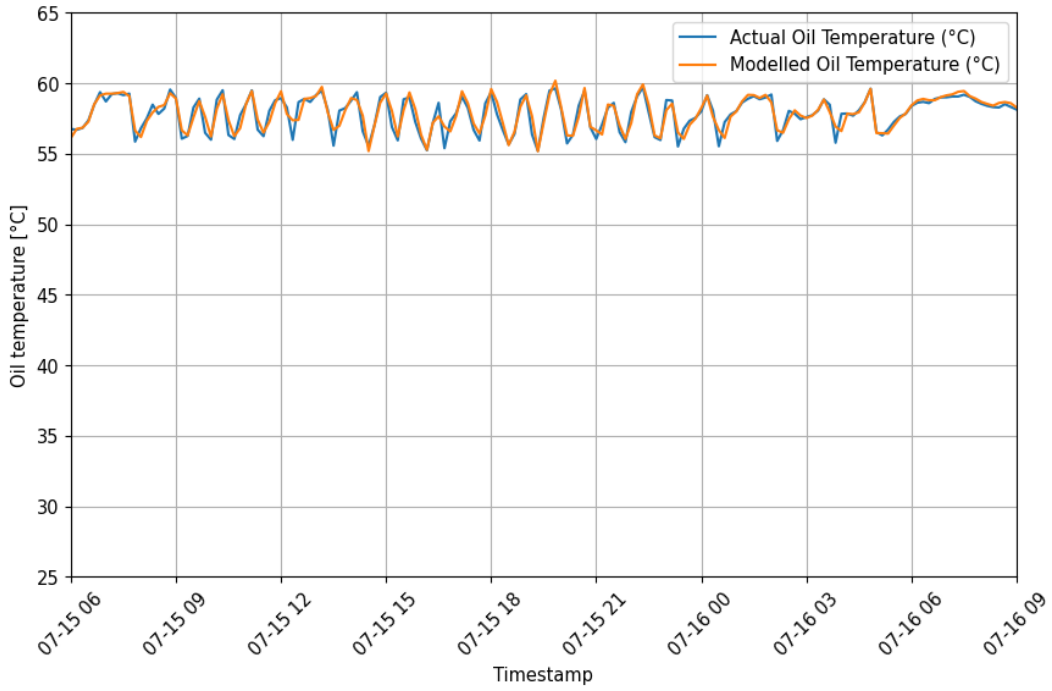


Figure 5.18: Comparison of actual and modelled temperatures over time for T01 from Penmanshiel dataset

This can also be explained by calculating the weights of the previous timestamp temperature for the 2 models. For the Penmanshiel dataset this weight is parameter β_1 from Equation 5.5 while for the EDP dataset is calculated as

$$w_{T,EDP} = \frac{a}{a + k + k_1 + k_2}. \tag{5.13}$$

In the case of EDP, the weight is 0.95 on average whereas in the case of Penmanshiel the weight is 0.8 on average. This means that for the case of the EDP dataset calculation of oil temperature is mainly based on the previous timestep temperature whereas for the case of the Penmanshiel dataset, the other signals used as inputs also contribute to the oil temperature calculation.

Chapter 6

Conclusions and Recommendations

This chapter includes the conclusions derived from this thesis by answering the research questions stated in the introduction. Then, recommendations for future work are also provided based on the outcome of the thesis.

6.1 Conclusions

The main objective of this thesis is to design a physics-based NBM of wind turbine gearbox oil temperature and use field data to validate its effectiveness. In order to achieve that, several steps are followed. First, a literature review is conducted which includes the different maintenance strategies, and the methods proposed in the literature for wind turbine condition monitoring using SCADA data. Then, the theory behind the main principles used in this thesis is provided in order to introduce the methodology describing how the proposed model works and the methods used to evaluate its effectiveness. Finally, the model is implemented in two case studies and its performance is assessed.

In Chapter 1, the research questions of this study are stated. These questions are answered in this chapter.

Why should a physics-based NBM be developed for gearbox oil temperature calculation?

Firstly, it is evident that the gearbox stands out as a crucial component in a wind turbine, contributing to prolonged downtime and incurring the highest replacement costs per failure. Recognising this, the development of effective methods for gearbox condition monitoring becomes imperative. NBMs using SCADA data have demonstrated good results for various components within the wind turbine gearbox and generator, particularly when integrated with expert systems. However, most of these studies rely on data-driven NBMs, which can yield results that are challenging to interpret. Furthermore, physics-based models have been implemented with trending, NBMs and damage modelling. However, the first requires human intervention and the third detailed information on the components of interest which are not usually available. Physics-based NBM using SCADA data have the advantage of providing interpretable results unlike data-driven NBMs.

This is also achieved without the need for detailed information on the gearbox by leveraging the availability of the SCADA data. Finally, some studies developing physics-based NBMs of different wind turbine components using SCADA data, suggest the development of a physics-based NBM of the gearbox. Consequently, the development of such a model emerges as a logical and promising endeavour.

How can a NBM of wind turbine gearbox oil temperature be developed using a physics-based approach?

This study proposes the model described in Chapter 4. The model combines the use of the first law of thermodynamics with historical SCADA data. First, the equation for the calculation of the gearbox oil temperature is formulated considering the balance between the heat generated in the gearbox from the losses in the bearings, the gears and the seals and the heat dissipated from the gearbox through the cooling system and through convection and conduction. Then, the equation is fitted to historical SCADA data using the least squares method to calculate its unknown parameters emerging from the unknown technical specifications of the gearbox. Finally, having the complete equation with known parameters it is possible to estimate the oil temperature on unseen data using the input signals from the SCADA system.

How can the accuracy of a physics-based NBM be evaluated?

The accuracy of the developed physics-based NBM has been assessed in several steps. First, the residuals, which are the differences between the actual observed temperatures and the temperatures predicted by the model, are analysed to assess the model's performance. Several metrics are calculated to quantify accuracy and they include RMSE, STDE, MAPE, and MSE. These metrics provide insight into how closely the model's predictions align with actual values. These metrics are computed separately for the training and testing periods to check for overfitting. Then, with the goal to develop a parsimonious model, the impact of each term in the heat balance equation is assessed to understand which components have the lowest influence on the temperature calculation, while AIC is used to evaluate the effect of reducing the number of parameters in the model. This helps determine whether simplifying the model improves or diminishes accuracy. Furthermore, cross-validation techniques are employed to assess how the model's performance varies with changes in training and testing periods. This provides insights into the model's generalisation capability. An investigation into the physical representation of the model's parameters is also conducted. Finally, to gain further insights, the physics-based model is compared with two data-driven models using the same set of metrics. This comparative analysis helps determine if the physics-based model is as accurate as the models proposed in the literature. By following these steps and considering a range of evaluation methods, the accuracy and reliability of the physics-based NBM can be comprehensively assessed.

What is the effect of each term of the heat balance equation in the calculation of temperature?

After fitting the equation to the SCADA data used for training, it became evident that certain parameters, specifically β_2, β_4, and β_{10} of Equation 4.37 which correspond to the heat dissipated to the nacelle, the load-dependent gear losses, and the seal losses, respectively, exhibited extremely low values. The reasons for these low values were investigated through both cross-correlation analysis and consideration of physical factors. The first parameter, β_2, represents the heat dissipated to the nacelle. Cross-correlation analysis revealed a very weak correlation between nacelle temperature and oil temperature. This weak correlation suggests that the nacelle's temperature has a limited influence on the temperature of the oil, which may explain the low value of this parameter. The second parameter, β_4, which accounts for load-dependent gear losses, exhibited an almost perfect correlation (close to 1) with another term in the heat balance equation, the load-dependent bearing losses. This high correlation indicates that these two terms are closely related, and as a result, only one of them significantly affects the output of the model. This redundancy could be the reason behind the negligible value of β_4. As for the third parameter, β_{10}, associated with seal losses, the low value might be attributed to the nature of heat generated by friction in the seals. It is plausible that this heat is not effectively absorbed by the oil but instead dissipates directly to the exterior of the gearbox through the attached shafts.

Do the calculated parameters of the model accurately represent the physical characteristics of the gearbox?

It is important to examine the parameters derived from the regression analysis to check for their consistency with the physical gearbox features. After comparing the values obtained from the model with the values calculated from a gearbox with known characteristics, it is observed that the values are in the same order of magnitude as the reference values, with the MAPE ranging from 1.8% to 70.8% highlighting the model's capability to capture and represent the gearbox's physical characteristics with good accuracy. Of course, it is expected to have some deviation as the reference turbine is different from the turbines in the dataset.

How does the performance of the physics-based model compare to existing data-driven NBM approaches?

The performance of the model is compared against 2 models from literature using ANNs for the calculation of temperature. These 2 models are the LR-ANN and the NARX-ANN. The physics-based model proposed in this study achieves better performance in all metrics with an RMSE 14.1% and 58.9% lower than NARX-ANN and LR-ANN, respectively. At the same time, the training time is significantly lower for the physics-based model as it needs 95% less time than the LR-ANN and around 70% less time than the NARX-ANN.

Can a physics-based NBM using SCADA data accurately predict the gearbox oil temperature?

In case study 1, the evaluation of the proposed physics-based NBM demonstrates its effectiveness for accurately predicting the gearbox oil temperature. This is proved by its low RMSE values, equal

to 0.554 and 0.562 for the training and testing periods, respectively. These values indicate minimal errors and no overfitting. Additionally, cross-validation reveals that variations in the training and testing periods have little impact on the model's performance. The mean RMSE of the folds is 0.593 with a CoV of 0.044 for TSS. Furthermore, the developed physics-based model outperforms two ANN models proposed in the literature in predicting the gearbox oil temperature. This underscores the model's ability to effectively utilise the heat balance principle and historical SCADA data to accurately calculate the unknown parameters of Equation 5.5 required for temperature prediction.

How is the performance of the physics-based model affected when the inlet oil temperature signal is not available in the dataset?

In case study 2 the dataset does not include the inlet oil temperature signal. For that reason, the model is modified and instead uses information from the log data also provided by the SCADA system. The resulting model shows an RMSE which 33% higher than that of the first case study. The weight of the previous timestep temperature which is present in the equation for the calculation of the temperature, is 0.95 which results in a model that gives a value almost the same as the previous timestep even when there are big changes in the temperature. This shows that the rest of the input signals are not sufficient to calculate the temperature of the oil and the inlet oil temperature is essential for a physics-based NBM of the oil temperature.

6.2 Recommendations

The purpose of this study is to develop a physics-based NBM that calculates the gearbox oil temperature. This is achieved as demonstrated in Chapter 5. However, for this model to contribute effectively to wind turbine condition monitoring, it is necessary to also prove its effectiveness for fault detection. Unfortunately, this aspect is not addressed in this study due to the absence of maintenance data for the wind turbines in the Penmashiel dataset. The following recommendations are proposed for future research:

- **Model validation against other datasets:**It is recommended to replicate the evaluation method employed for the physics-based model across multiple datasets containing the necessary input signals. This step is crucial as the present study relies on a single dataset and the wind turbines within this dataset are of the same type.
- **Fault Detection:** It is recommended to develop and test fault detection methods in conjunction with the physics-based model proposed in this study. Various fault detection techniques are available in the literature, which uses the residuals of NBMs. These methods can range from straightforward approaches involving predefined threshold values for the residuals [50], to more complex probabilistic approaches, such as control charts and the Welch's t-test[52], [84] or even health score systems [83]. The different methods should be compared to identify which one has a better performance.
- **Expert system development:** As explained in Chapter 2, the most effective condition monitoring methods combine the use of NBMs with expert systems which use fuzzy rules and membership functions to diagnose anomalies, predict remaining lifetime, and plan maintenance.

Such a system is necessary in order to use effectively the proposed NBM in real-world wind farms, so it needs to be developed to accompany the model proposed in this study.

- **Comparative analysis on fault detection:** In addition to comparing the performance of the physics-based model with data-driven methods in predicting oil temperature, it is suggested to conduct an additional comparison. This comparison should focus on assessing the accuracy of fault detection of the physics-based model and data-driven NBM like the ANN models presented in this study. This comparison is vital since it aligns with the primary objective of every condition monitoring system.
- **Investigate different physics-based models:** An analysis based on the first law of thermodynamics can be performed in other components of the wind turbine such as the generator and bearings. In addition, different physics-based models can be developed which utilise the available signals from the SCADA system and the physical behaviour of the gearbox such as physics-informed neural networks. However, this requires expertise in both the physics of the system and machine learning.

Bibliography

[1] Z. Liu, Z. Deng, S. Davis, and P. Ciais, "Monitoring global carbon emissions in 2022," *Nature Reviews Earth & Environment 2023 4:4*, vol. 4, no. 4, pp. 205–206, Mar. 2023, ISSN: 2662-138X. DOI: 10.1038/s43017-023-00406-z. [Online]. Available: https://www.nature.com/articles/s43017-023-00406-z.

[2] European Commission, "Investing in a climate-neutral future for the benefit of our people," Tech. Rep., 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0562.

[3] L. Cozzi, P. Frankl, B. Wanner, H. Bahar, and T. Spencer, *Tripling renewable power capacity by 2030 is vital to keep the 1.5°C goal within reach – Analysis - IEA*, Jul. 2023. [Online]. Available: https://www.iea.org/commentaries/tripling-renewable-power-capacity-by-2030-is-vital-to-keep-the-150c-goal-within-reach.

[4] IRENA, "FUTURE OF WIND Deployment, investment, technology, grid integration and socio-economic aspects IRENA," 2019. [Online]. Available: https://www.irena.org/-/media/files/irena/agency/publication/2019/oct/irena_future_of_wind_2019.pdf.

[5] REN21, "RENEWABLES 2022 GLOB AL STATUS REPORT," Tech. Rep., 2022. [Online]. Available: https://www.ren21.net/gsr-2022/.

[6] R. Wiser, M. Bolinger, B. Hoen, *et al.*, "Land-Based Wind Market Report: 2021 Edition Summary," Tech. Rep., 2021.

[7] IRENA, "RENEWABLE POWER GENERATION COSTS IN 2022," Abu Dhabi, Tech. Rep., Aug. 2023. [Online]. Available: https://www.irena.org/Publications/2023/Aug/Renewable-Power-Generation-Costs-in-2022.

[8] IRENA, "Renewable power generation costs in 2021," Abu Dhabi, Tech. Rep., 2021. [Online]. Available: www.irena.org.

[9] E. Artigao, S. Martín-Martínez, A. Honrubia-Escribano, and E. Gómez-Lázaro, "Wind turbine reliability: A comprehensive review towards effective condition monitoring development," *Applied Energy*, vol. 228, pp. 1569–1583, Oct. 2018, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2018.07.037. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261918310651.

[10] A. Stetco, F. Dinmohammadi, X. Zhao, *et al.*, "Machine learning methods for wind turbine condition monitoring: A review," *Renewable Energy*, vol. 133, pp. 620–635, Apr. 2019, ISSN: 18790682. DOI: 10.1016/J.RENENE.2018.10.047.

[11] C. Dao, B. Kazemtabrizi, and C. Crabtree, "Wind turbine reliability data review and impacts on levelised cost of energy," *Wind Energy*, vol. 22, no. 12, pp. 1848–1871, Dec. 2019, ISSN: 10991824. DOI: 10.1002/WE.2404/V1/REVIEW2.

[12] M. D. Reder, E. Gonzalez, and J. J. Melero, "Wind Turbine Failures - Tackling current Problems in Failure Data Analysis," *Journal of Physics: Conference Series*, vol. 753, no. 7, p. 072 027, Sep. 2016, ISSN: 1742-6596. DOI: 10.1088/1742-6596/753/7/072027. [Online]. Available: https://iopscience-iop-org.tudelft.idm.oclc.org/article/10.1088/1742-6596/753/7/072027%20https://iopscience-iop-org.tudelft.idm.oclc.org/article/10.1088/1742-6596/753/7/072027/meta.

[13] J. Carroll, A. Mcdonald, and D. Mcmillan, "Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines," 2015. DOI: 10.1002/we.1887. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/we.1887.

[14] C. Warren, *Enhancing Wind Turbine Reliability*, Aug. 2022. [Online]. Available: https://eprijournal.com/enhancing-wind-turbine-reliability/.

[15] H. Malik, N. Fatema, and A. Iqbal, *Intelligent Data-Analytics for Condition Monitoring: Smart Grid Applications*. Elsevier, Jan. 2021, pp. 1–251, ISBN: 9780323855105. DOI: 10.1016/B978-0-323-85510-5.00011-9. [Online]. Available: http://www.sciencedirect.com:5070/book/9780323855105/intelligent-data-analytics-for-condition-monitoring.

[16] E. Hau, "Wind turbines: Fundamentals, technologies, application, economics," *Wind Turbines: Fundamentals, Technologies, Application, Economics*, pp. 1–783, 2006. DOI: 10.1007/3-540-29284-5/COVER.

[17] S. Sheng, "Gearbox Typical Failure Modes, Detection, and Mitigation Methods (Presentation), NREL (National Renewable Energy Laboratory)," 2014.

[18] NREL, *Gearbox Reliability Database*, 2016. [Online]. Available: https://grd.nrel.gov/stats.

[19] E. Al-Ahmar, M. E. Benbouzid, Y. Amirat, and S. E. Elghali, "DFIG-based wind turbine fault diagnosis using a specific discrete wavelet transform," *Proceedings of the 2008 International Conference on Electrical Machines, ICEM'08*, 2008. DOI: 10.1109/ICELMACH.2008.4800033.

[20] H. Peng, H. Zhang, Y. Fan, L. Shangguan, and Y. Yang, "A Review of Research on Wind Turbine Bearings' Failure Analysis and Fault Diagnosis," *Lubricants 2023, Vol. 11, Page 14*, vol. 11, no. 1, p. 14, Dec. 2022, ISSN: 2075-4442. DOI: 10.3390/LUBRICANTS11010014. [Online]. Available: https://www.mdpi.com/2075-4442/11/1/14/htm%20https://www.mdpi.com/2075-4442/11/1/14.

[21] M. Mcdade and R. E. Geartech, "Wind Turbine Gearbox Failure Modes – A Brief (Presentation), NREL (National Renewable Energy Laboratory)," 2011.

[22] V. L. Jantara and M. Papaelias, "Wind turbine gearboxes: Failures, surface treatments and condition monitoring," *Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets*, pp. 69–90, Jan. 2020. DOI: 10.1016/B978-0-08-101094-5.00005-8.

[23] H. Farhat, "Maintenance: Availability and reliability," *Operation, Maintenance, and Repair of Land-Based Gas Turbines*, pp. 89–106, Jan. 2021. DOI: 10.1016/B978-0-12-821834-1.00009-5.

[24] A. Kelly, "Preventive maintenance decision-making Part 1: Principles, concepts and techniques," *Plant Maintenance Management Set*, pp. 85–118, Jan. 2006. DOI: 10.1016/B978-075066995-5.50037-0.

[25] N. H. Kim, J. H. Choi, and D. An, "Prognostics and health management of engineering systems: An introduction," *Prognostics and Health Management of Engineering Systems: An Introduction*, pp. 1–345, Jan. 2016. DOI: 10.1007/978-3-319-44742-1.

[26] H. Badihi, Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja, "A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 754–806, Jun. 2022, ISSN: 15582256. DOI: 10.1109/JPROC.2022.3171691.

[27] M. Martinez-Luengo, A. Kolios, and L. Wang, "Structural health monitoring of offshore wind turbines: A review through the Statistical Pattern Recognition Paradigm," *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 91–105, Oct. 2016, ISSN: 1364-0321. DOI: 10.1016/J.RSER.2016.05.085.

[28] J. Tautz-Weinert and S. J. Watson, "Using SCADA data for wind turbine condition monitoring - A review," *IET Renewable Power Generation*, vol. 11, no. 4, pp. 382–394, Mar. 2017, ISSN: 17521424. DOI: 10.1049/IET-RPG.2016.0248.

[29] C. Crabtree, D. Zappala, and Tavner P.J., "Survey of Commercially Available Condition Monitoring Systems for Wind Turbines," Tech. Rep., May 2014.

[30] M. Benbouzid, T. Berghout, N. Sarma, S. Djurović, Y. Wu, and X. Ma, "Intelligent Condition Monitoring of Wind Power Systems: State of the Art Review," *Energies 2021, Vol. 14, Page 5967*, vol. 14, no. 18, p. 5967, Sep. 2021, ISSN: 1996-1073. DOI: 10.3390/EN14185967. [Online]. Available: https://www.mdpi.com/1996-1073/14/18/5967/htm%20https://www.mdpi.com/1996-1073/14/18/5967.

[31] A. Verma, D. Zappalá, S. Sheng, and S. J. Watson, "Wind turbine gearbox fault prognosis using high-frequency SCADA data," *Journal of Physics: Conference Series*, vol. 2265, no. 3, p. 032 067, May 2022, ISSN: 1742-6596. DOI: 10.1088/1742-6596/2265/3/032067. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032067%20https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032067/meta.

[32] Z. János Viharos, C. István Sidló, A. A. Benczúr, *et al.*, ""Big Data" Initiative as an IT Solution for Improved Operation and Maintenance of Wind Turbines," in *Make your vision reality*, Vienna, Austria: European Wind Energy Association (EWEA), 2013, pp. 184–188.

[33] P. J. Tavner, "Review of condition monitoring of rotating electrical machines," *IET Electric Power Applications*, vol. 2, no. 4, pp. 215–247, 2008, ISSN: 17518660. DOI: 10.1049/IET-EPA:20070280.

[34] B. Roscher and R. Schelenz, "Usability of SCADA as predictive maintenance for wind turbines," *Forschung im Ingenieurwesen/Engineering Research*, vol. 85, no. 2, pp. 173–180, Jun. 2021, ISSN: 14340860. DOI: 10.1007/S10010-021-00454-1/TABLES/3. [Online]. Available: https://link.springer.com/article/10.1007/s10010-021-00454-1.

[35] R. Pandit, D. Astolfi, J. Hong, D. Infield, and M. Santos, "SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends," *Wind Engineering*, vol. 47, no. 2, pp. 422–441, Apr. 2023, ISSN: 2048402X. DOI: 10.1177/0309524X221124031/ASSET/IMAGES/LARGE/10.1177{_}0309524X221124031-FIG6.JPEG. [Online]. Available: https://journals-sagepub-com.tudelft.idm.oclc.org/doi/10.1177/0309524X221124031.

[36] I. M. Black, M. Richmond, and A. Kolios, "Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management," 2021. DOI: 10.1080/14786451.2021.1890736.

[37] 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, Jul. 2013, ISSN: 1099-1824. DOI: 10.1002/WE.1521. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.1521%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1521%20https://onlinelibrary.wiley.com/doi/10.1002/we.1521.

[38] Y. Qiu, Y. Feng, J. Sun, W. Zhang, and D. Infield, "Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data," *IET Renewable Power Generation*, vol. 10, no. 5, pp. 661–668, May 2016, ISSN: 1752-1424. DOI: 10.1049/IET-RPG.2015.0160. [Online]. Available: https://onlinelibrary-wiley-com.tudelft.idm.oclc.org/doi/full/10.1049/iet-rpg.2015.0160%20https://onlinelibrary-wiley-com.tudelft.idm.oclc.org/doi/abs/10.1049/iet-rpg.2015.0160%20https://ietresearch-onlinelibrary-wiley-com.tudelft.idm.oclc.org/doi/10.1049/iet-rpg.2015.0160.

[39] M. Wilkinson, B. Darnell, T. Van 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, ISSN: 1752-1424. DOI: 10.1049/IET-RPG.2013.0318. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2013.0318%20https://onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2013.0318%20https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-rpg.2013.0318.

[40] B. Corley, J. Carroll, and A. McDonald, "Thermal modelling of a small wind turbine gearbox for condition monitoring," *The Journal of Engineering*, vol. 2019, no. 18, pp. 5335–5339, Jul. 2019, ISSN: 2051-3305. DOI: 10.1049/JOE.2018.9282. [Online]. Available: https://www.researchgate.net/publication/334087910_Thermal_modelling_of_a_small_wind_turbine_gearbox_for_condition_monitoring.

[41] J. Durand De Gevigney, C. Changenet, F. Ville, and P. Velex, "Thermal modelling of a back-to-back gearbox test machine: Application to the FZG test rig," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 226, no. 6, pp. 501–515, Jun. 2012, ISSN: 13506501. DOI: 10.1177/1350650111433243.

[42] B. Corley, J. Carroll, and A. McDonald, "Fault detection of wind turbine gearbox using thermal network modelling and SCADA data," *Journal of Physics: Conference Series*, vol. 1618, no. 2, Sep. 2020, ISSN: 17426596. DOI: 10.1088/1742-6596/1618/2/022042. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1618/2/022042.

[43] B. Corley, S. Koukoura, J. Carroll, and A. McDonald, "Combination of thermal modelling and machine learning approaches for fault detection in wind turbine gearboxes," *Energies*, vol. 14, no. 5, Mar. 2021, ISSN: 19961073. DOI: 10.3390/EN14051375. [Online]. Available: https://www.researchgate.net/publication/349794383_Combination_of_Thermal_Modelling_and_Machine_Learning_Approaches_for_Fault_Detection_in_Wind_Turbine_Gearboxes.

[44] Z. Zhang and A. Kusiak, "Monitoring wind turbine vibration based on SCADA data," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 134, no. 2, May 2012, ISSN: 01996231. DOI: 10.1115/1.4005753/455682. [Online]. Available: https://asmedigitalcollection.asme.org/solarenergyengineering/article/134/2/021004/455682/Monitoring-Wind-Turbine-Vibration-Based-on-SCADA.

[45] X. Chesterman, T. Verstraeten, P. J. Daems, A. Nowé, and J. Helsen, "Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms," *Wind Energy Science*, vol. 8, no. 6, pp. 893–924, Jun. 2023, ISSN: 23667451. DOI: 10.5194/WES-8-893-2023.

[46] G. Helbing and M. Ritter, "Deep Learning for fault detection in wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 98, pp. 189–198, Dec. 2018, ISSN: 1364-0321. DOI: 10.1016/J.RSER.2018.09.012.

[47] J. Tautz-Weinert and S. J. Watson, "Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection," *Journal of Physics: Conference Series*, vol. 753, no. 7, p. 072 014, Sep. 2016, ISSN: 1742-6596. DOI: 10.1088/1742-6596/753/7/072014. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/753/7/072014%20https://iopscience.iop.org/article/10.1088/1742-6596/753/7/072014/meta.

[48] 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, no. 6, pp. 552–568, Aug. 2006, ISSN: 0166-3615. DOI: 10.1016/J.COMPIND.2006.02.011.

[49] W. Garlick, R. Dixon, and S. Watson, "A model-based approach to wind turbine condition monitoring using SCADA data," May 2009.

[50] M. Schlechtingen and I. Ferreira Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1849–1875, Jul. 2011, ISSN: 0888-3270. DOI: 10.1016/J.YMSSP.2010.12.007.

[51] C. McKinnon, A. Turnbull, S. Koukoura, J. Carroll, and A. McDonald, "Effect of Time History on Normal Behaviour Modelling Using SCADA Data to Predict Wind Turbine Failures," *Energies 2020, Vol. 13, Page 4745*, vol. 13, no. 18, p. 4745, Sep. 2020, ISSN: 1996-1073. DOI: 10.3390/EN13184745. [Online]. Available: https://www.mdpi.com/1996-1073/13/18/4745/htm%20https://www.mdpi.com/1996-1073/13/18/4745.

[52] 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, ISSN: 13552511. DOI: 10.1108/JQME-06-2016-0028/FULL/PDF.

[53] A. Turnbull, J. Carroll, and A. McDonald, "Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure," *Wind Energy*, vol. 24, no. 3, pp. 197–211, Mar. 2021, ISSN: 1099-1824. DOI: 10.1002/WE.2567. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.2567%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2567%20https://onlinelibrary.wiley.com/doi/10.1002/we.2567.

[54] X. Chesterman, T. Verstraeten, P. J. Daems, F. Perez Sanjines, A. Nowé, and J. Helsen, "The detection of generator bearing failures on wind turbines using machine learning based anomaly detection," *Journal of Physics: Conference Series*, vol. 2265, no. 3, p. 032 066, May 2022, ISSN: 1742-6596. DOI: 10.1088/1742-6596/2265/3/032066. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032066%20https://iopscience.iop.org/article/10.1088/1742-6596/2265/3/032066/meta.

[55] F. Jamil, T. Verstraeten, A. Nowé, C. Peeters, and J. Helsen, "A deep boosted transfer learning method for wind turbine gearbox fault detection," *Renewable Energy*, vol. 197, pp. 331–341, Sep. 2022, ISSN: 0960-1481. DOI: 10.1016/J.RENENE.2022.07.117.

[56] A. Meyer, "Early fault detection with multi-target neural networks," Jun. 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2106.08957.

[57] 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, Jul. 2010, ISSN: 1099-1824. DOI: 10.1002/WE.360. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.360%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.360%20https://onlinelibrary.wiley.com/doi/10.1002/we.360.

[58] M. Sepulveda, J. Shek, P. R. Thies, E. Oterkus, P. Davies, and M. Spring, "Physics-Based Gearbox Failure Model for Multi-MW Offshore Wind Turbines," Jun. 2017. DOI: 10.1115/OMAE2017-62257. [Online]. Available: https://doi.org/10.1115/OMAE2017-62257.

[59] V. L. Jantara Junior, H. Basoalto, and M. Papaelias, "A damage mechanics approach for lifetime estimation of wind turbine gearbox materials," *International Journal of Fatigue*, vol. 137, p. 105 671, Aug. 2020, ISSN: 0142-1123. DOI: 10.1016/J.IJFATIGUE.2020.105671.

[60] Y. Qiu, Y. Feng, P. Tavner, P. Richardson, G. Erdos, and B. Chen, "Wind turbine SCADA alarm analysis for improving reliability," *Wind Energy*, vol. 15, no. 8, pp. 951–966, Nov. 2012, ISSN: 1099-1824. DOI: 10.1002/WE.513. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.513%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.513%20https://onlinelibrary.wiley.com/doi/10.1002/we.513.

[61] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," *Renewable Energy*, vol. 36, no. 1, pp. 16–23, Jan. 2011, ISSN: 0960-1481. DOI: 10.1016/J.RENENE.2010.05.014.

[62] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, Jan. 2013, ISSN: 1568-4946. DOI: 10.1016/J.ASOC.2012.08.033.

[63] B. Chen, Y. N. Qiu, Y. Feng, P. J. Tavner, and W. W. Song, "Wind turbine SCADA alarm pattern recognition," *IET Conference Publications*, vol. 2011, no. 579 CP, p. 163, 2011. DOI: 10.1049/CP.2011.0164. [Online]. Available: https://www.researchgate.net/publication/260742118_Wind_turbine_SCADA_alarm_pattern_recognition.

[64] A. Pliego Marugán and F. P. García Márquez, "Advanced analytics for detection and diagnosis of false alarms and faults: A real case study," *Wind Energy*, vol. 22, no. 11, pp. 1622–1635, Nov. 2019, ISSN: 1099-1824. DOI: 10.1002/WE.2393. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.2393%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2393%20https://onlinelibrary.wiley.com/doi/10.1002/we.2393.

[65] Y. Qiu, Y. Feng, and D. Infield, "Fault diagnosis of wind turbine with SCADA alarms based multidimensional information processing method," *Renewable Energy*, vol. 145, pp. 1923–1931, Jan. 2020, ISSN: 0960-1481. DOI: 10.1016/J.RENENE.2019.07.110.

[66] Y. A. Çengel, *Heat transfer : a practical approach*. New York: McGraw-Hill, 2004, ISBN: 9780071236447. [Online]. Available: https://www.worldcat.org/title/224329445.

[67] K. Michaelis, B. R. Höhn, and M. Hinterstoißer, "Influence factors on gearbox power loss," *Industrial Lubrication and Tribology*, vol. 63, no. 1, pp. 46–55, 2011, ISSN: 00368792. DOI: 10.1108/00368791111101830.

[68] B.-R. Höhn, K. Michaelis, and M. Hinterstoißer, "OPTIMIZATION OF GEARBOX EFFICIENCY," *Goriva i maziva : časopis za tribologiju, tehniku podmazivanja i primjenu tekućih i plinovitih goriva i inžinjerstvo izgaranja*, vol. 48, no. 4, pp. 441–461, Dec. 2009, ISSN: 0350-350X. [Online]. Available: https://hrcak.srce.hr/47911.

[69] C. M. Fernandes, P. M. Marques, R. C. Martins, and J. H. Seabra, "Gearbox power loss. Part I: Losses in rolling bearings," *Tribology International*, vol. 88, pp. 298–308, Aug. 2015, ISSN: 0301-679X. DOI: 10.1016/J.TRIBOINT.2014.11.017.

[70] C. M. Fernandes, P. M. Marques, R. C. Martins, and J. H. Seabra, "Gearbox power loss. Part II: Friction losses in gears," *Tribology International*, vol. 88, pp. 309–316, Apr. 2015, ISSN: 0301679X. DOI: 10.1016/J.TRIBOINT.2014.12.004.

[71] C. M. Fernandes, M. Hammami, R. C. Martins, and J. H. Seabra, "Power loss prediction: Application to a 2.5 MW wind turbine gearbox," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 230, no. 8, pp. 983–995, Aug. 2016, ISSN: 2041305X. DOI: 10.1177/1350650115622362/ASSET/IMAGES/LARGE/10.1177{_}1350650115622362-FIG5.JPEG. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1350650115622362.

[72] P. M. Marques, C. M. Fernandes, R. C. Martins, and J. H. Seabra, "Power losses at low speed in a gearbox lubricated with wind turbine gear oils with special focus on churning losses," *Tribology International*, vol. 62, pp. 186–197, Jun. 2013, ISSN: 0301-679X. DOI: 10.1016/J.TRIBOINT.2013.02.026.

[73] ISO, *Gears — Thermal capacity — Part 2: Thermal load-carrying capacity (ISO/TR 14179-2:2001)*, 2001. [Online]. Available: https://www.iso.org/standard/30109.html.

[74] B. Srinivasan Venkatesan, "Modeling, Simulation and Correlation of Drag losses in a Power Transfer Unit of an All-Wheel Drive System," KTH Royal Institute of Technology, Stockholm, Tech. Rep., 2020.

[75] C. Changenet, G. Leprince, F. Ville, and P. Velex, "A note on flow regimes and churning loss modeling," *Journal of Mechanical Design, Transactions of the ASME*, vol. 133, no. 12, Dec. 2011, ISSN: 10500472. DOI: 10.1115/1.4005330/455842. [Online]. Available: https://asmedigitalcollection.asme.org/mechanicaldesign/article/133/12/121009/455842/A-Note-on-Flow-Regimes-and-Churning-Loss-Modeling.

[76] S. Abdan, N. Stosic, A. Kovacevic, I. Smith, and N. Asati, "Analysis of rolling bearing power loss models for twin screw oil injected compressor," *IOP Conference Series: Materials Science and Engineering*, vol. 604, no. 1, p. 012 013, Aug. 2019, ISSN: 1757-899X. DOI: 10.1088/1757-899X/604/1/012013. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/604/1/012013%20https://iopscience.iop.org/article/10.1088/1757-899X/604/1/012013/meta.

[77] C. M. Fernandes, P. M. Marques, R. C. Martins, and J. H. Seabra, "Gearbox power loss. Part III: Application to a parallel axis and a planetary gearbox," *Tribology International*, vol. 88, pp. 317–326, Aug. 2015, ISSN: 0301-679X. DOI: 10.1016/J.TRIBOINT.2015.03.029.

[78] *Viscosity - Absolute (Dynamic) vs. Kinematic*. [Online]. Available: https://www.engineeringtoolbox.com/dynamic-absolute-kinematic-viscosity-d_412.html.

[79] M. A. Rao, *Rheology of Fluid and Semisolid Foods* (Food Engineering Series). Boston, MA: Springer US, 2007, ISBN: 978-0-387-70929-1. DOI: 10.1007/978-0-387-70930-7. [Online]. Available: https://link.springer.com/10.1007/978-0-387-70930-7.

[80] M. Liu, X. Zhou, L. Yang, and X. Ju, "A novel Kalman-filter-based battery internal temperature estimation method based on an enhanced electro-thermal coupling model," *Journal of Energy Storage*, vol. 71, p. 108 241, Nov. 2023, ISSN: 2352-152X. DOI: 10.1016/J.EST.2023.108241.

[81] T. M. Le, B. Fatahi, H. Khabbaz, and W. Sun, "Numerical optimization applying trust-region reflective least squares algorithm with constraints to optimize the non-linear creep parameters of soft soil," *Applied Mathematical Modelling*, vol. 41, pp. 236–256, Jan. 2017, ISSN: 0307-904X. DOI: 10.1016/J.APM.2016.08.034.

[82] M. A. Branch, T. F. Coleman, and Y. Li, "Subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems," *SIAM Journal of Scientific Computing*, vol. 21, no. 1, pp. 1–23, 1999, ISSN: 10648275. DOI: 10.1137/S1064827595289108. [Online]. Available: http://www.siam.org/journals/sisc/21-1/28910.html.

[83] L. A. Lima, A. Blatt, and J. Fujise, "Wind Turbine Failure Prediction Using SCADA Data," *Journal of Physics: Conference Series*, vol. 1618, no. 2, p. 022 017, Sep. 2020, ISSN: 1742-6596. DOI: 10.1088/1742-6596/1618/2/022017. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1618/2/022017%20https://iopscience.iop.org/article/10.1088/1742-6596/1618/2/022017/meta.

[84] Y. Wang and D. Infield, "Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring," *IET Renewable Power Generation*, vol. 7, no. 4, pp. 350–358, Jul. 2013, ISSN: 1752-1424. DOI: 10.1049/IET-RPG.2012.0215. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2012.0215%20https://onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2012.0215%20https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-rpg.2012.0215.

[85] A. Meyer, "Multi-target normal behaviour models for wind farm condition monitoring," *Applied Energy*, vol. 300, p. 117 342, Oct. 2021, ISSN: 0306-2619. DOI: 10.1016/J.APENERGY.2021.117342.

[86] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, Dec. 2012, ISSN: 0960-1481. DOI: 10.1016/J.RENENE.2012.04.020.

[87] J. Delgadillo and D. Atzil-Slonim, "Artificial intelligence, machine learning and mental health," *Encyclopedia of Mental Health*, pp. 132–142, Jan. 2023. DOI: 10.1016/B978-0-323-91497-0.00177-6.

[88] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: an empirical study on performance estimation methods," *Machine Learning*, vol. 109, no. 11, pp. 1997–2028, Nov. 2020, ISSN: 15730565. DOI: 10.1007/S10994-020-05910-7/FIGURES/33. [Online]. Available: https://link-springer-com.tudelft.idm.oclc.org/article/10.1007/s10994-020-05910-7.

[89] Packt, *Cross-Validation strategies for Time Series forecasting*, May 2019. [Online]. Available: https://hub.packtpub.com/cross-validation-strategies-for-time-series-forecasting-tutorial/.

[90] L. Tarassenko, *A guide to neural computing applications*. Elsevier Science, 1998, ISBN: 03407-05892. [Online]. Available: https://books.google.mk/books?id=28CWNxEx60kC&printsec=frontcover#v=onepage&q&f=false.

[91] H. Kukreja, B. N, S. C S, and K. S, "An introduction to artificial neural network," *IJARIIE*, vol. 1, no. 5, pp. 27–30, 2016. [Online]. Available: https://ijariie.com/AdminUploadPdf/AN_INTRODUCTION_TO_ARTIFICIAL_NEURAL_NETWORK_1399.pdf.

[92] Wallstreetmojo, *Artificial Neural Network - Meaning, Types, Examples, Applications*. [Online]. Available: https://www.wallstreetmojo.com/artificial-neural-network/.

[93] 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, Aug. 2017, ISSN: 1099-1824. DOI: 10.1002/WE.2102. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/we.2102%20https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2102%20https://onlinelibrary.wiley.com/doi/10.1002/we.2102.

[94] Y. Cui, P. Bangalore, and L. B. Tjernberg, "An anomaly detection approach based on machine learning and scada data for condition monitoring of wind turbines," *2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings*, Aug. 2018. DOI: 10.1109/PMAPS.2018.8440525.

[95] W. Xie, A. Noble, and A. Zisserman, "Layer Recurrent Neural Networks," in *ICLR*, 2017. [Online]. Available: https://www.semanticscholar.org/paper/Layer-Recurrent-Neural-Networks-Xie-Noble/7a6896f17e815c8d4cd3d1300b1150499f8c6e91.

[96] *Layer recurrent neural network - MATLAB layrecnet - MathWorks Benelux*. [Online]. Available: https://nl.mathworks.com/help/deeplearning/ref/layrecnet.html.

[97] *Design Time Series NARX Feedback Neural Networks - MATLAB & Simulink - MathWorks Benelux*. [Online]. Available: https://nl.mathworks.com/help/deeplearning/ug/design-time-series-narx-feedback-neural-networks.html.

[98] C. Nutakor, A. Kłodowski, J. Sopanen, A. Mikkola, and J. I. Pedrero, "Planetary gear sets power loss modeling: Application to wind turbines," *Tribology International*, vol. 105, pp. 42–54, Jan. 2017, ISSN: 0301-679X. DOI: 10.1016/J.TRIBOINT.2016.09.029.

[99] C. Plumley, "Penmanshiel Wind Farm Data," Feb. 2022. DOI: 10.5281/ZENODO.5946808. [Online]. Available: https://doi.org/10.5281/zenodo.5946808#.ZGYMBTAvI9M.mendeley.

[100] *Senvion MM82/2050 - Manufacturers and turbines - Online access - The Wind Power*. [Online]. Available: https://www.thewindpower.net/turbine_en_464_senvion_mm82-2050.php.

[101] *Penmanshiel Wind Farm — THE PROJECT — Penmanshiel Wind Farm*. [Online]. Available: http://penmanshiel-windfarm.co.uk/the-project/penmanshiel-wind-farm/.

[102] A. Zaher, S. D. McArthur, D. G. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009, ISSN: 10991824. DOI: 10.1002/WE.319.

[103] EDP, *Data — edp.com*, Jul. 2018. [Online]. Available: https://www.edp.com/en/innovation/open-data/data.

[104] C. Sequeira, A. Pacheco, P. Galego, and E. Gorbeña, "Analysis of the efficiency of wind turbine gearboxes using the temperature variable," *Renewable Energy*, vol. 135, pp. 465–472, May 2019, ISSN: 0960-1481. DOI: 10.1016/J.RENENE.2018.12.040.

[105] Moventas, *PLH-1900.1 GEARBOX*. [Online]. Available: www.moventas.com/services.

[106] The Engineering ToolBox, *Metals - Specific Heats*, 2003. [Online]. Available: https://www.engineeringtoolbox.com/specific-heat-metals-d_152.html.

Appendix

A.1 Parameter calculation from reference turbine

The calculation of the parameters of the Equation 4.31 is demonstrated below for the case of the gearbox to estimate some reference values. Starting from parameter a, it is calculated as shown in Equation A.1. The assumptions for these parameters are the mass of the gearbox M [105], the uniform material and specific heat capacity c_p for all the gearbox [106] and that the timestep is constant and 10 minutes.

$$a = \frac{C_p}{dt} = \frac{c_p \cdot M}{dt} = \frac{0.49kJ/(kgK) \cdot 23000kg}{10 \cdot 60sec} = 18.8kW/K \tag{A.1}$$

For parameter k_1, the calculation is based on [38] where some information about the cooling system and the heat dissipation is provided. For this parameter, the density of the oil ρ_{oil}, the heat capacity of the cooling medium c_{oil} and the flow rate $\dot{V}$ are needed. In the study from which k_1 is calculated, the wind turbine has a nominal power output of 1.5MW. If we assume that the efficiency of the gearbox is the same for the 2MW wind turbines in Penmashiel dataset, this means that the cooling system needs to dissipate heat that is increased linearly with nominal power. In addition, this study states that the cooling system operates constantly whereas in the case of the turbines in Penmashiel, the cooling system turns on and off, so the same amount of heat needs to be dissipated in less time. By assuming a duty cycle DC of 0.6 for the cooling system, k_1 can be calculated as shown in Equation A.2.

$$k_1 = \rho_{oil} \cdot \dot{V} c_{oil} \cdot \frac{P_2}{P_1} \cdot \frac{1}{DC} = 900kg/m^3 \cdot 1.05 \cdot 10^{-3} m^3/sec \cdot 2kJ/(kgK) \cdot \frac{2}{1.5} \cdot \frac{1}{0.6} = 4.2kW/K \tag{A.2}$$

Continuing with the gearbox losses, the parameters are calculated based on the gearbox in [71]. First,

the parameters for the no-load losses of the gears P_{VZ0} are calculated according to Equation 3.10.

$$\begin{aligned} C_{Sp} &= \left(\frac{4h_e}{3h_c}\right)^{1,5} \frac{2h_c}{l_c} \\ C_1 &= 0,063\left(\frac{h_e}{10}\right) + 0,0128\left(\frac{b}{10}\right)^3 \\ C_2 &= \frac{h_e}{800} + 0,2 \end{aligned} \tag{A.3}$$

h_e is the immersion depth of the gear, h_c the radius of the larger gear, l_c the hydraulic length and b the width of the gear. From Equation 4.31, the parameters a_2 and a_3 are calculated as shown in Equation A.4. This calculation is only done for the third stage of the gearbox as it can not be applied for the planetary stages. In addition, [71] states that the churning losses from low speed gears such as the planetary stages can be disregarded.

$$\begin{aligned} a_2 &= C_{Sp} \cdot C_1 \cdot i_3 \cdot \frac{\pi}{30} \\ a_3 &= C_2 \cdot \frac{v_t}{n_{rated}} \end{aligned} \tag{A.4}$$

The calculated values are $a_2 = 0.0009$ and $a_3 = 0.41$

The no-load bearing losses are calculated based on Equation 3.11. However, as explained in subsection 4.2.1, only the bearings of the last stage are taken into account as the losses in the other stages are insignificant. The equation is modified for the intermediate and the high speed shaft according to Equation A.6 and Equation A.5 respectively.

$$\begin{aligned} &P_{VL0,hss} = 10^{-10} f_0 \left(\nu_{oil} n_{hss}\right)^{2/3} d_m^3 \frac{\pi n_{hss}}{30} = 10^{-10} f_0 (\nu_{oil} \cdot i_1 i_2 \cdot n_{rotor})^{2/3} d_m^3 \frac{\pi \cdot i_1 i_2 \cdot n_{rotor}}{30} \Rightarrow \\ &P_{VL0,hss} = b_{3,hss} f_0 \left(\nu_{oil} n_{rotor}\right)^{2/3} \cdot n_{rotor} \\ &b_{3,hss} = 10^{-10} \left(i_1 i_2\right)^{2/3} \cdot d_m^3 \frac{\pi \cdot i_1 i_2}{30} \end{aligned} \tag{A.5}$$

$$\begin{aligned} &P_{VL0,is} = 10^{-10} f_0 \left(\nu_{oil} n_{is}\right)^{2/3} d_m^3 \frac{\pi n_{is}}{30} = 10^{-10} f_0 \left(\nu_{oil} \cdot i_1 i_2 i_3 \cdot n_{rotor}\right)^{2/3} d_m^3 \frac{\pi \cdot i_1 i_2 i_3 \cdot n_{rotor}}{30} \Rightarrow \\ &P_{VL0,is} = b_{3,is} f_0 \left(\nu_{oil} n_{rotor}\right)^{2/3} \cdot n_{rotor} \\ &b_{3,is} = 10^{-10} \left(i_1 i_2 i_3\right)^{2/3} \cdot d_m^3 \frac{\pi \cdot i_1 i_2 i_3}{30} \end{aligned} \tag{A.6}$$

In Table A.1 the parameters of each bearing are calculated but also the summed value of b_3 from Equation 4.37.

Table A.1: Parameters for bearing no-load losses

	HSS			IS		
	Bearing 1	Bearing 2	Bearing 3	Bearing 1	Bearing 2	Bearing 3
b_3	$2.8 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$
f_0	5	5	6	5	6	6
sum	$7.5 \cdot 10^{-3}$					

For the load depended bearing losses P_{VLP} there are not sufficient information to be calculated however the parameter b_2 is known to range between 1 and 1.5 so $b_2 \in [1, 1.5]$ [73].

A.2 Parameters from the model

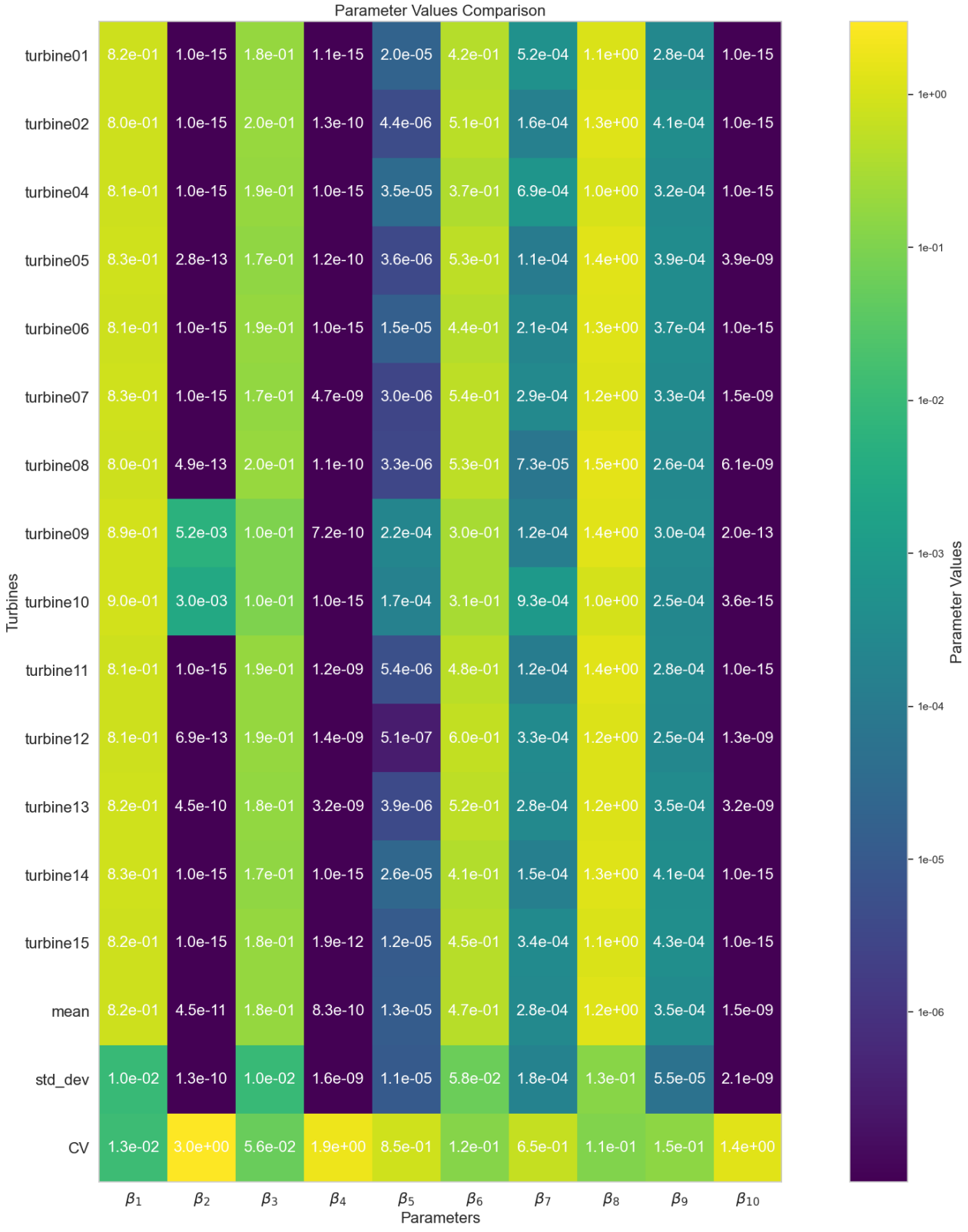


Figure A.1: Calculated parameters for all turbines when Equation 4.37 is fitted in one year of data

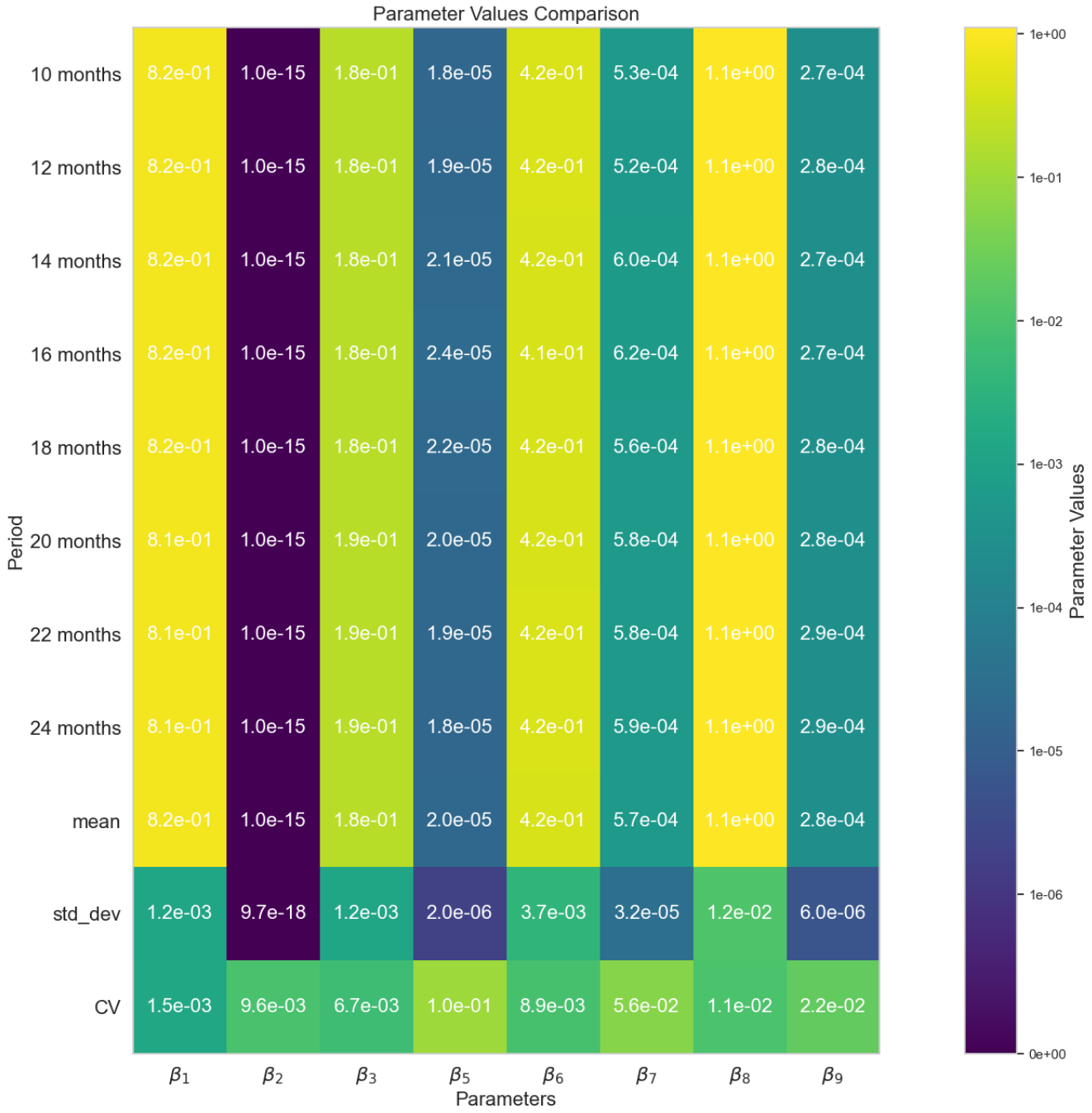


Figure A.2: Calculated parameters for turbine T01 when Equation 5.5 is fitted in different training period of data using TSS

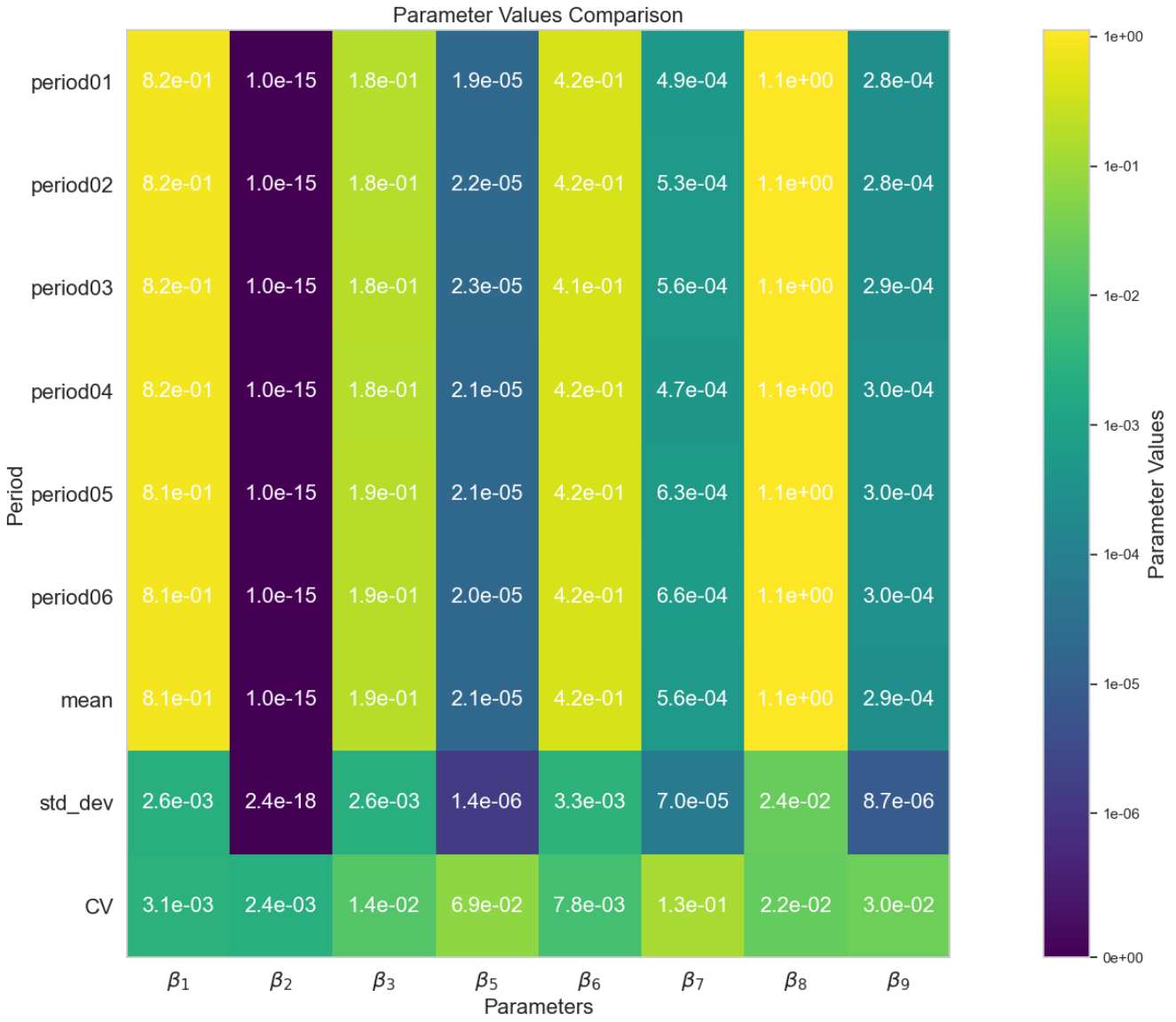


Figure A.3: Calculated parameters for turbine T01 when Equation 5.5 is fitted in different training period of data using BTSS

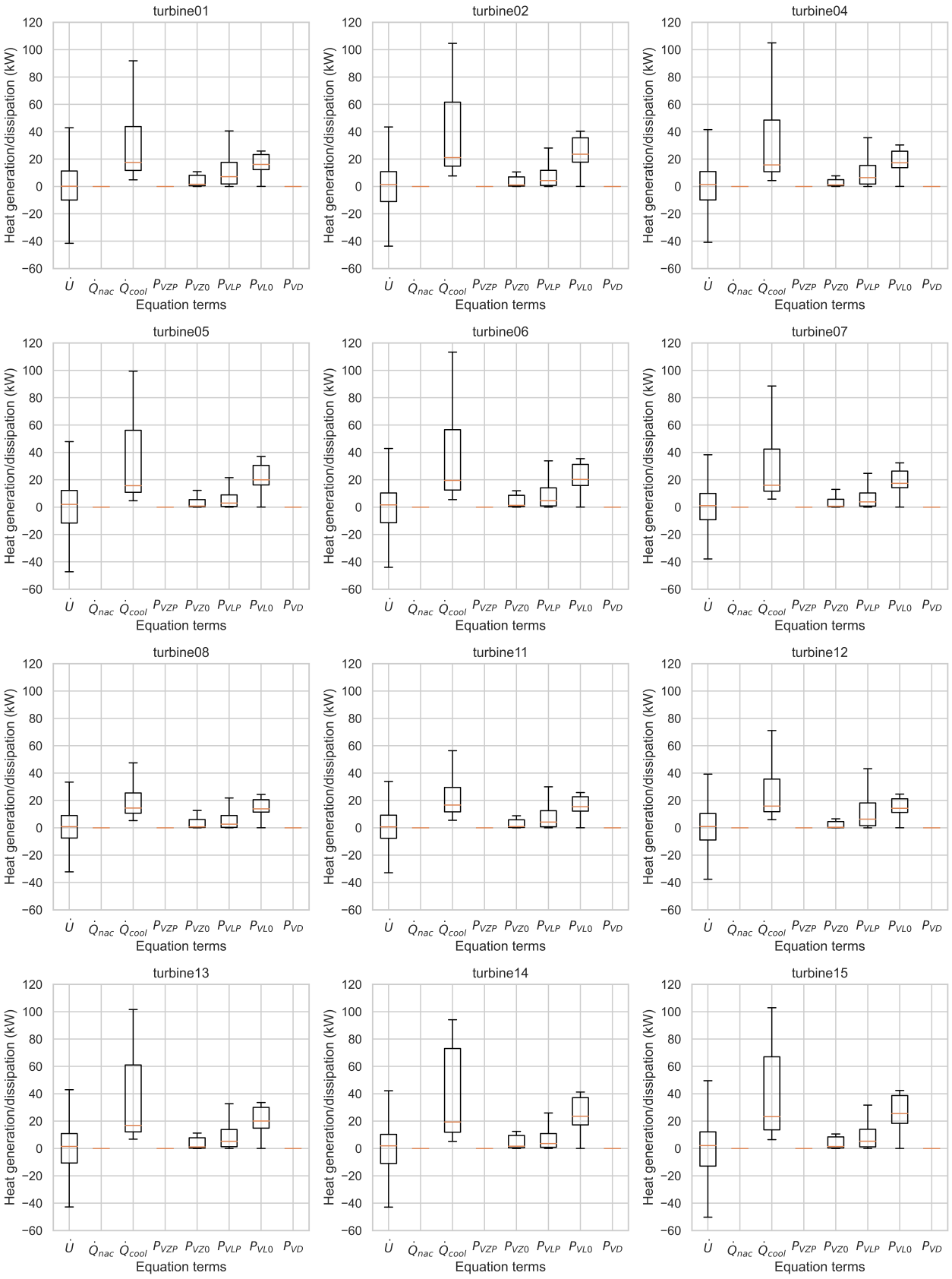


Figure A.4: Heat generation and dissipation terms of heat balance Equation 4.28 for all 12 turbines

A.3 Wind turbine failures

Table A.2: Failures of wind turbines components of EDP dataset

Turbine_ID	Component	Timestamp	Remarks
T11	GENERATOR	2016-03-03T19:00	Electric circuit error in generator
T06	HYDRAULIC_GROUP	2016-04-04T18:53	Error in pitch regulation
T07	GENERATOR_BEARING	2016-04-30T12:40	High temp. in generator bearing (replaced sensor)
T09	GENERATOR_BEARING	2016-06-07T16:59	High temp. generator bearing
T07	TRANSFORMER	2016-07-10T03:46	High temp. transformer
T06	GENERATOR	2016-07-11T19:48	Generator replaced
T01	GEARBOX	2016-07-18T02:10	Gearbox pump damaged
T06	GENERATOR	2016-07-24T17:01	Generator temp. sensor failure
T09	GENERATOR_BEARING	2016-08-22T18:25	High temp. generator bearing
T07	TRANSFORMER	2016-08-23T02:21	High temp. transformer. Transformer refrigeration repaired
T06	GENERATOR	2016-09-04T08:08	High temp. generator error
T06	GENERATOR	2016-10-02T17:08	Refrigeration system and temp. sensors in generator replaced
T09	GEARBOX	2016-10-11T08:06	Gearbox repaired
T09	GENERATOR_BEARING	2016-10-17T09:19	Generator bearings replaced
T11	HYDRAULIC_GROUP	2016-10-17T17:44	Hydraulic error in brake circuit
T06	GENERATOR	2016-10-27T16:26	Generator replaced
T09	GENERATOR_BEARING	2017-01-25T12:55	Generator bearings replaced
T11	HYDRAULIC_GROUP	2017-04-26T18:06	Hydraulic error in brake circuit
T07	HYDRAULIC_GROUP	2017-06-17T11:35	Oil leakage in Hub
T01	TRANSFORMER	2017-08-11T13:14	Transformer fan damaged
T06	HYDRAULIC_GROUP	2017-08-19T09:47	Oil leakage in Hub
T07	GENERATOR_BEARING	2017-08-20T06:08	Generator bearings damaged
T07	GENERATOR	2017-08-21T14:47	Generator damaged
T01	TRANSFORMER	2017-08-11T13:14	Transformer fan damaged
T06	GEARBOX	2017-10-17T08:38	Gearbox bearings damaged
T06	HYDRAULIC_GROUP	2017-08-19T09:47	Oil leakage in Hub
T07	GENERATOR_BEARING	2017-08-20T06:08	Generator bearings damaged
T07	GENERATOR	2017-08-21T14:47	Generator damaged
T07	HYDRAULIC_GROUP	2017-06-17T11:35	Oil leakage in Hub
T07	HYDRAULIC_GROUP	2017-10-19T10:11	Oil leakage in Hub
T09	GEARBOX	2017-10-18T08:32	Gearbox noise
T09	GENERATOR_BEARING	2017-01-25T12:55	Generator bearings replaced
T09	HYDRAULIC_GROUP	2017-09-16T15:46	Pitch position error related GH
T11	HYDRAULIC_GROUP	2017-04-26T18:06	Hydraulic error in brake circuit
T11	HYDRAULIC_GROUP	2017-09-12T15:30	Hydraulic error in brake circuit

A.4 Data-sheets and Reference wind turbine gearbox specifications

Wind Turbine Characteristics

Technical information

POWER	
Rated power (kW)	2 000
Cut-in wind speed (m/s)	4
Rated wind speed (m/s)	12
Cut-out wind speed (m/s)	25
Wind class (IEC)	IEC II (7.5 – 8.5 m/s)
ROTOR	
Diameter (m)	90
Swept area (m^2)	6 362
Number of blades	3
Rotor speed, max (rpm)	14.9
Tip speed (m/s)	70
Power density 1 (W/m^2)	314.4
Power density 2 (m^2/kW)	3.2
GEARBOX	
Type	Planetary/spur
Stages	3
GENERATOR	
Type	Asynchronous
Speed, max (rpm)	2 016
Voltage (V)	690
Grid frequency (Hz)	50
TOWER	
Hub height (m)	80
Type	Steel tube
Shape	Conical
Corrosion protection	Painted

The following tables include information about the reference gearbox found in [71]

Table A.3: Rotational and tangential speed on the gear mesh for an input speed of 20 rpm and maximum Hertz pressures for the reference gearbox

		Stage 1		Stage 2		Stage 3
Property	Unit	P/S	P/R	P/S	P/R	Helical
n	rpm	111.4	34.9	610.4	190.6	610.4
v_t	m/s	1.867	0.974	6.302	3.251	24.933
p_0	MPa	1028	699	921	624	567

P/S: Planet/Sun, P/R: Planet/Ring.

Table A.4: Gear geometric properties of the reference wind turbine gearbox

	Stage 1			Stage 2			Stage 3	
Parameter	Sun	Planet	Ring	Sun	Planet	Ring	Pinion	Wheel
z	21	35	-96	23	38	-103	117	35
b	320	320	331.5	168.4	168.4	177.4	245	240
i	5.587			5.464			3.343	
m	16			9			7	
α_z	20			20			20	
β_z	10			10			10	
a'	476			290			550	
x_z	0.71	0.8031	0.2093	0.6464	0.7693	-0.0639	0.769	0.7176
S_F	1.68	1.19	1.89	1.98	1.39	2.18	2.74	2.91
S_H	1.09	1.15	1.79	1.18	1.22	2.25	2.02	1.99

Table A.5: Rolling bearings of the reference wind turbine gearbox

Stage	Rolling bearing	Location	Quantity
Stage 1	SKF NU 20/800 ECMA	carrier	1
	SKF NU 1080 MA	carrier	1
	SKF NU 2340 ECMA	planets	3
	SKF NU 2340 ECMA	planets	3
Stage 2	SKF NU 244 ECMA	carrier	1
	SKF NU 1060 MA	carrier	1
	SKF NNCF 4930 CV	planets	3
	SKF NNCF 4930 CV	planets	3
Stage 3	SKF NU 1060 MA	pinion shaft	1
	SKF 32960	pinion shaft	1
	SKF 32960	pinion shaft	1
	SKF NU 1036 ML	wheel shaft	1
	SKF NUP 236 ECMA	wheel shaft	1
	NSK QJ 1036	wheel shaft	1