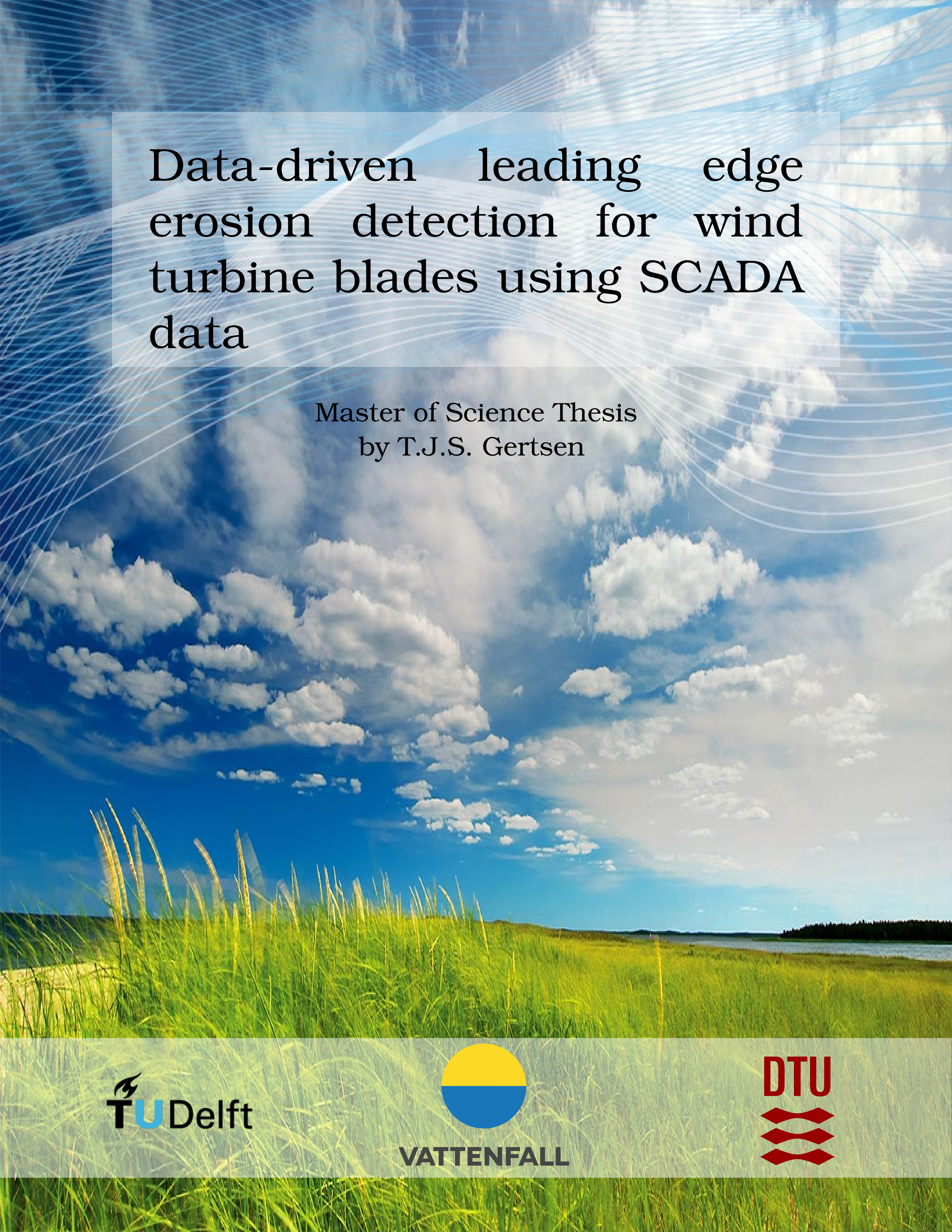

Data-driven leading edge erosion detection for wind turbine blades using SCADA data

Master of Science Thesis

by T.J.S. Gertsen

Data-driven leading edge erosion detection for wind turbine blades using SCADA data

by

T.J.S. Gertsen

to obtain the degree of:
Master of Science in Aerospace Engineering at the Delft University of Technology and
Master of Science in Engineering at the Technical University of Denmark
to be defended publicly on Tuesday August 10, 2021 at 13:00 PM.

Student number:	4224590 (TU Delft)	193042 (DTU)
Project duration:	November 1, 2020 – July 21, 2021	
Thesis committee:	Dr.ir. D. Zarouchas,	chairman & supervisor (TU Delft)
	ir. W.A. Timmer,	examiner (TU Delft)
	N.K. Dimitrov,	supervisor (DTU)
	I. Antoniou,	censor (DTU)
	M. Florian,	supervisor (Vattenfall)

TUDelft

Data-driven leading edge erosion detection for wind turbine blades using SCADA data

Department of Wind Energy Master Report

Thomas Johan Sebastiaan Gertsen

DTU Wind Energy-M-0435

July 2021

DTU Wind Energy
Department of Wind Energy

Author: Thomas Johan Sebastiaan Gertsen

Title: Data-driven leading edge erosion detection for wind turbine blades using SCADA data

DTU Wind Energy-M-0435
July 2021

Project period:
November 2020 – July 2021

ECTS: 45

Education: Master of Science

Supervisor:
Nikolay Krasimirov Dimitrov
DTU Wind Energy

Dimitrios Zaruchas
TUDelft

Remarks:
This report is submitted as partial fulfillment of the requirements for graduation in the above education at the Technical University of Denmark.

Technical University of Denmark
Department of Wind Energy
Frederiksborgvej 399
4000 Roskilde
Denmark
www.vindenergi.dtu.dk

Preface

This thesis is the final part in fulfilling the program of the European Wind Energy Master (EWEM) consisting of a MSc in Wind Energy Engineering at the Technical University of Denmark (DTU) and a MSc in Aerospace Engineering at Delft University of Technology (TU Delft). During the last years of my study I developed an interest and appreciation for the power of coding and data. Especially, the courses of both my University supervisors inspired me to find a thesis subject related to structural health monitoring since this would combine my background in material science and interest in data science. I would like to thank them both, Dimitrios Zarouchas and Nikolay Dimitrov, for guiding me through the process from start to end. They kept me sharp with the right questions and suggestions. Furthermore, this research could not have taken place without the cooperation of all the nice people within the Analytics team of Vattenfall. In special I would like to thank Mihai Florian, my daily supervisor from Vattenfall, who was always available to answer my questions and help me out if necessary. The European Wind Energy Master is a special program taking place at different Universities in different countries. One thing throughout the program stayed the same, however, and that was the group of students with which we had a lot of fun. One person in particular needs to be mentioned: Alex Sanchez Bravo. We were condemned to each other since we were the only two of the Structures and Composites profile. Together we made it to the end. Back in the days spending lots of hours in the library and sometimes even long nights, with pizza, to finish the assignments of the FEM-heavy course. During the thesis, despite we were graduating in two different countries, we kept each other motivated and sharp by starting-up every morning at 9 o'clock with a coffee in front of our webcams. Thanks Alex! I am looking forward to the wine and cheese tastings we promised ourselves.

Rotterdam, the Netherlands, July 21, 2021
T.J.S. Gertsen

Summary

Wind turbines are operating in harsh environmental conditions, especially offshore. An implication of these conditions, caused by the impact of precipitation, is the development of leading edge erosion (LEE). This leads to degraded blade surfaces that result in lower aerodynamic performance. Leading edge erosion is researched in many ways but remote detection is still underdeveloped. Therefore, this thesis investigates the possibility to develop a LEE detection method by analysing real-life data from operational turbines in the field. High frequency data is aggregated to 10-minute statistical data. It is then filtered and corrected for events that are known to cause deviation from normal behaviour. Where possible, the IEC61400-12-1 standard is followed to comply with industry standards.

The hypothesis of this thesis is that, due to the lower aerodynamic performance caused by LEE, apart from the power signal the pitch and rotor speed signal should show signatures of degradation as well. Therefore, the three SCADA[1] signals; power, pitch angle and rotor speed are chosen to monitor the performance of the turbine. A reference period is defined from the start of a turbine's lifetime where it is assumed to be free from LEE. By using the binning method, a reference curve is extracted from this period. Residuals can be computed for incoming measurements that are translated along a linear segment within each bin to the bin-center. Subsequently a normalisation is applied to convert all bins to a standard normal distribution.

Deviating performance of the three different signals of interest are tracked by using control charts. Monitoring the exponentially weighted moving average (EWMA) of the normalised residuals, gives a good indication of long term distribution shifts. Control limits are computed to define a threshold where the signal is said to be out of control.

A Monte Carlo simulation is used to generate synthetic data with an artificial degradation to verify the model. It is necessary to verify the model and evaluate its performance for each individual turbine since it is dependent on the variance of the data in the reference period. Depending on this reference period, the model shows consistently low average detection times for the rotor speed signal up to the minimum degradation tested of 0.5%. The power and pitch signal showed reliable detection, with low false negative rates, for degradation values from 5% and $2-5\%$ respectively.

Underperformance is detected for all three signals using real-life data from operational turbines. However, it can not be validated that this periods are caused by LEE. In fact, the results are suggesting that LEE impact is not significant enough to cause a deviation in the signals compared to the first year of operation which is taken as normal behaving reference period and assumed to be free of LEE. Therefore, it is recommended to perform a validation study on periods of data where LEE is detected during an inspection followed by a repair to investigate the reaction and behaviour of the different signals.

[1] Supervisory Control And Data Acquisition (SCADA), a well-known term in wind industry for the system that gathers all the data of a turbine

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ADT Average Detection Time

AEP Annual Energy Production

API Application Programming Interface

CDF Cumulative Distribution Function

CDS Climate Data Store

CFD Computational Fluid Dynamics

CM Condition Monitoring

EWMA Exponentially Weighted Moving Average

FDM Fault Diagnosis Methods

FNR False Negative Rate

FPR False Positive Rate

HAWT Horizontal Axis Wind Turbine

IEC International Electrotechnical Commission

ISA International Standard Atmosphere

LCL Lower Control Limit

LEE Leading Edge Erosion

MA Moving Average

MC Monte-Carlo

NTF Nacelle Transfer Function

OEM Original Equipment Manufacturer

PPF Percent Point Function

ROI Return On Investment

RUL Remaining Useful Life

SCADA Supervisory Control And Data Acquisition

SHM Structural Health Monitoring

SQL Structured Query Language

TI Turbulence Intensity

TSO Transmission System Operator

UCL Upper Control Limit

Symbols

α Pitch angle [°]

λ Smoothing factor [-]

C_d Drag coefficient [-]

C_l Lift coefficient [-]

C_p Power coefficient [-]

1 | Introduction

Yearly global reports show that wind energy is one of the main energy sources that humankind relies on for supplying an ever growing part of the total energy-demand [Lee and Zhao, 2020]. In Europe alone, it already supplies 15% of the total energy demand [Komusanac et al., 2020]. The vast majority of machines used to capture this wind energy, by converting the kinetic energy of the wind into electrical energy, are horizontal axis wind turbines (HAWT). These wind turbines play a vital role in the transition to a fossil free energy market worldwide. Driving down the costs will keep contributing to the competitiveness of wind turbines within the energy market and therefore to cleaner energy. Unexpected maintenance is one of the causes for extra costs that negatively impacts the return of investment (ROI) for a wind turbine operator.

Numerous papers are written in the area of structural health monitoring (SHM), fault diagnosis methods (FDM) and remaining useful life (RUL) estimations for wind turbines to overcome these unexpected maintenance events [Amirat et al., 2009, Si et al., 2011, Chun Piu Lau et al., 2012, Dhanraj et al., 2016]. Especially, condition monitoring (CM) for generators and gearboxes are well represented in literature [De Azevedo et al., 2016, Hossain et al., 2018]. A plausible explanation is that these techniques are already used in other, much older, industries. Nowadays, data to analyse abnormal behaviour of generators and gearboxes, such as oil temperatures or vibration sensors, are available in the supervisory control and data acquisition (SCADA) system for most turbines [Zaher et al., 2009]. Depending on different studies these two components are responsible for a fair amount of failures and downtime. However, depending on which publication one looks at, a component that appears to be responsible for as much or even more failures and downtime are the blades [Pinar Pérez et al., 2013]. The blades are the most important part in the structure to convert the kinetic energy from the wind into a rotating motion of the whole rotor which in the end gets translated into electrical energy by the generator. Failure of the blades to perform this task or to perform this task less efficient will reduce the amount of energy that is harvested by the turbine. Different damage detection methods have been developed to monitor the integrity of wind turbine blades [Li et al., 2015, Du et al., 2020]. These techniques, however, are not yet widely implemented in industry which means that the vast majority of blade inspections are still traditionally done by visual inspections of experts [Nielsen et al., 2020]. Nevertheless, multiple strategies are developed to cope with this rather qualitative data taken from expert maintenance reports [Leimeister and Kolios, 2018, Nielsen and Sørensen, 2018]. A close link with the applicability of SHM, FDM and CM on blades is with the underlying driver to prevent underperformance of a wind turbine. A phenomenon which is still quite underdeveloped in literature is leading edge erosion (LEE). Damages like cracks and pits that compromise the structural integrity do not necessarily influence the power output and thereby the performance. LEE on the other hand, is not a cause for direct failure and will only compromise the structural integrity in rare occasions when the erosion is accumulated over multiple years and advanced up to a state where the laminate is exposed. Moreover, LEE gradually influences the performance and therefore the annual energy production (AEP)

and ROI of a wind turbine in the long run [Schramm et al., 2017]. Multiple research areas are explored to investigate and mitigate the impact of LEE. These areas are categorised in three directions; prevention, imitation and prediction of LEE. A fourth area for which not much literature can be found, is identified namely the detection of LEE. Therefore, this thesis tries to develop a LEE detection method in order to measure the impact. A data-driven approach is chosen by exploiting the vast amounts of SCADA data produced by turbines nowadays. Successful detection of LEE would make it possible to measure the impact on annual energy production (AEP). From this perspective of underperformance, a business case could be made to calculate if it is worthwhile to repair the blade(s) or not.
The report is structured in the following chapters:

- Chapter 2 covers a summary of the literature study that is performed. As mentioned above several research areas are distinguished to give structure to the variety of literature that is found. This chapter is concluded with a more precise description of the research gap and what methods from literature can be used to build upon.

- Chapter 3 states the formulated research questions and objectives that follow out of the literature study. Multiple sub objectives are mentioned to break up the bigger goal into smaller pieces.

- Chapter 4 covers the data exploration. Different data sources are gathered and discussed in detail. A description of each signal that will be used in the analysis, including frequency, is given.

- Chapter 5 describes how the data is prepared in order to get a data-set that is considered to be in 'normal' operational state. An aggregation of the high frequency SCADA data is used to construct 10-minute intervals which is a standard format used in the wind industry. The IEC61400-12-1 standard is used as a guideline for the rejections and corrections to comply with industry standards. Several techniques are used and developed to filter the data.

- Chapter 6 is devoted to the modelling part. Three signals of interest; power, pitch angle and rotor speed are described which are analysed with a residual model. Control charts are constructed to monitor these signals.

- Chapter 7 is focused on the verification of the model. Two simulation techniques are developed to generate synthetic data. An artificial degradation is used to evaluate the performance of the model.

- Chapter 8 covers the application of the model on real data. Several turbines are evaluated to get a broader view of the actual results and therefore the applicability of the model on real life data.

- Chapter 9 is the final part of this thesis where conclusions are drawn from the results. Shortcomings of this research together with the opportunities are brought together in the recommendations that are made.

2 | Literature study

In this chapter the literature study will be presented together with the current state-of-the-art. Different research fields in leading edge erosion are identified and summarised. The research gap that this thesis tries to fill is presented in the end.

2.1 Current research fields

Leading edge erosion is currently one of the major issues the wind turbine industry is facing [Bartolomé and Teuwen, 2019]. Research shows that the impact velocity of precipitation is the main factor causing the erosion [Keegan et al., 2013]. It is typically described as a monotonic process that starts with small pitting in the coating and without maintenance, finally results in delamination of the structure. Mainly the outer parts of the blades, the tip sections, are affected due to the high relative speeds that are reached. Especially offshore, ever growing rotor diameters cause tip-speeds to increase making the problem even more relevant [Yang et al., 2014, Herring et al., 2019].

Prevention

Multiple perspectives can be identified to contribute to a solution for decreasing or eliminating the negative impact, on the power production of a wind turbine, caused by LEE. The first solution at hand is to improve or modify the blade leading edge material to protect it from erosion in the first place [Chen et al., 2019]. Research is performed in enhancing the protective (gel) coatings that are applied directly into the mould while producing the blades [Slot et al., 2015]. Another option to strengthen the leading edge is by applying dedicated protective tapes or aluminum shielding.

Imitation

A second field of research is aimed at imitating the physical phenomenon by making use of special lab experiment set-ups [Zhang et al., 2015], numerical methods [Keegan, 2014], CFD analysis [Han et al., 2018] or wind tunnel tests [Sareen et al., 2014]. Once a relationship is established, impact calculations on the annual energy production (AEP) are performed. A common finding in all these experiments is that LEE causes a decrease in lift and an increase in drag. The quantification of this drop in lift and increase in drag, however, varies through different studies. The wind tunnel tests performed by Sareen et al. [2014] for three different classes of erosion degradation, three stages of severity per class and for three Reynolds numbers each, showed very clear C_l/C_d and C_l/α curve differences. A drag increase in the range from 6-500% is reported depending on the studied severity of erosion. Substantial decrease of the lift coefficient is mainly experienced at higher angles of attack. It is argued that the annual energy production (AEP) loss ranges from 3-25%. Care must be taken since an assumption is made that the erosion is equally distributed over the whole blade-span whereas literature and field research shows that it is predominantly present at the tips of the blades. By making use of CFD simulations with an eroded blade tip, Han et al. [2018] reported a lift reduction and drag increase up to 53% and 314% respectively. This in turn resulted in an AEP reduction of 2-3.7%.

Prediction

A third area of research is focused on prediction of LEE. Martinez et al. [2019] developed a machine learning (ML) framework by using an extensive database with 84000 defects from approximately 1100 turbines in 17 wind farms. It was concluded that rain is a significant factor in the development of LEE which confirms previous research from Keegan et al. [2013]. However, clear quantified predictions of LEE were not presented. Letson et al. [2020] proposed a generalised framework to design an erosion atlas by making use of open-source noncommercial data. A tool was developed to predict the erosion potential per location to facilitate anticipation in planning of future wind farms.

2.2 Research gap: Detection

A research gap is identified concerning the detection of LEE. Therefore, this is proposed as fourth pillar of research where this thesis will focus on. Inspections of turbines are not planned on a regular basis since, especially off-shore, this comes with high costs. Successful identification and detection of LEE would not only open up the opportunities for operation and maintenance strategies to minimise the impact on ROI, but also monitoring possibilities of the development of LEE which could be useful in the earlier mentioned research areas. Furthermore, it could serve as a sound basis to make predictions on further progression of LEE.

Although a gap is identified there is some research that touches upon the subject and developed some very useful techniques that could be used as a starting point. Caselitz and Giebhardt [2005] developed algorithms to monitor the condition of the rotor including surface roughness. They verified their algorithm describing an icing condition. However, no condition was mentioned where leading edge erosion played a role. Butler et al. [2013] demonstrated the potential of using SCADA data to monitor (power) performance by taking one year of historical data to define an assumed fault-free baseline. This could be called a normal behaviour model (NBM) which is more often used in condition monitoring. Gaussian processes with two inputs, namely 10 minute averages of both wind speed and air density, are used as NBM. Cambron et al. [2016] proposed a method for monitoring the generator by analysing the power curve. The research presented also touches upon detection of underperformance and thereby surface erosion. As Butler et al. [2013] suggested they filter out the data-points above the nominal wind speed to remove interference with pitch control. A reference power curve is constructed using the binning method described in the IEC 61400-12-1 standard[1]. Not only the wind speed is corrected for air density but also a correction factor for turbulence is implemented based on Albers et al. [2007]. Besides that, the measured power values are translated such that the power curve is linearised in each segment in order to use one control chart. Important to mention is that the validation data is simulated with the assumption that all the bins suffer the same relative shift due to underperformance. Although, this could be valid to simulate the underperformance of a generator, it will not hold for underperformance of the blades caused by LEE since earlier discussed research clearly shows wind speed dependent impact due to aerodynamic effects. Nevertheless, it is used to simulate both a step and ramp change where it is claimed that the ramp change amongst other things represents the effect of blade erosion. With this simulated ramp shift a change of 1% could be detected over an average of 234 days. A short section is devoted to an application on real data. A quick conclusion is drawn that the underperformance, which was visible in the control chart for the wind turbine generator, could be explained by leading edge erosion. Although this shows that the method could be useful for detecting LEE, it is concluded that the method itself does not provide the real source of underperformance. Together with the study done by Astolfi et al. [2020] which shows a clear impact of an aged gearbox on the underperformance, it becomes clear that an effort must be taken to separate different factors leading to underperformance of a wind turbine.

Despite the research results on the effects of LEE, that show better signatures in the rated power region due to higher aerodynamic losses, most publications discussed above discard the data which is in the rated power area because of pitch interference. It is interesting to investigate if it is possible to come up with a normalisation of pitch effect or using a pitch reference curve to monitor the degradation in order to minimise data-waste and maximise impact detection. Successful implementation would open up the possibility to design a hybrid model for both the non-rated (partial load) and rated area (full load).

[1] IEC 61400-12-1:2017 Wind energy generation systems - Part 12-1: Power performance measurements of electricity producing wind turbines

3 | Research framework

As stated in the literature review, a research gap is identified concerning detection of LEE. First, a research hypothesis is formed. Thereafter, an approach is outlined to fill the research gap by formulating research questions and objectives set out to be answered in this thesis.

3.1 Research hypothesis

The literature study showed that leading edge erosion causes the lift coefficient (C_l) and drag coefficient (C_d) to decrease and increase respectively. By drawing a vector decomposition of this impact it is straightforward to conclude that the tangential vector, which causes the rotor to rotate, decreases due to the tilted resultant vector (see figure 3.1). A smaller tangential vector means less torque which will result in less generated power for the same wind speed. The additional hypotheses formed in this thesis, is that at the same wind speeds:

- the rotor speed will decrease in the below-rated area, due to LEE, since the pitch angle is kept constant.
- the pitch angle will decrease in the above-rated area due to LEE since the rotor speed and power output are kept constant. Another way to state it is that the pitch angle required to produce the same (rated) power, with respectively lower and higher C_l and C_d values, should be lower.

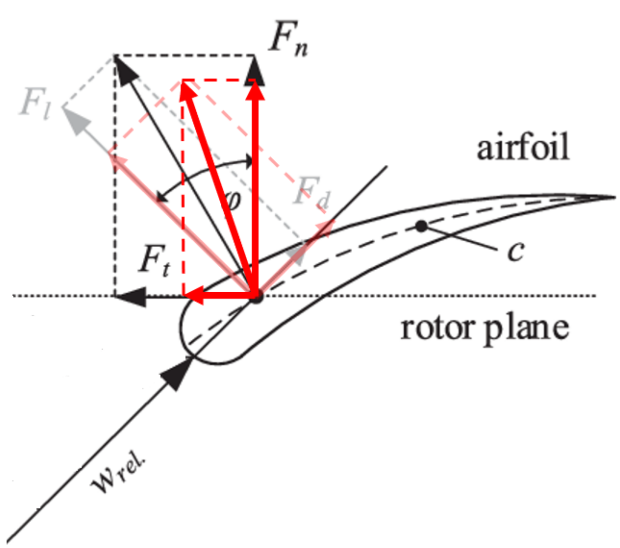

Figure 3.1: Example of vector decomposition with a smaller lift vector and bigger drag vector in red.[1]

[1]Base figure taken from Rommel et al. [2020] under CC BY 4.0 license. Changes are made.

These two signals, rotor speed and pitch angle, are considered to be a more direct indicator of LEE. Solely a lower power output could also be caused by defects or degradation in the generator or gearbox.

3.2 Research question(s)

In order to fill the research gap that is outlined in chapter 2 the main research question is formulated as follows:

> *To what extent is it possible to detect and quantify underperformance of a variable-speed pitch-controlled wind turbine due to leading edge erosion using SCADA data?*

Multiple sub-questions can be formulated to contribute to the answer of the main question:

1. To what extent is it possible to isolate the impact of LEE on the performance?
2. Is it possible to find LEE signatures in a wind speed versus pitch angle curve?
3. Is it possible to find LEE signatures in a wind speed versus rotor speed curve?
4. Is it possible to detect a monotonic shift in the data despite the complex environment and the induced uncertainty in SCADA signals?

Due to the possibility that multiple wind turbine components can degrade over time it is possible that not only LEE is contributing to the underperformance of the wind turbine (subquestion 1). In more common power curve analysis research, the above rated wind speed region is discarded due to the interference of pitching the blades. As mentioned in chapter 2 this research tries to investigate the possibility to use the pitch and rotor speed signal as monitoring entities to exclude other possible degradation factors such as the generator or gearbox (subquestion 2 & 3). Since the development of leading edge erosion is irreversible and not self-healing it is expected that the effect on performance will be monotonic (subquestion 4).

3.3 Research objective(s)

The ultimate objective of this thesis research that follows out of the identified research gap together with the research questions stated in the above section, is:

> *To design a leading edge erosion indicator in order to calculate the energy production loss by exploiting SCADA data using data-driven models.*

In order to achieve this main objective, multiple sub-goals are defined. First, the available data needs to be explored, processed and structured. When the right data is gathered, known deviations such as yaw errors, pitch errors, idling phase and transition phase data-points must be filtered out in order to maximise the accuracy. Corrections must be applied to account for different air densities and turbulence intensities. After constructing a reference curve to compare new measurements with, residuals can be calculated. The residuals will be scattered and need to be normalised in order transform all residuals within different bins to a standard normal distributions. Underperformance can then be identified by constructing control charts, with control limits, to monitor the shift in a distribution. A novel LEE indicator can be designed to relate the underperformance with energy production loss.

4 | Data exploration

The following chapter describes the data that is used during this thesis research. Different data sources are coupled to make this research possible which is desribed in section 4.1. The most extensive used database is the one from Vattenfall where the SCADA signals and alarms are fetched from (section 4.2). To make use of homogeneous environmental data, across wind farms, the Climate Data Store of the European Union is used to fetch hourly data from the ERA5 reanalysis (section 4.3). Various inspection reports were available and are discussed in section 4.4.

4.1 Availability

This thesis research is performed in cooperation with Vattenfall[1]. The author was part of the analytics team within the off-shore wind unit. That created a lot of opportunities to explore the data, owned by Vattenfall. Full access was provided to SCADA data and partial access to inspection data. Both quantitative and qualitative sources were present. The SCADA data, which is a common collective name in the wind industry for the data signals generated by each individual turbine, was stored in a database which could be entered via Structured Query Language (SQL). Inspection reports were available which made it possible to check whether a turbine suffered from LEE or not. In addition to the data of Vattenfall the Climate Data Store (CDS) from the European Union is used to fetch environment data. The CDS database is available through an Application Programming Interface (API) which makes it easy to request the data through Python.

4.2 SCADA data

Supervisory Control And Data Acquisition (SCADA) is a broad term in the wind industry that describes in general almost all information that is taken from a wind turbine. To narrow down the actual data used from this extensive system section 4.2.1 describes all relevant high frequency signals which are direct measurements from the turbine itself. Whereas section 4.2.2 describes logged time intervals with registered events.

4.2.1 High frequency signals

The data was available in a long format table. To speed up the query to request the data for the life-time of a turbine the code automatically separates it into smaller requests per month which are aggregated and optimised to reduce the storage size by approximately 98%. This makes it possible to store it efficiently in local cache files. The following signals were fetched:

[1] Utility company active in multiple countries in western Europe: `https://group.vattenfall.com/`

Nacelle mounted anemometer wind speed

The nacelle mounted anemometer wind speed is available for every turbine which makes it easy to work with. There are, however, some drawbacks. As the measurement device is located behind the blades it is influenced by a turbulent flow. This is partly solved by using a Nacelle Transfer Function (NTF) which is determined by the OEM during the instalment of a turbine. The NTF is nothing more then determining a linear relation between an anemometer which is located in the free stream wind speed, usually on a metmast, and the nacelle mounted anemometer. By using the nacelle anemometer of each individual turbine it does not matter, in the scope of this research, if it suffers from a systematic error since historic data is used as a reference which should include this error as well. Since a ten minute average is used in this research the random error reduces too. Furthermore, the moving averages constructed in the final control charts will act as a low pass filter reducing the frequency and thereby improving the stability.

Yaw angle

The yaw angle is a calibrated signal with north as zero point. A clockwise and anti-clockwise direction is used to describe the spectrum from $0°$ to $180°$ and $0°$ to $-180°$.

Wind direction

To measure the wind direction a weather vane is located at the top of the nacelle. This vane measures the direction with respect to the nacelle orientation which is measured and logged in the yaw angle signal. By combining these two measurements the wind direction can be computed.

Power

The power output of a turbine is measured since this is the end product and purpose of energy harvesting. Furthermore, this signal is also used to control certain aspects of the turbine such as the pitch angle.

Pitch angle

When the power is reaching its rated value[2] the pitch system is activated. Actuators are pressurised in order to rotate the blades "away" from the wind to relieve loads and thereby keeping the power output stable. The effective rotation is logged as pitch angle to know how many degrees a blade is pitched at a certain moment in time. Every individual blade has its own signal which means that there are three different pitch signals.

[2]The maximum power a turbine is allowed to produce, a design limit.

4.2.2 Other signals

Other signals were available per event and are present to flag known deviations from normal. A time range with a begin and end timestamp masks the affected time period. The two signals that were used are:

Curtailment

Curtailment is defined as restricting the power output of a turbine or farm. Curtailment can be applied by different parties. The Transmission System Operator (TSO) can enforce curtailment due to capacity problems of the grid or other grid associated issues. An OEM can enforce curtailment, for example, due to mechanical problems with an individual turbine. Finally, the operator itself can also curtail farms or turbines, not only for mechanical issues but for financial reasons as well. The curtailment flags in the database have a begin and end timestamp.

Alarms

Every turbine is prone to defects and errors. By using pre-programmed checks and protocols most of these defects and errors are captured by specific alarms. All these alarms are captured in the database with a begin and end timestamp.

4.3 CDS data

The Climate Data Store (CDS)[3] from the European Union is freely accessible and provides amongst other things a reanalysis with hourly weather data from 1979 up till now, called ERA5. In order to use consistent weather data for different wind parks several signals are fetched through the ERA5 API. These environmental signals are used to correct for changing conditions in order to make the SCADA signals as comparable as possible through time. To correct for air density the pressure and relative humidity are needed. The relative humidity can be calculated with the temperature and dew point temperature. CDS has an excellent data catalogue with high quality parameter descriptions:

Sea level pressure

"This parameter is the pressure (force per unit area) of the atmosphere at the surface of the Earth, adjusted to the height of mean sea level. It is a measure of the weight that all the air in a column vertically above a point on the Earth's surface would have, if the point were located at mean sea level. The units of this parameter are pascals (Pa)."[4]

[3] https://cds.climate.copernicus.eu

Temperature

"This parameter is the temperature of air at 2m above the surface of land, sea or inland waters. It has units of kelvin (K)."[4]

Dew point temperature

"This parameter is the temperature to which the air, at 2 metres above the surface of the Earth, would have to be cooled for saturation to occur. It is a measure of the humidity of the air. Combined with temperature and pressure, it can be used to calculate the relative humidity. This parameter has units of kelvin (K)."[4]

Hourly data from the ERA5 reanalysis is used since it is the finest sub set available. This means that later on a mapping from one hour to 10-minute SCADA signal measurements is needed.

4.4 Inspection reports

The partial access to inspection reports revealed blade defects, such as leading edge erosion, that were supplemented with pictures. It was possible to relate the defects to a specific turbine blade through an internal identification code describing the park and turbine location. Since no access was available for service and repair data it was not possible to validate the model with real-life events and to prove or disapprove the detection of leading edge erosion at certain moments in time. However, suggestive qualitative conclusions are drawn by taking the available pictures of leading edge erosion, together with the date of inspection, and comparing the signals of different turbines within that specific park.

[4] `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview`

5 | Data preparation

Multiple steps are needed to go from raw data to data that is suitable for modelling. At first some practical modifications were needed in order to make it easier to work with the data (section 5.1). Next is to aggregate the high frequency signals to 10-minute statistical data (section 5.2). It is necessary to reject certain data to be sure that later on the inputs that go into the model are free from (known) errors and non-physical values (section 5.3). In order to extract a degradation pattern, as purely as possible, caused by leading edge erosion, it is important to exclude other external factors that might cause a deviation from a normal reference state. Certain phenomena are corrected (section 5.4) whereas other external factors need to be filtered out (section 5.5). Various rejection, correction or filtering strategies are taken from or inspired by the IEC 61400-12-1 standard. The impact of all preparations are shown in example figures of a power curve. The power curve is a widely used relation in the wind industry to characterise a turbine's power output. It is the relation between the power (P) produced by the turbine and the wind speed (v) at that time which is described by equation 5.1.

$$\mathrm{P} = \frac{1}{2}\rho \mathrm{C_p} \mathrm{A} \mathrm{v}^3 \tag{5.1}$$

Where A is the swept area of the rotor, ρ the air density and C_p the power coefficient. However, due to operational constraints this relation only holds between the cut-in and rated wind speed as shown in figure 5.1. The cut-in wind speed is the moment when the turbine is able to start-up. According to the cubic law (eq. 5.1) the power increases up till the rated wind speed where it reaches rated power which can be seen as a design limit. From that point the turbine will pitch its blades to keep the power output steady on rated power until it reaches its cut-out speed where the turbine is shut down.

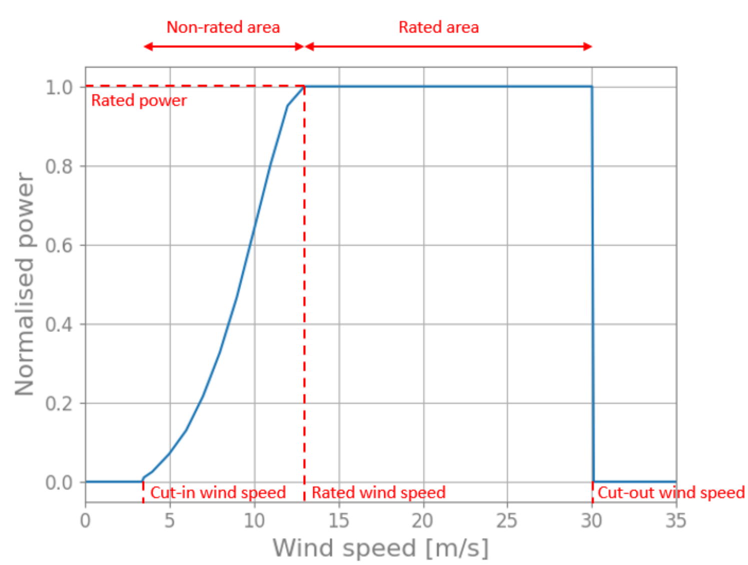

Figure 5.1: Theoretical power curve with characteristic points.

5.1 Modifications

Some of the raw signals, queried from the Vattenfall database and the CDS, were not suitable for direct use. Therefore, a few modifications were necessary to transform it into useful signals.

Angular SCADA signals

There are multiple angular signals that are used in the model which are prone to discontinuity errors when the data is aggregated. Therefore, the signals are transformed into vectors before the aggregation takes place. The composition of both the horizontal and vertical vector are then used to reconstruct the mean angular value. The angular SCADA signals that were transformed with this procedure are:

- Wind direction
- Yaw angle

Pitch signal

The pitch signal is logged for every single blade. To reduce the complexity of the problem a decision is made to reduce these three signals into one pitch signal by taking the mean. More in depth analysis of detecting leading edge erosion for each individual blade might be possible in the future by not doing so.

Environmental signals from CDS

As mentioned in 4.3, three environmental signals are taken from the Climate Data Store (CDS). Both temperature and pressure need to be corrected for a height difference to be representative at hub-height of the turbine under investigation. According to the International Standard Atmosphere (ISA) the temperature can be corrected as follows:

$$\mathrm{T_{corrected}} = \mathrm{T_{sea\text{–}level}} + (\mathrm{h} \cdot \mathrm{L}) \tag{5.2}$$

with

$$\mathrm{L} = -0.0065\ [^{\circ}\mathrm{C/m}]$$

where T is the temperature in celsius, h is the hub-height in meters and L is the lapse-rate in degree celsius per meter. The pressure is corrected using the barometric formula:

$$\mathrm{P_{corrected}} = \mathrm{P_b} \left(\frac{\mathrm{T_b} + \mathrm{L_b} \cdot (\mathrm{h} - \mathrm{h_b})}{\mathrm{T_b}} \right)^{\frac{-\mathrm{g_0} \cdot \mathrm{M}}{\mathrm{R} \cdot \mathrm{L_b}}} \tag{5.3}$$

where b is a subscript for reference values, P is the pressure in Pascal, T is the temperature in Kelvin and h is the height in meters. The following constants are applicable for the barometric formula:

Symbol	Value	Description
L_b	-0.0065 [K/m]	lapse rate
g_0	9.80665 [m/s^2]	gravitational constant
M	0.0289644 [kg/mol]	molar mass of Earth's air
R	8.3144598 [J/(mol · K)]	universal gas constant

The relative humidity is calculated using the temperature and dew point temperature by computing the ratio of both saturation vapor pressures using the Buck equation:

$$P_s = 6.1121 \cdot \exp\left(\left(18.678 - \frac{T}{234.5}\right)\left(\frac{T}{257.14 + T}\right)\right) \tag{5.4}$$

where P_s is the saturation vapor pressure and T is the temperature.

5.2 Aggregation

The high frequency SCADA signals are aggregated. It is common in the wind industry to aggregate this high frequency data to 10-minute statistical data for multiple reasons. This means that sections of 10 minutes are aggregated to extract the mean, standard deviation, minimum value, maximum value and a count of how much measurements are aggregated. This improves the speed in which one can handle vast amounts of data but also reduces the random error caused by the complex environment in which turbines are operating. It smooths the signal so to say. Furthermore, it is not rare that the data is incomplete and therefore aggregation makes the data set more robust. The drawback of aggregating data is that one loses information of fast changing phenomena (higher frequencies). Since this research is looking into the long term trend or shift of a distribution from a specific signal this drawback is accepted.

A more practical advantage of this aggregation is the handling size of the data set for the complete lifetime of a turbine. Transforming the high frequency data into 10-minute statistical data decreases the size by approximately 95%. By using another datatype optimisation in python it was possible to decrease the size even further by more than 50%. This combination made it possible to compress a month of data in approximately 1MB. So for a turbine lifetime of ten to fifteen years the data set comprises $120-180$MB which is still easy to handle on an average laptop nowadays.

5.3 Rejections

There are a different reasons to reject certain measurements. This section elaborates on these reasons.

Missing data

With almost no exception databases are not 100% complete which means that missing data needs to be handled or ignored. Depending on the signal and the amount of missing data points in a row a filter is applied. The following list of crucial signals that need to be present is composed in order to be able to fulfil the analysis:

- Wind speed
- Power
- Pitch

Another possibility to consider is to preserve more data by interpolating missing values. A drawback of interpolation is that it becomes an estimation instead of a measurement. Since the analysis spans over a range of years instead of days or months it is chosen to stick to filtering instead of interpolation.

Operating range

Every turbine has an operating range imposed by the design limits. This translates into a cut-in and cut-out wind speed. At cut-in wind speed the turbine is able to start producing power. The cut-out wind speed is a maximum at which the blades are already pitched at its maximum and therefore the only option to relief loads is to shut down the entire turbine. Both wind speeds are enforced to keep a data set that falls in between these two values.

Free wind sector(s)

Turbines are not rarely surrounded by obstacles or neighbouring turbines. Since this analysis is focused on off-shore wind parks the latter is accounted for. By retrieving the geographic coordinates of all turbines in the park it is possible to compute the free wind sector(s) for the turbine of interest. The method is based on the IEC 61400-12-1 standard, Annex A. A schematic overview is given in fig. 5.2.

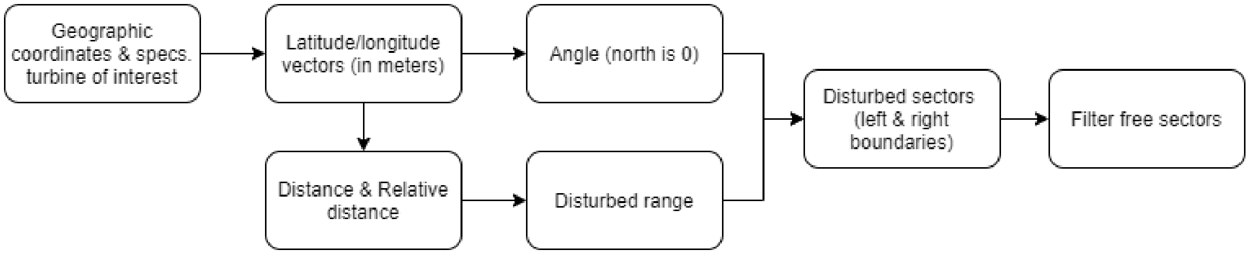

Figure 5.2: Schematic overview of filtering the free wind sector(s).

A complete wind rose example can be seen in figure 5.3 whereas figure 5.4 shows a "left-over" wind rose after applying the filtering method.

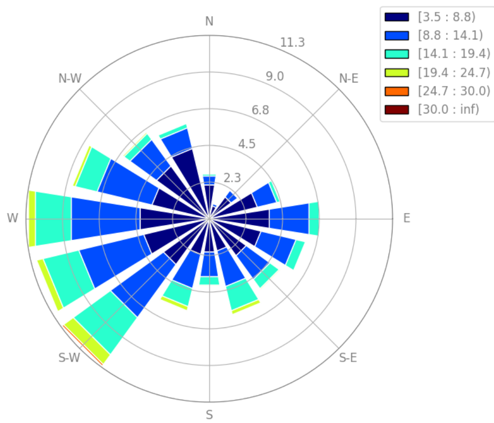

Figure 5.3: Example of a wind rose for a specific park.

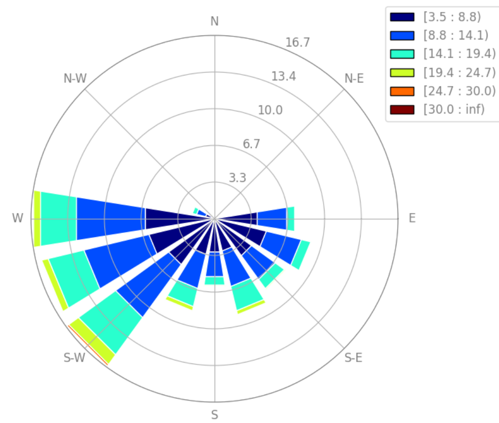

Figure 5.4: Example of the "left-over" wind rose after applying the free wind sector filter.

5.4 Corrections

In order to bring all measurements to a comparable state two corrections are applied according to the IEC 61400-12-1 standard.

5.4.1 Air density correction

The air density varies through different seasons which makes it necessary to correct for it. The IEC 61400-12-1 standard prescribes a procedure to correct for different air densities by correcting the wind speed using a sea-level reference air density with the following equation:

$$V_C = V_M \left(\frac{\rho}{1.225}\right)^{\frac{1}{3}} \tag{5.5}$$

with

$$\rho = 1.225 \left(\frac{288.15}{T}\right)\left(\frac{B}{1013.3}\right) \tag{5.6}$$

Where V_C and V_M are, respectively, the corrected and measured wind speed in m/s. ρ is the air density with T as ambient temperature in Kelvin and B as atmospheric pressure in mbar. An example of the impact is shown in fig. 5.5. It appeared that, for the wind farm investigated in this research, the correction had a small impact.

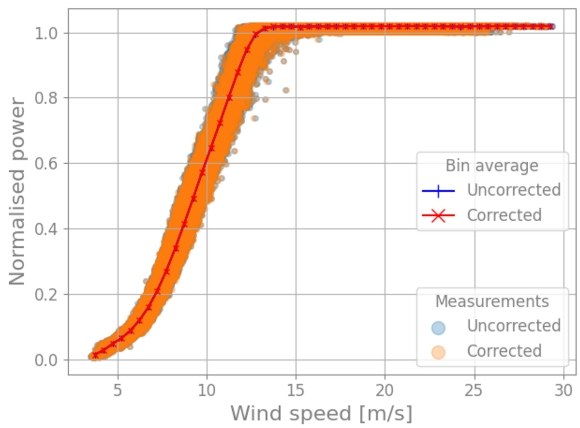

Figure 5.5: Example power curve with and without air density correction.

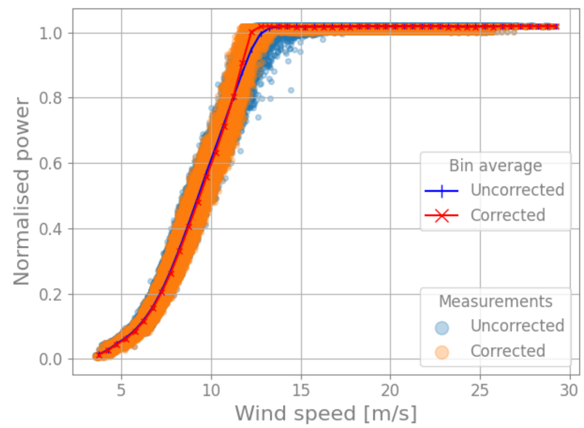

Figure 5.6: Example power curve with and without turbulence correction.

5.4.2 Turbulence correction

A turbulence correction is applied to counteract the negative side-effect of 10-minute averaging of the data. The goal is to make measurements at different turbulence intensities more comparable. Turbulence is defined as the unsteady movement of air. In the wind energy industry this is quantified using the Turbulence Intensity (TI) measure defined as:

$$TI = \frac{\sigma_v}{v} \tag{5.7}$$

Where v is the average wind speed and σ_v the wind speed standard deviation of a 10-minute period. As explained in the beginning of this chapter, the relation between power (P) and wind speed (v) is explained by:

$$P = \frac{1}{2}\rho C_p A v^3 \tag{5.8}$$

Where A is the swept area of the rotor, ρ the air density and C_p the power coefficient. As this relation is non-linear, the 10-minute averaging skews the measured power output, below rated wind speed, in two ways:

- The lower area of the power curve where the power increases proportionally with the wind speed (the so-called ankle); the power output increases with an increasing turbulence intensity.

- The upper area of the power curve where the power does not increase proportional to the wind speed anymore (the so-called knee); the power output decreases with an increasing turbulence intensity.

The method to correct for this turbulence is nicely described in the IEC 61400-12-1 standard, Annex M. With this method, depending on the lifetime of the turbine, one would need to integrate hundreds of thousands measurements to correct the complete data set. Using ordinary integration functions in python is therefore impossible. By making use of a Monte-Carlo method this process is still possible to perform on an ordinary laptop. A more elaborate explanation of the programming solution is documented in appendix section A.2. An example of the impact is shown in fig. 5.6.

Assumptions made in this normalisation procedure are[1]:

- At each moment the wind turbine follows a certain power curve that is independent of the turbulence intensity.

- Wind speed fluctuations over the entire rotor area are characterised by the TI at hub height.

5.5 Filtering

The goal with filtering the data is to get a set of measurements that are in a "normal" comparable state. Therefore, measurements that are deviating from this normal state caused by external factors, are filtered out. There are different causes why a measurement is not considered as normal. These causes and the ways of filtering are described in the following sections.

Alarms

As discussed in section 4.2.2 all kinds of different alarms will occur during the lifetime of turbine. These alarms are logged in the database with a begin and end timestamp in order to be able to filter the data such that known erroneous measurements can be removed. The time span between these two timestamps is converted to affected 10-minute intervals to be able to flag the 10-minute SCADA measurements. An example for an easy programming solution is documented in appendix A.1. Figure 5.7 shows an example of a turbine where the affected intervals, caused by alarms, are flagged. For the scope of this research the flagged

[1] IEC 61400-12-1, Annex M: Power performance measurement of electricity producing wind turbines

measurements are removed to be sure that no affected data is present. Alternatively, one could track down certain alarm codes, by using an alarm guide, to be more specific and make a more precise distinction between useful and not useful data.

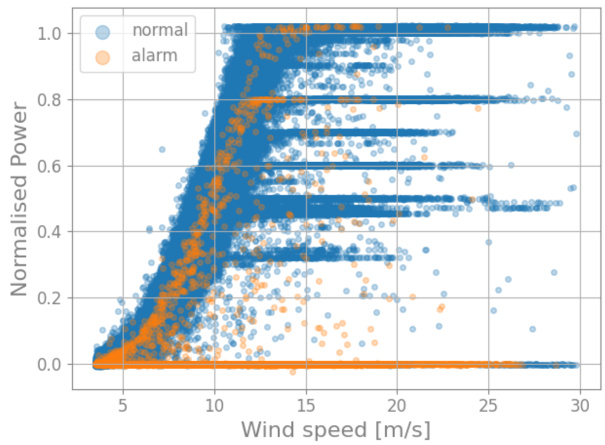

Figure 5.7: Example of flagged alarm measurements.

Curtailment

In the non-rated area of the power curve the rotor speed is increasing with higher wind speeds and the pitch angle of the blades should be constant under an optimised design angle which is often slightly negative. Curtailment data is filtered by applying logicals based on pitch activity by checking the data for not exceeding a pitch angle of higher than zero degrees and a maximum power that is below the rated power. Since curtailed data is not representative for the analysis all these flagged points are filtered out (see figure 5.8).

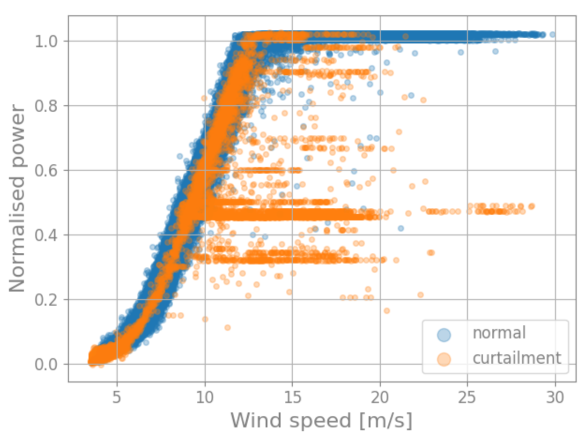

Figure 5.8: Example of flagged curtailments.

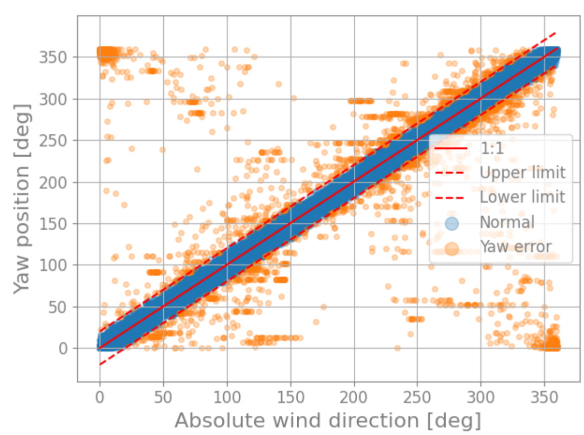

Figure 5.9: Example of flagged yaw misalignments.

Yaw misalignment

By taking the absolute difference between the wind direction and the yaw angle it is possible to compute the yaw misalignment. This measure indicates the degrees of misalignment of the turbine with the wind. From an aerodynamic point of view this will cause different performance than a fully aligned turbine. A threshold of 5 degrees is set to flag a measurement as yaw misalignment. Figure 5.9 shows an example of the variety of measurements. For illustration purposes a threshold of 20 degrees is used to construct this figure.

Non-operational phases

From an operational point of view one can imagine that a turbine has different stages in which it operates. The fact that it has a cut-in and cut-out wind speed means that there are limits to which the turbine can operate and produce energy. Three phases are specified:

- Idling phase
- Transient phase
- Operational phase

The idling and transient phase must be flagged as it is not qualified as normal data. Idling is defined as a rotating turbine without delivering power. This is done when the turbine is not allowed or able to enter the operational phase, as alternative to enabling the brake in order to relieve the rotor from unnecessary loads. For the sake of simplicity, measurements that are both not rotating and not producing energy are also marked as idling phase. Figure 5.10 shows flagged measurements that are identified as falling in the idling phase. The logical rules that are used for every 10-min interval are:

$$\text{Power}_{\text{max}} <= 1 \text{ [kW]}$$
$$\text{Power}_{\text{min}} <= 1 \text{ [kW]}$$

The transition from idling to operational is called the transient phase. It is tracked by looking at the minimum and maximum power values of a 10-minute measurement. The maximum power should be above 1kW where the minimum should be less than or equal to 1kW. In figure 5.11 an example of measurements that fall in the transient phase are shown.

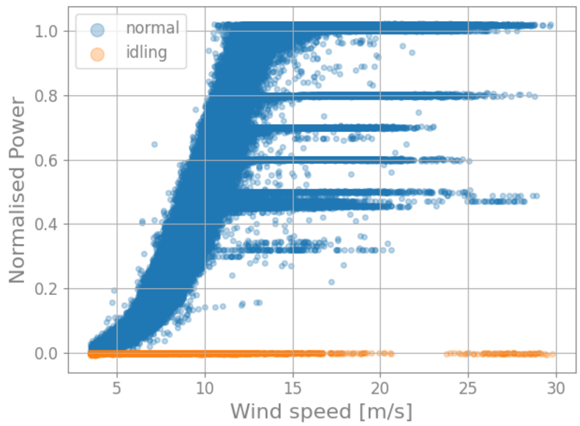

Figure 5.10: Example of flagged idling phase measurements.

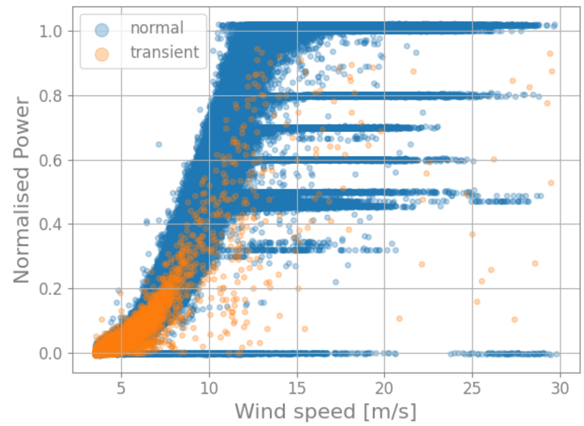

Figure 5.11: Example of flagged transient phase measurements.

Outliers

This research is focused on longer term distribution shifts and not on single measurements that deviate. Therefore, as last step in the filtering process the Local Outlier Factor (LOF) algorithm, proposed by Breunig et al. [2000], is used to remove outliers. The algorithm is based on local density where local is defined by its 'k' nearest neighbors. A value of 20 is chosen for 'k' which resulted in approximately 1% of additional data that was filtered. Figure 5.12 shows an example of the effect.

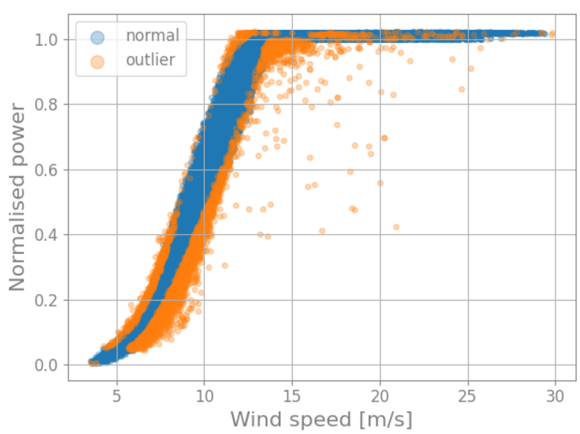

Figure 5.12: Example of flagged outliers by the local outlier factor (LOF) algorithm.

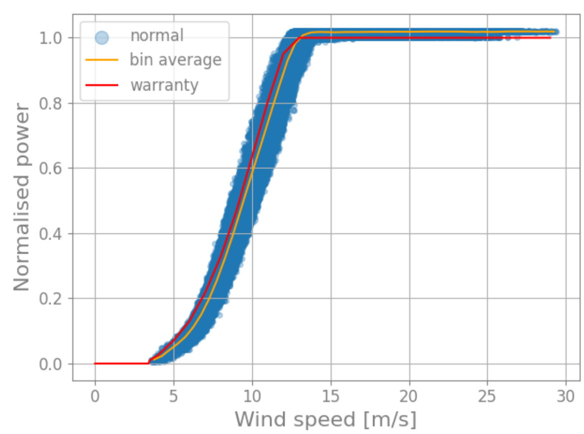

Figure 5.13: Example of a final clean power curve after all filters.

Final clean power curve

Finally a clean power curve is obtained as can be seen in figure 5.13. The actual 10-minute measurements follow the same trend as the warranty curve but do show a certain spread. The bin average line shows the average value of this spread for a width of 0.5m/s bins.

6 | Data modelling

After extensive preparation of the data, the modelling can start. First, the motivation behind choosing the three signals of interest is explained. Thereafter, the method to compute residuals and at last, the monitoring method using control charts.

6.1 Signals of interest

Three SCADA signals are chosen to monitor the performance of a turbine: (active) power, rotor speed and pitch angle. Active power is the most direct signal to indicate underperformance since the goal of a wind turbine is to produce an "x" amount of power at wind speed "y". However, a degradation in the power signal can be caused by many things and therefore it is hard to relate it directly to leading edge erosion. The research of Sareen et al. [2014] shows that due to LEE the C_l and C_d values[1] of a blade at the same wind speeds will drastically decrease and increase respectively. The hypotheses described in section 3.1, and repeated here, are stating that due to the loss of efficiency at the same wind speed:

- the rotor speed will decrease in the below-rated area, due to LEE, since the pitch angle is kept constant.
- the pitch angle will decrease in the above-rated area due to LEE since the rotor speed and power output are kept constant. Another way to state it is that the pitch angle required to produce the same (rated) power, with respectively lower and higher C_l and C_d values, should be lower.

Therefore, the two signals, rotor speed and pitch angle, are suspected to be a more direct measure of the performance of the blades and could therefore potentially show a better signature of LEE. For each signal a region of interest is defined as can be derived from the hypotheses mentioned above. The region of interest for both the power and rotor speed signal is the below-rated area since pitch interference needs to be excluded. It naturally follows that for the pitch signal the above-rated area is defined as the region of interest.

6.2 Residual model

In order to compute residuals a reference is needed. The complex environment in which a turbine is operating makes it impossible to draw conclusions from individual measurements. Although the data is prepared extensively, random variation will always be present. The solution to overcome this is to take the data of the first operational year as a reference. It is assumed that in this first year of operation, no leading edge erosion is present or has developed yet. In this way the reference is a distribution covering the random variation

[1] Lift and drag coefficient, respectively.

that is "normal" for the complex environment. The binning method is chosen to construct a reference curve. Bins of 0.5m/s are taken as advised in the IEC61400-12-1 standard. An advantage of this method is that it is straightforward and easy to explain which makes it more attractive to industry.
After the reference curve is constructed all measurements within each bin are translated, along a linearised segment from the left bin edge to the right bin edge, to the bin-center (see fig. 6.1-6.3). It is assumed that the data is normally distributed after the translation.

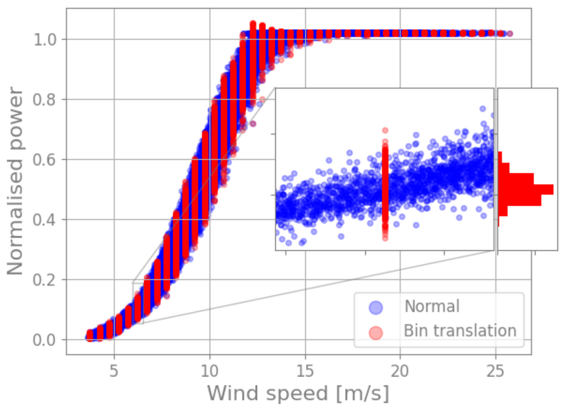

Figure 6.1: Example of bin translation for power signal.

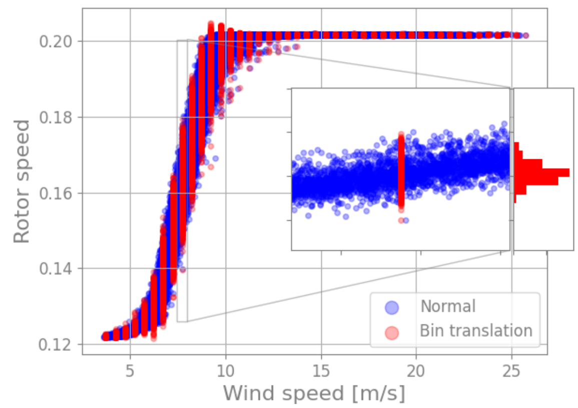

Figure 6.2: Example of bin translation for rotor speed signal.

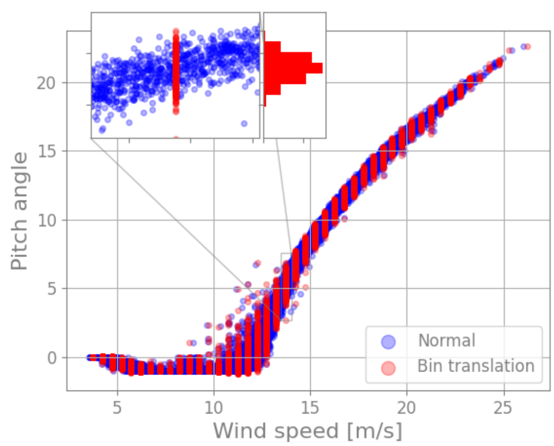

Figure 6.3: Example of bin translation for pitch signal.

Residuals can now be computed by subtracting the bin-center of the reference curve from the translated bin values. In order to treat all residual data in one monitoring system each bin is normalised to get standard normally distributed residuals.

6.3 Control charts

The end goal of using control charts, to monitor the normalised residuals, is to eliminate variability in the process by assigning causes to "special variation" (outside control limits) in contrast to "normal variation" (within control limits) which is inherently present in real life processes. In other words it distinguishes abnormal variation from background noise. For the purpose of detecting leading edge erosion, control charts are used to monitor the three SCADA signals mentioned in section 6.1; (active) power, rotor speed and pitch angle.
As mentioned in section 6.1 each signal has its region of interest for which the data is filtered before computing the control charts. Since this research is focused on a long term shift in the distribution of the signal it is additionally smoothed with a moving average (MA) over a month (30 days). This was necessary to reduce the noise to an acceptable level. A drawback of this operation is that the signal reacts slower to a change and thereby increases the chance on false negatives (Type II error). However, after the incubation of mass loss the phenomenon of LEE is characterised by a monotonic linear process (Bartolomé and Teuwen [2019]) which means type II errors induce a higher average detection time (ADT) instead of no detection. Nevertheless, up to which extent it is possible to detect degradation is dependent on the variability of the reference year. An exponentially weighted moving average (EWMA) is finally computed to monitor the signal in a control chart. In statistical process control (SPC) the EWMA is a well known method and firstly introduced in 1959 by S. Roberts, later on republished in Roberts [2000]. It is characterised by the following equation:

$$z_i = \lambda \bar{x}_i + (1 - \lambda) z_{i-1} \tag{6.1}$$

with

$$0 < \lambda \leq 1$$

where z_i is the value of the EWMA at time period i, $\bar{x}_i$ is the incoming measurement from the process and λ the smoothing factor. The smoothing factor can be tweaked to find a balance in minimising the ADT and both the type I & II errors. z_1 is set to be the average of the first day of 10-min, which comes down to 144 measurements. The control limits are computed according to the following equations:

$$UCL = \mu_0 + k \cdot \sigma_{z_i} \tag{6.2}$$

$$LCL = \mu_0 - k \cdot \sigma_{z_i} \tag{6.3}$$

with

$$\sigma_{z_i} = \sigma_{\bar{x}} \sqrt{\frac{\lambda}{(2 - \lambda)}} \tag{6.4}$$

where μ_0 is zero since the residuals are standard normally distributed, k is chosen to be 3 which is often used in statistical process control, σ_{z_i} is the standard deviation of z_i and $\sigma_{\bar{x}}$ is the process standard deviation taken from the reference year. UCL and LCL are abbreviations for the upper and lower control limit respectively. Figure 6.4 shows an example of a control chart for the power signal. Essentially the control chart is a continuous way of testing a null hypothesis. The null hypothesis is defined as the process being in a state of statistical control. When the signal is in between the two control limits it is failing to reject the null hypothesis. Crossing one of the control limits is equivalent to rejection of the null hypothesis that says it is in statistical control [Montgomery, 2020].

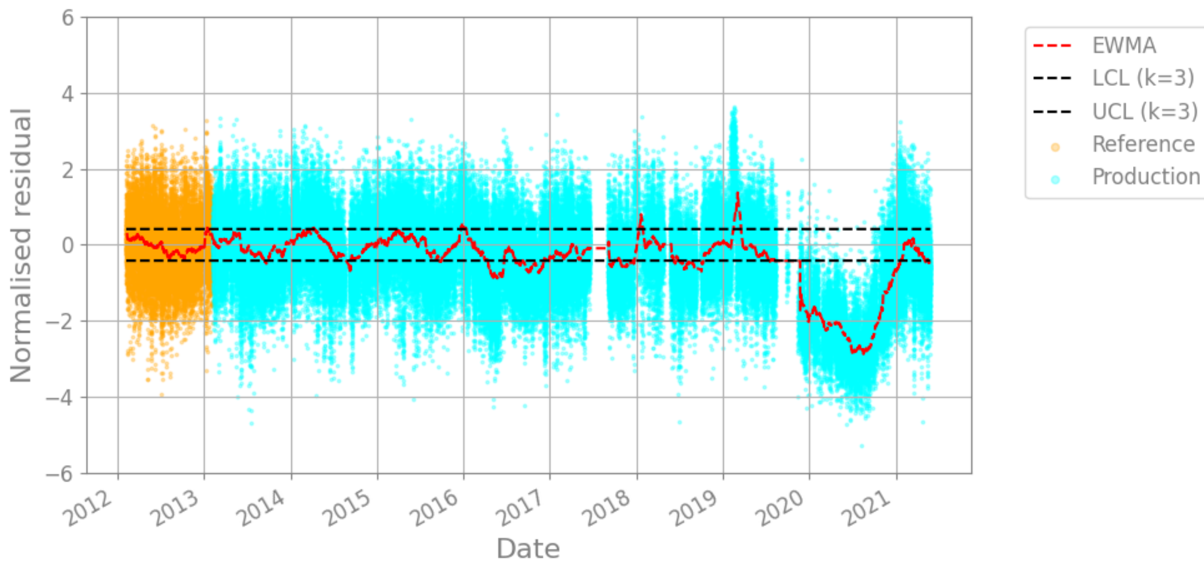

Figure 6.4: Example of power signal control chart.

To rephrase it in a more practical way; if the EWMA is crossing the upper control limit (UCL) it is said to overperform. This means that the signal (power, pitch angle or rotor speed) is on average higher than in the reference period. For crossing the lower control limit (LCL) the exact opposite is true; the signal is on average lower than in the reference period and therefore said to underperform.

6.4 Dependencies

The model has three dependencies that influence the performance:

- Variance of the residuals in the reference period
- Length of the reference period
- EWMA smoothing parameter λ

The first one is inherent to the data and the other two are adjustable. Adjusting the performance of the model comes down to a combination of smoothing the signal and setting the width of the control limits. The variance of the residuals, during the reference period, is leading for the width of the control limits. It induces the variance of the EWMA signal as outlined in equation 6.4. The smoothing parameter λ is an additional factor that influences the width of the control limits and can be used to find a balance between Type I & II errors. A more elaborate study on the effect of λ is performed in section 7.4. Finally, it is chosen in this study to take the first year of data as reference period in order to cover all seasons. This makes sure that all seasonal phenomena, that are not yet filtered out or normalised, are represented in the reference curves created which makes the model more robust. In the ideal situation all possible seasonal effects are filtered out during the data preparation phase. If that is the case one could examine the possibility to use shorter reference periods.

7 | Model verification

A verification of the model is necessary in order to proof that it is performing as it should. Moreover, it is interesting to know how well the model is performing quantitatively. Two techniques, with different assumptions, are used to generate synthetic data. Section 7.1 elaborates on the first technique using a Monte Carlo (MC) simulation with standard normally distributed random samples. The second technique using a Fourier simulation with randomised phases is explained in section 7.2. Section 7.3 elaborates on two variations of implementing an artificial degradation with the preferred Monte Carlo simulation in order to show the impact from different perspectives. A sensitivity study for the model smoothing parameter λ is performed in section 7.4. It is important to note that a verification study needs to be performed for every individual turbine since the performance of the model is dependent on the reference period. The example figures generated in this chapter are all based on one turbine.

7.1 Monte Carlo simulation

The Monte Carlo method is based on repeated random sampling in order to simulate a random process. By using this method it is assumed that the normalised residuals in the reference period are random, independent and normally distributed and can be modelled by simulating white noise. Figure 7.1 shows the steps that are taken to generate synthetic data-sets.

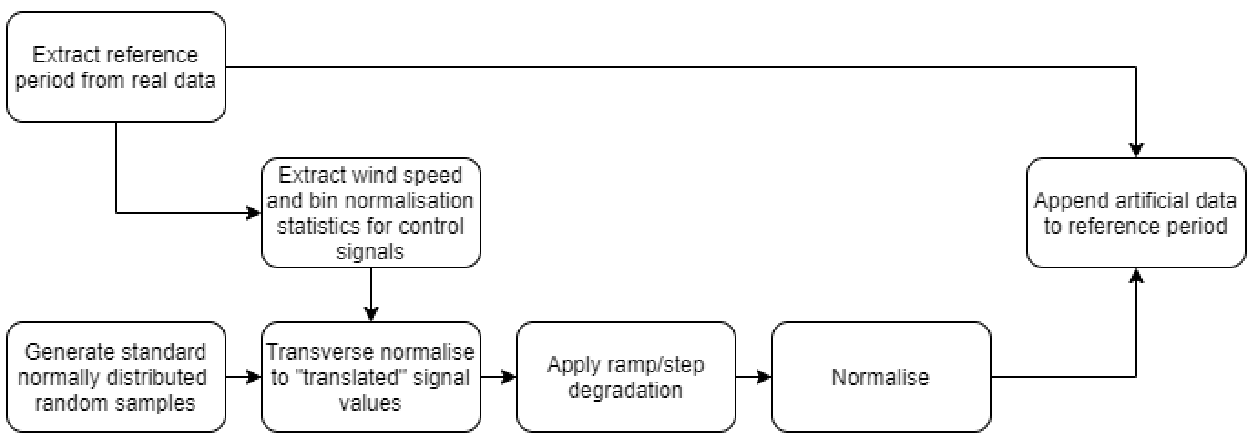

Figure 7.1: Monte Carlo simulation scheme.

Three years of data are simulated for each iteration. A degradation is applied in the second simulated year and is kept steady for the following (third) simulated year. This process is repeated 100 times to produce reliable statistics for the performance values that are tracked. The following performance values are computed:

- Average detection time (ADT)

- False negative rate (FNR)
- False positive rate (FPR)

The false positive rate is computed by using the same simulation without the introduction of an artificial degradation. As discussed in section 6.3, the false positive rate strongly depends on the smoothing factor λ that is chosen. Section 7.4 will elaborate on the discussion of choosing the right λ-value.

7.2 Fourier simulation

It could be argued that over or underperformance is dependent on the environmental conditions that have a possible time-dependency. The MC-simulation assumes uncorrelated normalised residuals (white noise) where all frequencies have equal intensities. A Fourier simulation on the other hand preserves time dependency of the data by preserving the variety in intensity per frequency. The process of generating synthetic data with the Fourier simulation is almost the same as with the Monte Carlo simulation except for the generation of the standard normal random samples. The scheme in figure 7.2 shows how the synthetic normalised residuals are generated which should be seen as a replacement for the first block in the bottom left of figure 7.1.

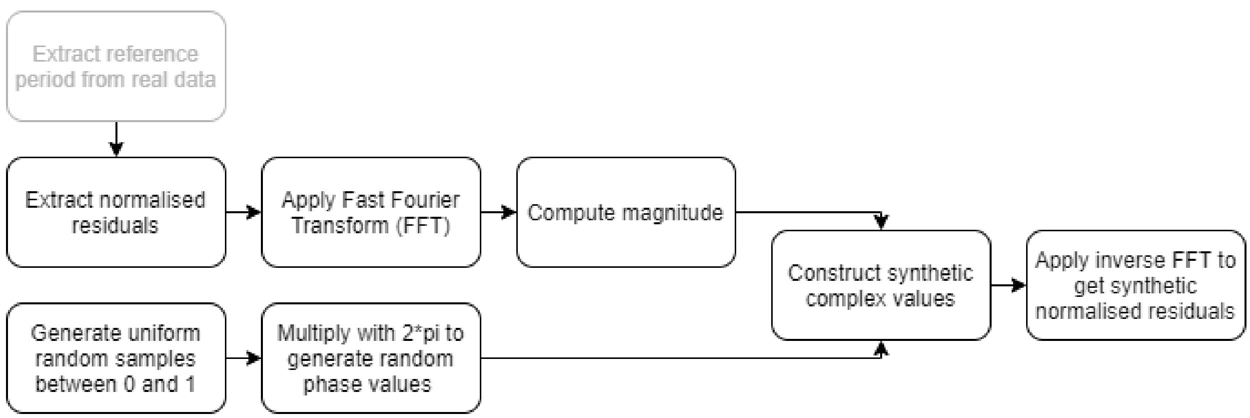

Figure 7.2: Fourier simulation scheme.

Because of the filtering process in section 5.5, the time series is not complete anymore. This means that the correlation in the signal gets lost. The maximum length of consecutive measurements in the data set, after filtering, is approximately 180 10-min measurements, i.e. 30 hours. This is too short to represent the majority of the spectrum and include low-frequency information. The loss of information is shown, in figures 7.3-7.8, by plotting the autocorrelation for an incomplete time series of the reference year and a complete time series for a sub-set in the reference year that has consecutive measurements.

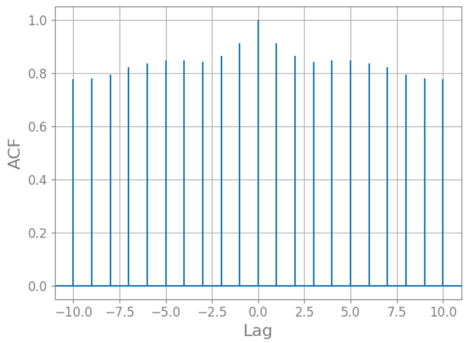

Figure 7.3: Autocorrelation plot for the power signal with a complete consecutive sub-set of the reference year.

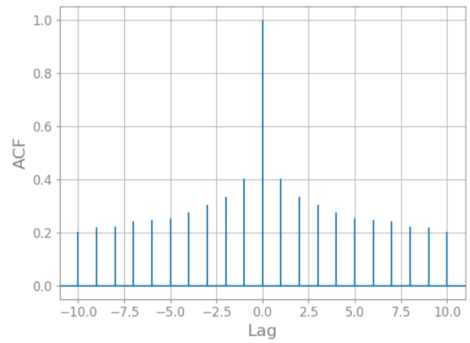

Figure 7.4: Autocorrelation plot for the power signal in the reference year (incomplete time series).

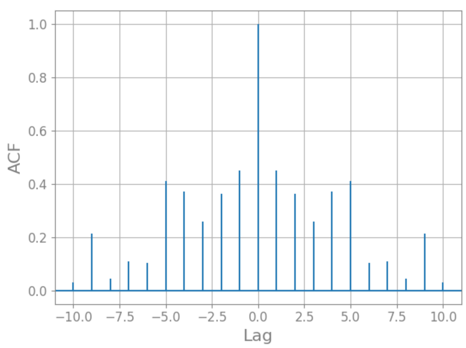

Figure 7.5: Autocorrelation plot for the pitch signal with a complete consecutive sub-set of the reference year.

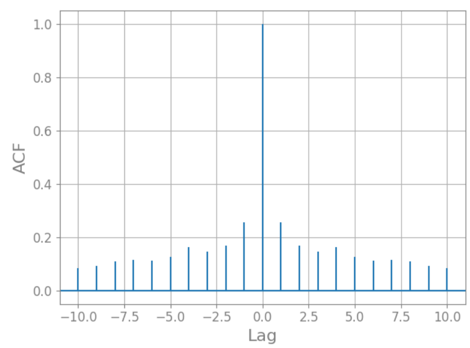

Figure 7.6: Autocorrelation plot for the pitch signal in the reference year (incomplete time series).

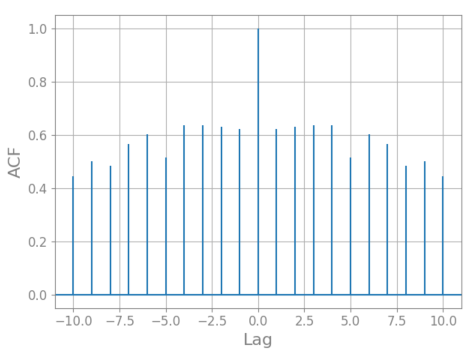

Figure 7.7: Autocorrelation plot for the rotor speed signal with a complete consecutive sub-set of the reference year.

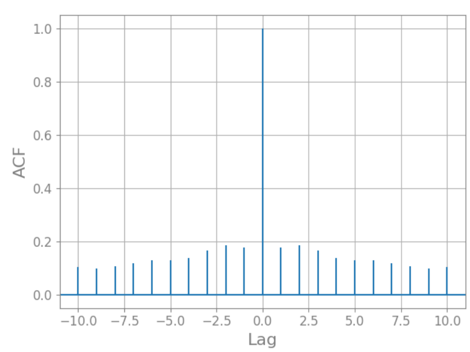

Figure 7.8: Autocorrelation plot for the rotor speed signal in the reference year (incomplete time series).

Due to the incompleteness of the data it is not possible to use a valid Fourier simulation. For this reason it is chosen to proceed with the Monte-Carlo method to simulate the normalised residuals as white noise.

7.3 Artificial degradation

Two different variations of the progression of degradation are simulated in order to show the impact from multiple perspectives. Firstly, a linear ramp degradation applied over one year and secondly a step degradation. The degradation is directly applied on the signal of interest (power, rotor speed and pitch) as shown in figure 7.1. Since the analysis is focused on the long term permanent impact of LEE, detection is defined as the signal that is out of control for a period of 1400 consecutive data entries (10-min measurements). This translates into approximately 10 days which can slightly vary because of missing or filtered measurements.

Simulated ramp degradation

Since the development of erosion is a gradual process it is believed that a ramp degradation is a good approximation of the reality. The degradation is applied linearly over one year. An example of average detection times can be seen in figure 7.9. The vertical error bars indicate the standard deviation of the simulation results. Not all signals have a value for every degradation step which is explained by a false negative rate (Type II error) of 100% which can be seen in figure 7.10. This means that these degradation values cannot be detected by the model.

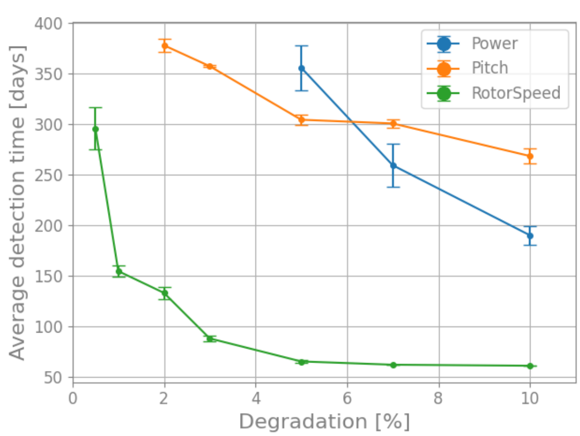

Figure 7.9: Example of average detection time for a range of simulated ramp degradation values with a MC-simulation.

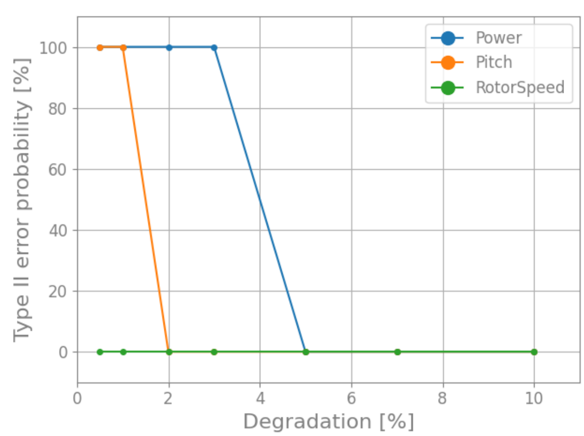

Figure 7.10: Example of false negative rates for a range of simulated ramp degradation values with a MC-simulation.

Simulated step degradation

The analysis of a step degradation is performed in order to get a better feeling for how fast the model reacts. Figure 7.11 shows the average detection times that are, not surprisingly, lower then with a ramp degradation. The false negative rates, shown in figure 7.12, are almost the same as with a ramp degradation. It is expected to converge when more iterations are performed since the smoothing factor lambda and the reference period are the same. Therefore the ramp degradation should only differ in performance of detection times and not in precision.

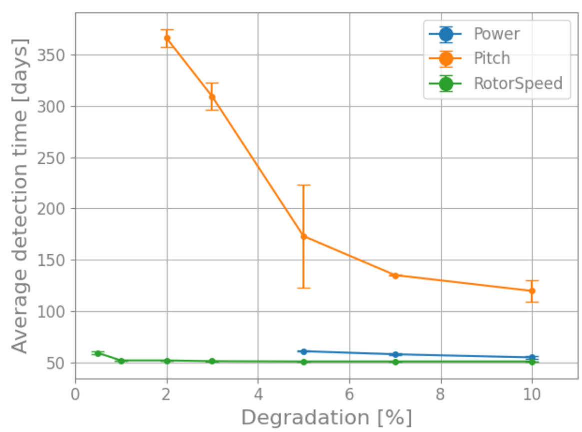

Figure 7.11: Example of average detection time for a range of simulated step degradations with a MC simulation.

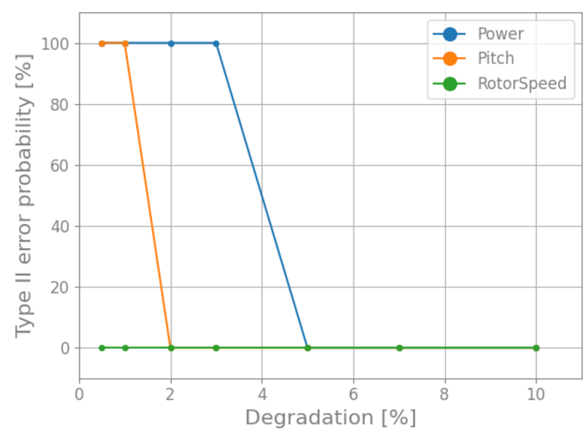

Figure 7.12: Example of false negative rates for a range of simulated step degradations with a MC simulation.

7.4 Sensitivity analysis

The smoothing parameter of the EWMA signal (λ) makes it possible to optimise the performance of the model to certain user preferences. A balance must be found between minimising the average detection time, the false positive rate (Type I errors) and the false negative rate (Type II errors). The latter is translated, in this case, to not detecting a certain degree of degradation. A sensitivity analysis is performed with different lambda values for all three signals of interest. Figure 7.13 shows the result for the power signal. The other two signals, pitch and rotor speed, showed the same trend and can be seen in appendix B. Interesting to note is that when $\lambda = 1$ the EWMA control chart reduces to a Shewhart control chart. Literature indicates that EWMA control charts perform better in detecting smaller shifts than Shewhart control charts. This difference in behaviour can clearly be seen in figure 7.13. However, as lambda is decreasing the false positive rate is increasing which is shown in figure 7.14. It is up to the end-user to decide what level of Type I errors is allowed and thereby accepting the fact that smaller degradations are not detected. It is chosen to continue with a lambda value of 0.3 for the remainder of this research as that seems to be around the change point where the false positive rate starts increasing exponentially. Moreover, for all three signals the false positive rate stays within 1%.

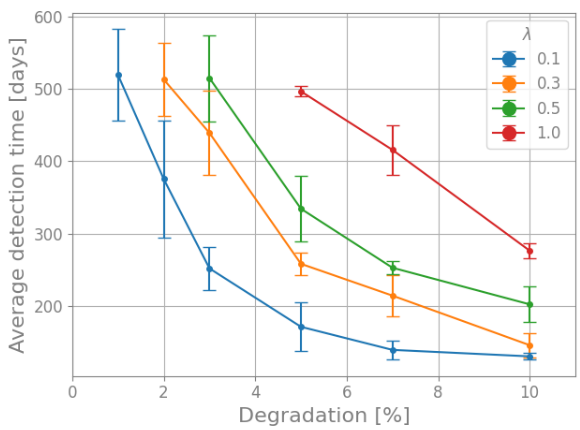

Figure 7.13: Lambda sensitivity for the power signal.

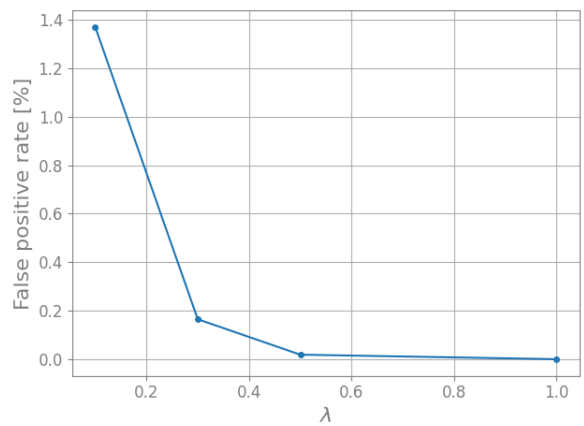

Figure 7.14: False positive rate (Type I errors) for the power signal with varying lambda values.

8 | Data evaluation

The verified model is applied on real data to evaluate the usability. Unfortunately, detailed information on the history of repair and maintenance events were not available during this research which made it impossible to validate actual detection of LEE. The code is generalised in such a way that it is possible to run an analysis for every turbine in the portfolio of Vattenfall. However, to limit the scope of this research several interesting turbines, from one park, are chosen to prevent an overload of information. The actual names of the turbines are replaced by random denotations. For internal use a mapping is given in the dedicated appendix for Vattenfall.

This specific wind farm is chosen because two inspection reports, from the end of 2017 and begin 2021, were available with clear pictures of leading edge erosion at several turbines. It is interesting to check whether, towards these inspection dates, a clear trend can be found for the turbines that suffer from LEE. A vertical blue line is drawn in the control charts to mark the dates of inspection. The model parameters were set as shown in table 8.1 based on section 6.4 and 7.4. Turbines at the edge of the farm are chosen in order to have a free wind sector.

Table 8.1: Model settings

Parameter	**Value**	**Unit**
reference period	1	[year]
EWMA smoothing factor (λ)	0.3	[–]

8.1 Turbine 1

The inspection documents of turbine 1 report several smaller damages and one substantial damage due to LEE for the inspection in 2017. Furthermore, damage and peeling of the foil is reported. Moreover, the most recent inspection in 2021 shows severe LEE. Pictures of these damages can be seen in appendix section C.1. After preparation of the data, 36.9% is left for the analysis. Figures of the preparation process, comparable to what is shown in chapter 5, can be found in C.1 .

Results

Figures 8.1-8.3 show the control charts of the power, pitch and rotor speed signal respectively.

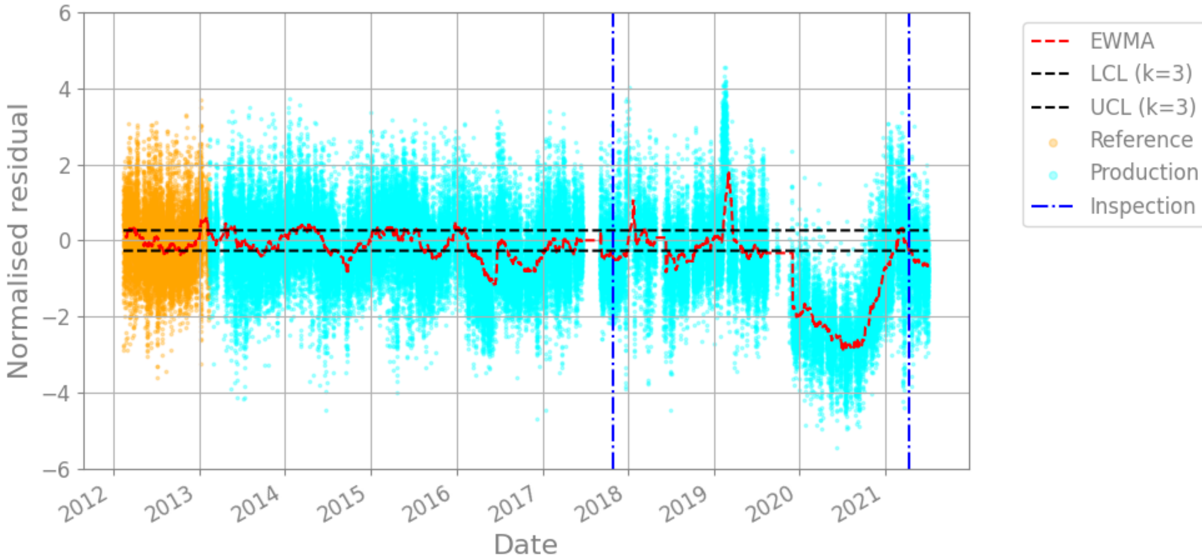

Figure 8.1: Power signal control chart, turbine 1.

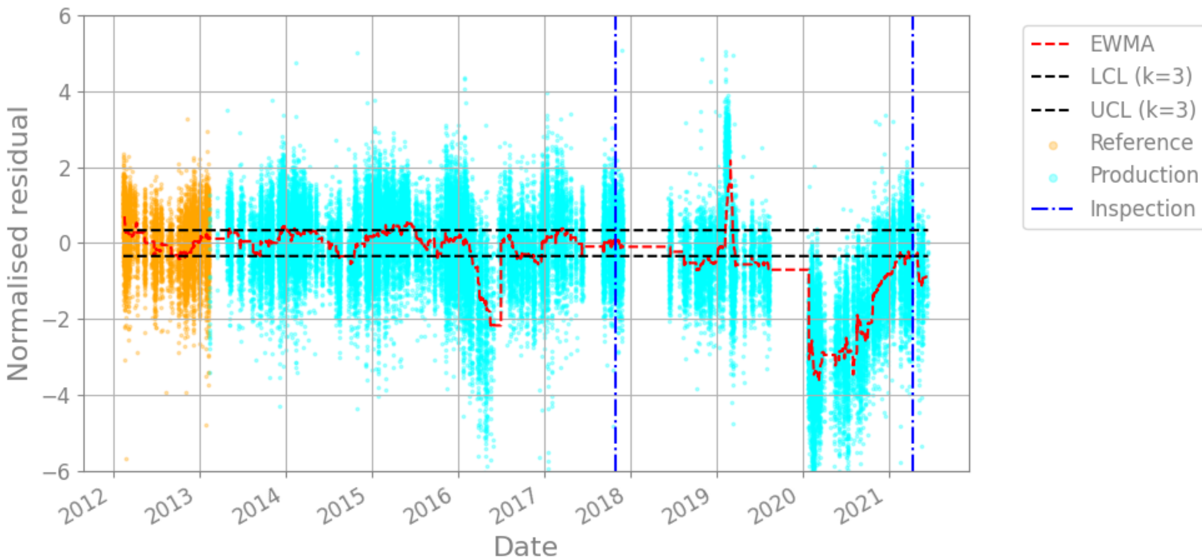

Figure 8.2: Pitch signal control chart, turbine 1.

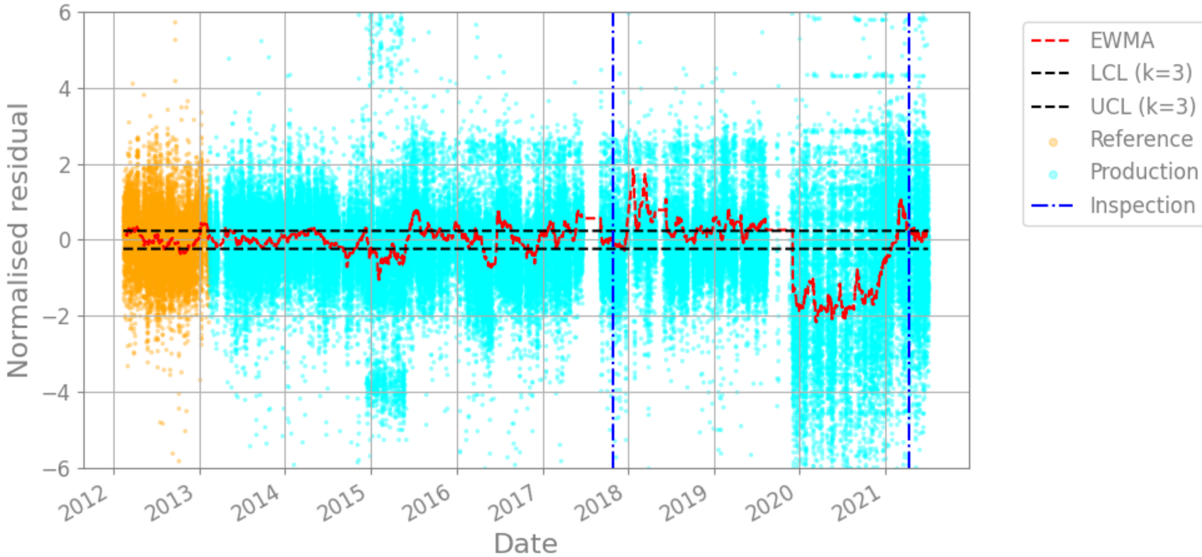

Figure 8.3: Rotor speed signal control chart, turbine 1.

Looking at the control charts one can see that despite the extensive data preparation and smoothing techniques used, the signal is still very unstable and shows a lot of variability. The power and pitch signal are clearly getting out of control at the start of 2016. The distribution made a shift downward which caused a crossing of the control limit. This means that for the same wind speeds both signals consistently show lower values. As explained in section 6.1 such an event could potentially be due to LEE. Due to the lack of access to repair and maintenance data, in this research, it is hard to explain the recovery and thereby providing evidence for LEE. More importantly, the signal does not show clear signs of degradation at the time of the inspection where there is hard evidence of LEE (see table 8.2). Furthermore, a big drop is detected at the end of 2019 which lasts until the end of 2020. It recovers slowly, against expectations, towards the second inspection where severe LEE and even voids are detected.

Table 8.2: Summary of turbine 1 control chart signal during inspections.

Signal	**Inspection**	**LEE state**	**In/Out of control**	**Monotonic behaviour (prior to inspection)**
Power	2017	Substantial	Out of control	No
	2021	Critical	Out of control	No
Pitch	2017	Substantial	In control	No
	2021	Critical	In control	No
Rotor speed	2017	Substantial	In control	No
	2021	Critical	In control	No

Verification

The Monte Carlo simulation without degradation factor generated false positive rates as shown in table 8.3. A lower lambda value could be used to realise lower average detection times. This would, however, increase the false positive rates as shown in section 7.4. It is up to the user to decide on what false positive rates are acceptable.

Table 8.3: False positive rates (Type I errors) with MC-simulation, turbine 1

Signal	Value	Unit
Power	0.0	[%]
Pitch angle	0.07	[%]
Rotor speed	0.0	[%]

The ramp degradation modelled with a Monte Carlo simulation is chosen as representative situation, as mentioned in chapter 7. Results for the step degradation are shown in appendix section C.1. Figure 8.4 shows the average detection times (ADT) for different degradation intensities. Missing ADT values for both pitch and power are explained by a 100% false negative rate shown in figure 8.5. This means that the model was not able to detect these degradation intensities.

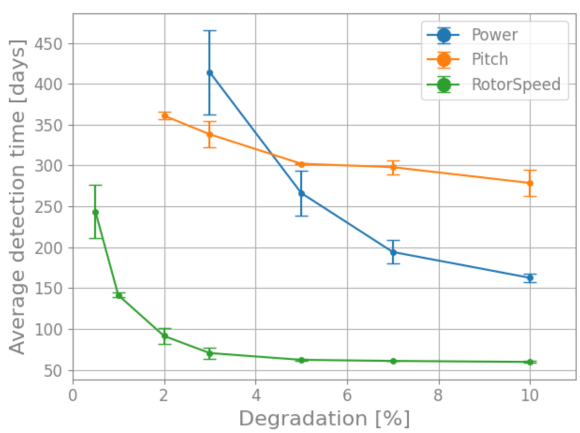

Figure 8.4: Average detection time performance of model for ramp degradation with Monte-Carlo simulation, turbine 1.

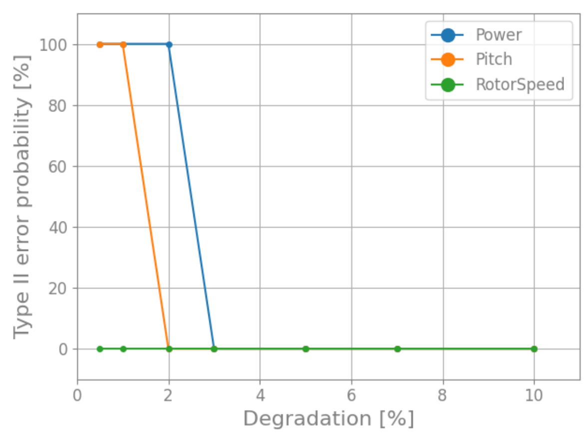

Figure 8.5: False negative rate (Type II error) performance of the model for a ramp degradation with Monte-Carlo simulation, turbine 1.

8.2 Turbine 2

Turbine 2 is located at the south-west corner of the wind farm which is the most optimal place according to the wind rose. 39.3% of the data is preserved after filtering which can be seen in figures C.22-C.30. The inspection in 2017 shows light LEE with additional peeling of the protecting leading edge foil. Severe to critical damage is detected in the inspection of 2021 where the core of one blade tip is visible.

Results

Figures 8.6-8.8 show the control charts of the power, pitch and rotor speed signal respectively.

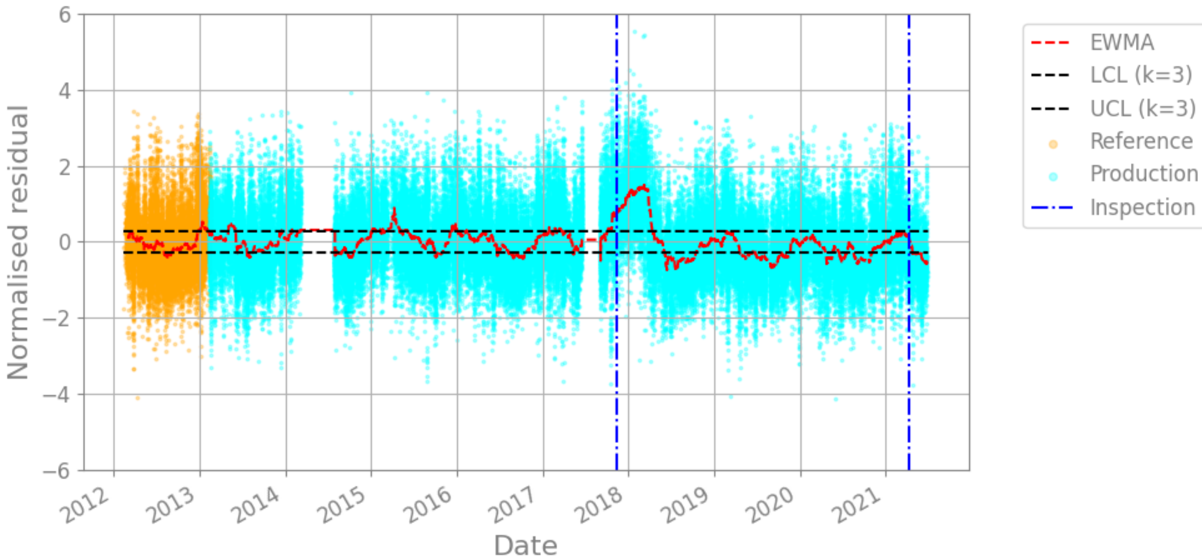

Figure 8.6: Power signal control chart, turbine 2.

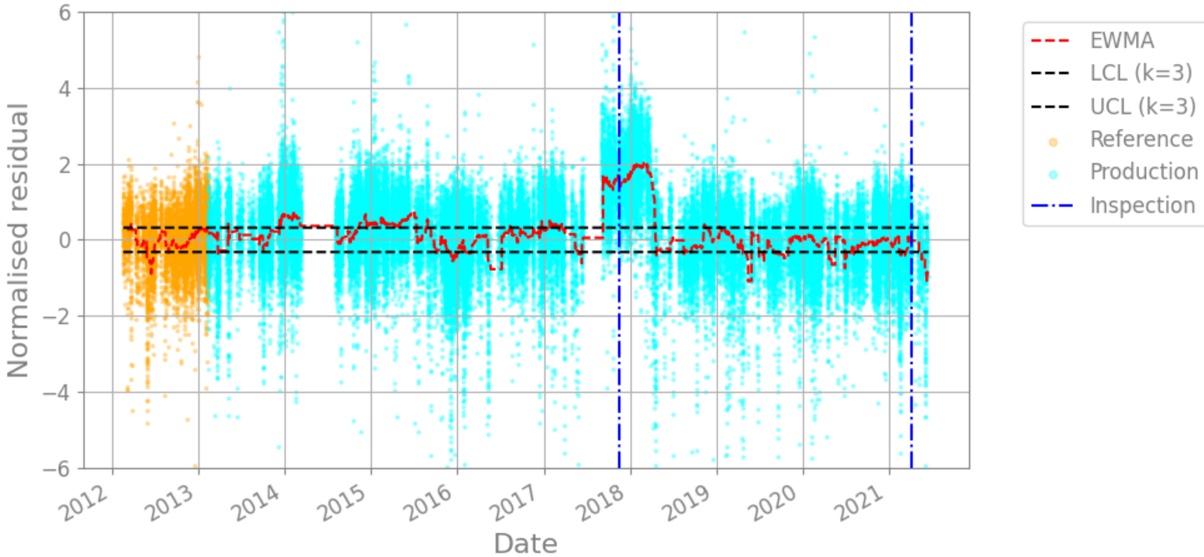

Figure 8.7: Pitch signal control chart, turbine 2.

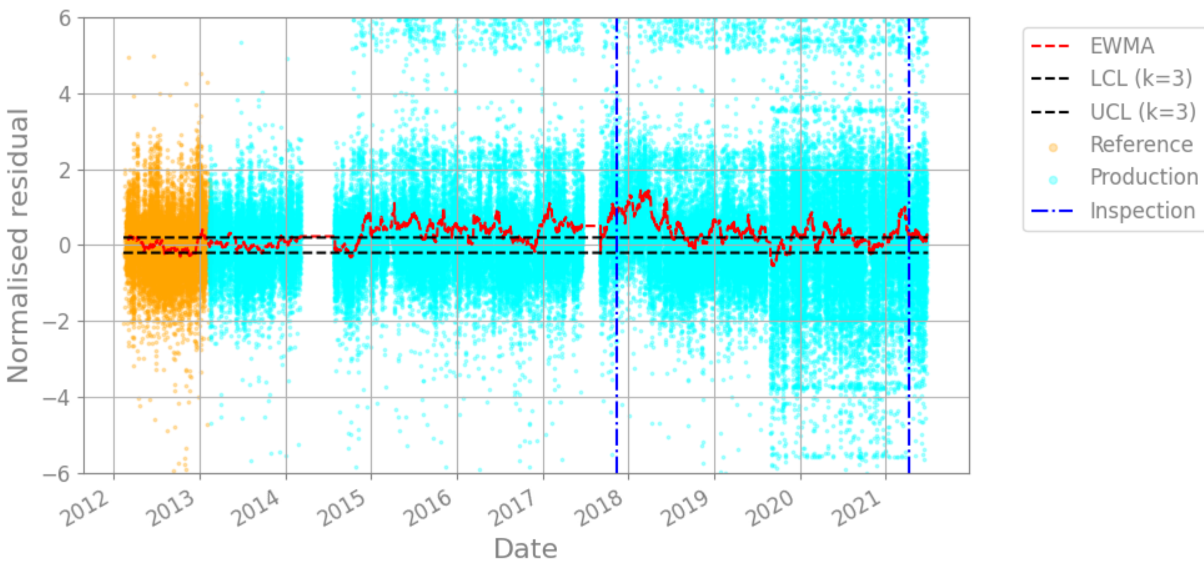

Figure 8.8: Rotor speed signal control chart, turbine 2.

During the inspection of 2017 all three signals are across the upper control limit. This means that, for the same wind speeds, those measurements are higher than during the reference period. Moreover, it is remarkable that, at the time of the inspection in May 2021 where severe and critical LEE is discovered, the power signal is in control and therefore no shift of the distribution is detected in this signal. In fact, the signal seems to be very similar to the years before. Also the pitch and rotor speed signal are not deviating substantially from other years. These findings are summarised in table 8.4.

Table 8.4: Summary of turbine 2 control chart signal during inspections.

Signal	**Inspection**	**LEE state**	**In/Out of control**	**Monotonic behaviour (prior to inspection)**
Power	2017	Light	Out of control	No
	2021	Critical	In control	No
Pitch	2017	Light	Out of control	No
	2021	Critical	In control	No
Rotor speed	2017	Light	Out of control	No
	2021	Critical	Out of control	No

Verification

The false positive rates computed with a Monte Carlo simulation are shown in table 8.5.

Table 8.5: False positive rates (Type I errors) with MC-simulation, turbine 2

Signal	**Value**	**Unit**
Power	0.5	[%]
Pitch angle	0.0004	[%]
Rotor speed	0.0006	[%]

The rotor speed signal is very stable in the reference period which is reflected in high performance of degradation detection with low average detection times up to a degradation of 0.5% (see figure 8.9). 3% degradation for the power signal still shows a low false negative rate but the average detection time suggests that it is detected due to randomness since it is much longer than a year (365 days) in which the degradation is implemented. So a reliable detection for the power is possible up to a degradation of 5% (see figure 8.10). The pitch signal shows reliable results up to 2% of degradation.

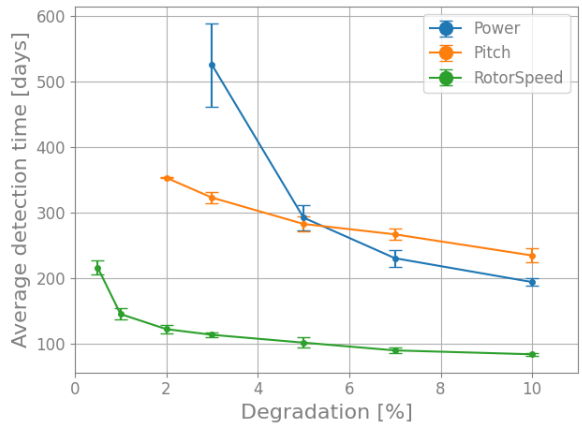

Figure 8.9: Average detection time performance of model for ramp degradation with Monte-Carlo simulation, turbine 2.

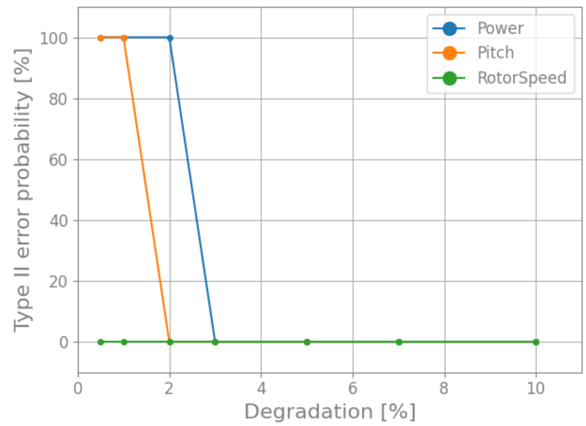

Figure 8.10: False negative rate (Type II error) performance of the model for a ramp degradation with Monte-Carlo simulation, turbine 2.

8.3 Turbine 3

Multiple substantial damages were reported for this turbine, at the inspection of end 2017, as can be seen in figure C.33-C.35. The erosion has advanced up to a state where the laminate is exposed to the environment. Leading edge erosion is also found at the inspection of April 2021 (fig. C.36-C.37). The wind rose of this farm is not in favour of this turbine what results in a smaller free wind sector. Only 23.5% of the data is left after filtering (fig. C.38). The impact of every filter is shown in figures C.39-C.46.

Results

Figures 8.11-8.13 show the control charts of the power, pitch and rotor speed signal respectively.

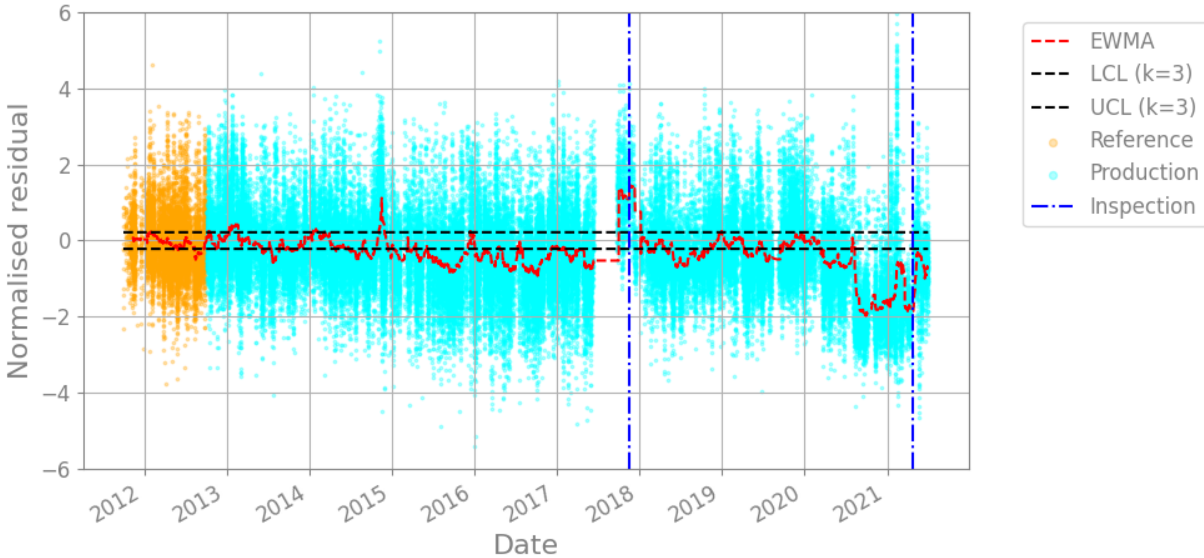

Figure 8.11: Power signal control chart, turbine 3.

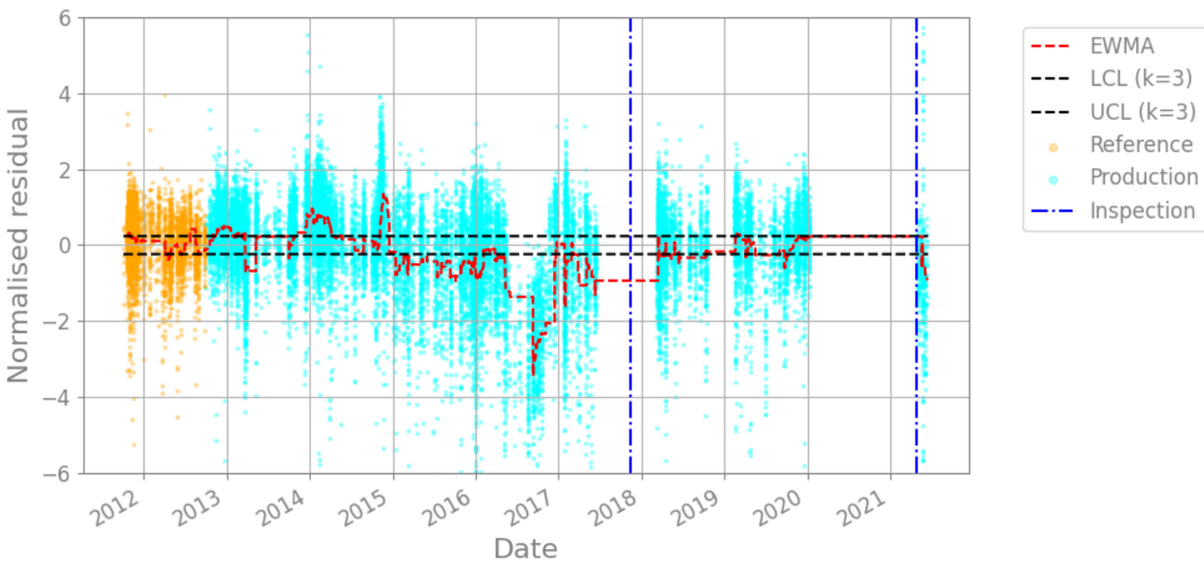

Figure 8.12: Pitch signal control chart, turbine 3.

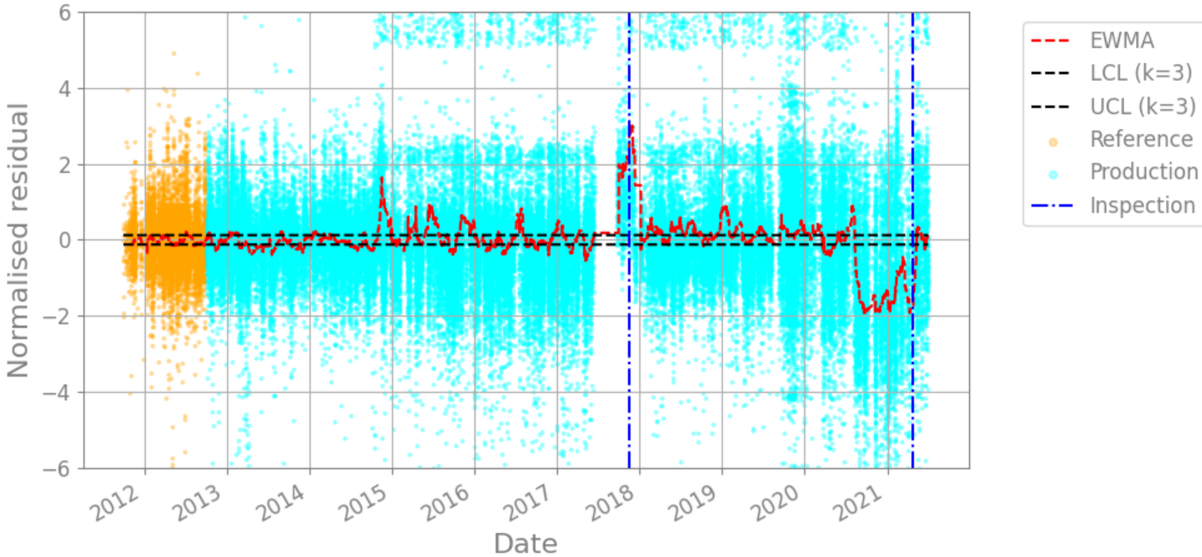

Figure 8.13: Rotor speed signal control chart, turbine 3.

At the time of the inspection in 2017 all three signals are out of control. However, the power and rotor speed signal show overperformance instead of expected underperformance. As it is a very sudden and temporary increase, it is suspected that something is wrong with the measurements or measurement devices. Furthermore, on average the power signal seems to be degraded from 2015 up till the sudden jump in 2017. Because of the variability in the signal it jumps in and out of control depending on the time of the year. During this period the pitch signal seems to be degraded on average as well, whereas the rotor speed signal stays noisy around the zero-line. A possible explanation could be that, according to Sareen et al. [2014], degradation effects are more visible at higher angles of attack (above rated area) and thus should be visible in the pitch signal earlier in this analysis according to the hypothesis of this research. From mid 2020 a large decrease in the power and rotor speed signal is visible. No measurements are available for the pitch control chart because everything is curtailed before it reaches rated power and therefore it is filtered by the model. The fact that the turbine is curtailed for a longer period suggests that there is a known problem. Interesting enough, the power and rotor speed signal seem to recover after the inspection date which is against expectations just as with turbine 1. The findings for the control charts during the inspection dates are summarised in table 8.6.

Furthermore, deviating groups of outliers are visible in the rotor speed signal from 2015 onward. After investigation this was the case at multiple turbines that are investigated. Section 8.4 elaborates on this issue.

Table 8.6: Summary of turbine 3 control chart signal during inspections.

Signal	Inspection	LEE state	In/Out of control	Monotonic behaviour (prior to inspection)
Power	2017	Severe	Out of control	No
	2021	Substantial	Out of control	No
Pitch	2017	Severe	Out of control	No
	2021	Substantial	-	-
Rotor speed	2017	Severe	Out of control	No
	2021	Substantial	Out of control	No

Verification

The simulations without degradation factor generated false positive rates as shown in table 8.7. Since lambda is kept constant during runs for different turbines it can be seen that a balance must be found between minimising both the ADT and FPR.

Table 8.7: False positive rates (Type I errors) with MC-simulation, turbine 3

Signal	Value	Unit
Power	0.2	[%]
Pitch angle	0.2	[%]
Rotor speed	0.8	[%]

Figure 8.14 and 8.15 indicate that a power degradation of more than 5% should always be detected for this turbine. One trend that can be seen is that a more stable reference period contributes on average to lower detection times.

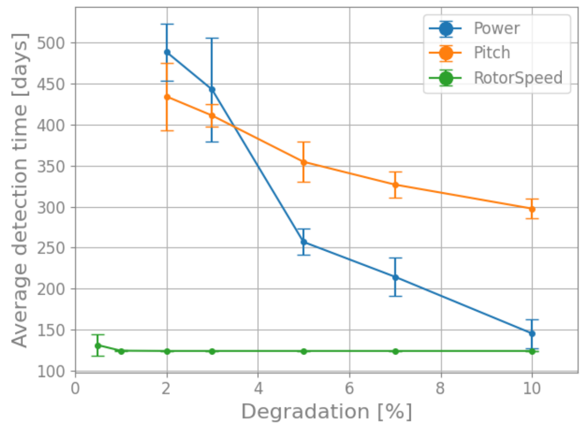

Figure 8.14: Average detection time performance of model for ramp degradation with Monte-Carlo simulation, turbine 3.

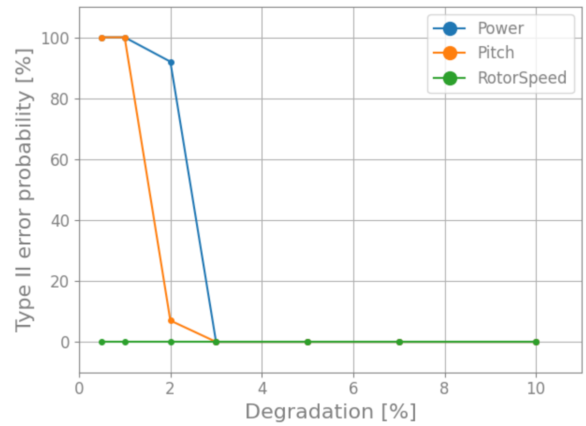

Figure 8.15: False negative rate (Type II error) performance of the model for a ramp degradation with Monte-Carlo simulation, turbine 3.

8.4 General findings

During the analysis of the rotor speed control charts, from different turbines, it was observed that consistent deviations were present from approximately 2015 onward. For example in the upper part of figure 8.13. Through investigation of individual bin control charts it became clear that most probably a software update was implemented. The lower wind speed bins, up to 6 m/s indicated a sudden step change in the mean rotor speed as can be seen in figure 8.16 and 8.17.

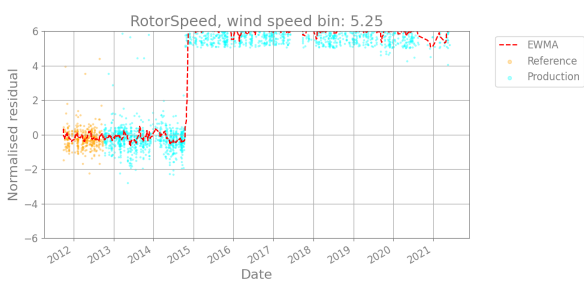

Figure 8.16: Rotor speed bin control chart, turbine 3, with sudden step change.

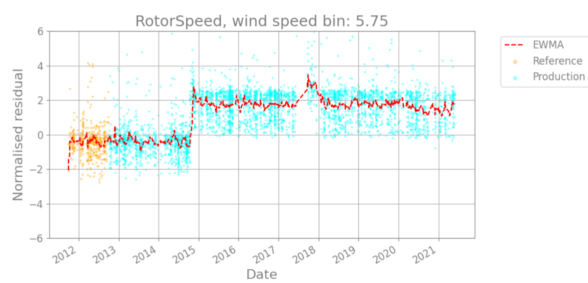

Figure 8.17: Rotor speed bin control chart, turbine 3, with sudden step change.

Such a sudden change increases the variance in the EWMA signal of the complete control chart. As described in section 6.3; control charts are useful to eliminate variability in the process by assigning causes to 'special variation'. This finding is a good example of how it can be implemented in practice. Figure 8.18 shows the control chart signal without this step change in comparison to the control chart that still includes the step change (fig. 8.19). For operational use it would be useful to correct for this new status quo instead of removing the whole period of deviating data.

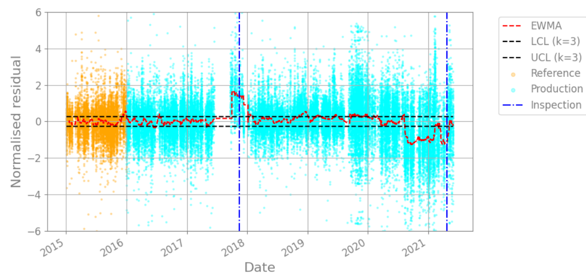

Figure 8.18: Rotor speed control chart, turbine 3, without sudden change in lower wind speed bins.

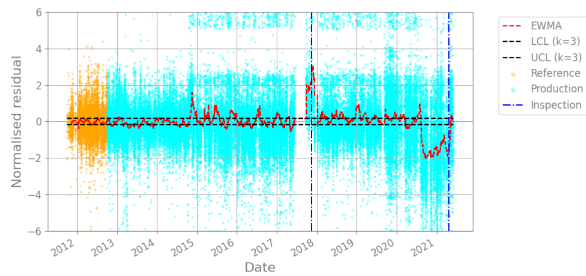

Figure 8.19: Rotor speed control chart, turbine 3, with sudden change in lower wind speed bins.

8.5 Discussion

To generalise the different out of control periods in the control charts it is suggested to divide them into three categories:

- *Short periods that are out of control*
  These are probably caused by unexplained variance in the signal that is not filtered out or corrected as discussed above.

- *Sudden step changes*
  These are probably explained by software updates or sudden errors in measurement devices.

- *Monotonic decreasing periods*
  This is the most interesting case to investigate as it is thought to be explained by degradation of sub-components of the turbine.

It is clear that significant changes of the signal distributions can be detected (out of control periods). However, without access to validation data, in the form of repair histories, it is hard to identify root cause events. Moreover, although hard evidence of LEE is present from inspection reports, no unambiguously monotonic behaviour can be detected prior to the inspection.
In some cases the EWMA is not oscillating around the zero line. Especially in the reference period where it is assumed that the turbine is still free from defects this would mean that there are still phenomena present that are not filtered out or corrected during the preparation phase. In several cases it even seems like seasonal behaviour. According to a wind resource

expert within Vattenfall it could be related to icing or low-level jets which are seasonal events.
The fact that the EWMA signal is not oscillating around the zero line means that there is still some dependency in the signal which is confirmed, for small time windows, by the autocorrelation plots in section 7.2. That also explains deviation of the turbine specific type I error rates compared to a probability belonging to three standard deviations away from the average. These probabilities are calculated by assuming independent normally distributed data. However, not satisfying this assumption in the reference period could contribute to different limits and therefore different probabilities of exceedance.

9 | Conclusion & Recommendations

Based on the results outlined in chapter 8 the research questions are answered in the conclusion which can be found in section 9.1. Furthermore, section 9.2 is devoted to recommendations that followed out of these conclusions. Specific recommendations for Vattenfall are available in a dedicated appendix.

9.1 Conclusion

The blades are a critical component of a wind turbine to convert the kinetic energy from the wind into electric energy. Due to the long exposure to harsh environmental conditions, development of leading edge erosion (LEE) on the blades is a common phenomenon nowadays. Performance degradation due to LEE is shown in previous research but detection of LEE for operational turbines is underdeveloped. This thesis investigates the possibility to develop a detection method by utilising the data that is gathered by the supervisory control and data acquisition (SCADA) system of the turbine.
Periods of underperformance are clearly detected in the lifetime data of several turbines for the three signals of interest; power output, pitch angle and rotor speed. However, due to the lack of access to validation data, these underperformances can not be linked to specific events and thereby not to leading edge erosion. In fact, the model results from several turbines, where clear pictures of (severe) LEE are available from on-site inspections, suggest that the impact of LEE is not significant compared to other unknown factors that cause deviation from normal behaviour. These outcomes question conclusions from literature where significant underperformance is said to be caused by leading edge erosion. This means that if a validation of the results in this research confirm that LEE does not have a significant impact on the performance, a repair would only be justified by concerns about the structural integrity of the blade.
Although an interesting correlation between a deviating power signal and both pitch and rotor speed signal seem to be present. It cannot be proven that both the pitch and rotor speed signal are good signatures for LEE. Monotonic behaviour of the signals is visible in some of the periods where underperformance is detected. However, for the turbines that are analysed in this research no clear monotonic period is present prior to the inspection dates where LEE is detected.
The verification of the model shows that the performance of detecting degradation is in the first place strongly dependent on the variability of the data in the reference period of the turbine. Furthermore, by tuning the model parameters a balance must be found, by the user, to minimise both the average detection time and type I errors (false positives). Average detection times are consistently low for the rotor speed signal up to the minimum degradation tested of 0.5%. For the pitch signal, reliable detection with a low false negative

rate is possible for a degradation of $2-5\%$. The power signal shows reliable detection for degradation of 5% or higher.
All these findings together show that the method is capable of detecting deviating behaviour of the power, pitch and rotor speed signal. By using control charts a distinction can be made between normal variation and special variation in the signals of interest. However, to assign causes to these periods of special variation, repair histories of turbines are needed in order to investigate the root cause. The method is calibrated on its own reference period and compares incoming data to detect relative shifts. Furthermore, the data is normalised for environmental conditions. Therefore the method can be used with arbitrary turbine types and geolocations. However, areas where the blades are affected by icing are excluded since this is not covered in the filtering process.

9.2 Recommendations

It would be interesting to investigate the behaviour of the different signals after repairs of LEE. A sudden jump could explain the impact caused by the LEE that was present before the repair. Moreover, this could be placed in perspective to other degradation in the history of the turbine to prioritise efforts to reduce underperformance. However, to explain these other deviating periods a complete history of the turbine repairs is needed. When this is available one can also try to explain the periods of underperformance that are present in the pitch and rotor speed signal. Especially the rotor speed signal would be an interesting indicator since it is very stable over time which makes it possible to detect small deviations.
Other recommendations can be made regarding the improvement of the model. It would be worthwhile to perform a variance reduction study on the impact of each filter. If a filter does not have much impact on the variance of the signal it would be possible to preserve data instead of removing it. Especially the impact of the free wind sector would be an interesting case. When it turns out that disturbed sectors do not have a large impact on the long term trend of the signal it can be considered to neglect this filter. This would make it possible to monitor turbines that are not on the boundaries of a wind farm. Furthermore, it is also mentioned in section 8.5 that additional unexplained phenomena and seasonality should be filtered out or corrected. It is known that some wind parks are equipped with ice-detection sensors or algorithms. This would be a good additional filter to add to the preparation phase of the model. More sophisticated residual models can also be explored. For example, artificial neural networks could give a more stable result since such a model is not discretised, it is able to explain variance and able to capture seasonality if present. This could potentially lead to a residual distribution with a smaller spread which makes it possible to detect smaller changes. As an alternative it is also possible to include the measured air density and turbulence intensity as a feature instead of applying a simplified correction.

References

Axel Albers, Tim Jakobi, Rolf Rohden, and Jürgen Stoltenjohannes. Influence of meteorological variables on measured wind turbine power curves. Technical report, Deutsche WindGuard Consulting GmbH, 2007. URL https://www.researchgate.net/publication/229015286.

Y. Amirat, M. E.H. Benbouzid, E. Al-Ahmar, B. Bensaker, and S. Turri. A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. *Renewable and Sustainable Energy Reviews*, 13(9):2629–2636, 12 2009. ISSN 13640321. doi: 10.1016/j.rser.2009.06.031.

Davide Astolfi, Raymond Byrne, and Francesco Castellani. Analysis of wind turbine aging through operation curves. *Energies*, 13(21), 10 2020. ISSN 19961073. doi: 10.3390/en13215623.

Luis Bartolomé and Julie Teuwen. Prospective challenges in the experimentation of the rain erosion on the leading edge of wind turbine blades. *Wind Energy*, 22(1):140–151, 1 2019. ISSN 10991824. doi: 10.1002/we.2272.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00*, pages 93–104, New York, New York, USA, 2000. ACM Press. ISBN 1581132174. doi: 10.1145/342009.335388. URL http://portal.acm.org/citation.cfm?doid=342009.335388.

Shane Butler, John Ringwood, and Frank O'Connor. Exploiting SCADA system data for wind turbine performance monitoring. In *Conference on Control and Fault-Tolerant Systems, SysTol*, pages 389–394, 2013. ISBN 9781479928552. doi: 10.1109/SysTol.2013.6693951.

P. Cambron, R. Lepvrier, C. Masson, A. Tahan, and F. Pelletier. Power curve monitoring using weighted moving average control charts. *Renewable Energy*, 94:126–135, 8 2016. ISSN 18790682. doi: 10.1016/j.renene.2016.03.031.

Peter Caselitz and Jochen Giebhardt. Rotor condition monitoring for improved operational safety of offshore wind energy converters. *Journal of Solar Energy Engineering*, 127(2): 253–261, 5 2005. ISSN 01996231. doi: 10.1115/1.1850485.

Junlei Chen, Jihui Wang, and Aiqing Ni. A review on rain erosion protection of wind turbine blades. *Journal of Coatings Technology and Research*, 16(1):15–24, 1 2019. ISSN 15470091. doi: 10.1007/s11998-018-0134-8.

Bill Chun Piu Lau, Eden Wai Man Ma, and Michael Pecht. Review of Offshore Wind Turbine Failures and Fault Prognostic Methods. In *Prognostics and System Health*

Management Conference, Beijing, China, 5 2012. IEEE. ISBN 9781457719110. doi: 10.1109/PHM.2012.6228954.

Henrique Dias Machado De Azevedo, Alex Maurício Araújo, and Nadège Bouchonneau. A review of wind turbine bearing condition monitoring: State of the art and challenges. *Renewable and Sustainable Energy Reviews*, 56:368–379, 4 2016. ISSN 18790690. doi: 10.1016/j.rser.2015.11.032.

Joshuva Arockia Dhanraj, V Sugumaran, and A Joshuva. Fault diagnostic methods for wind turbine: A review. *Journal of Engineering and Applied Sciences*, 11(7), 4 2016. ISSN 1819-6608. URL `www.arpnjournals.com`.

Y. Du, S. Zhou, X. Jing, Y. Peng, H. Wu, and N. Kwok. Damage detection techniques for wind turbine blades: A review. *Mechanical Systems and Signal Processing*, 141, 2020. ISSN 08883270. doi: 10.1016/j.ymssp.2019.106445.

Woobeom Han, Jonghwa Kim, and Bumsuk Kim. Effects of contamination and erosion at the leading edge of blade tip airfoils on the annual energy production of wind turbines. *Renewable Energy*, 115:817–823, 2018. ISSN 18790682. doi: 10.1016/j.renene.2017.09.002.

Robbie Herring, Kirsten Dyer, Ffion Martin, and Carwyn Ward. The increasing importance of leading edge erosion and a review of existing protection solutions. *Renewable and Sustainable Energy Reviews*, 115, 11 2019. ISSN 18790690. doi: 10.1016/j.rser.2019.109382.

Md Liton Hossain, Ahmed Abu-Siada, and S. M. Muyeen. Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review. *Energies*, 11(5), 2018. ISSN 19961073. doi: 10.3390/en11051309.

M. H. Keegan, D. H. Nash, and M. M. Stack. On erosion issues associated with the leading edge of wind turbine blades. *Journal of Physics D: Applied Physics*, 46(38), 9 2013. ISSN 00223727. doi: 10.1088/0022-3727/46/38/383001.

Mark Hugh Keegan. *Wind Turbine Blade Leading Edge Erosion: An investigation of rain droplet and hailstone impact induced damage mechanisms*. PhD thesis, University of Strathclyde, Glasgow, 2014.

Ivan Komusanac, Guy Brindley, and Daniel Fraile. Wind energy in Europe in 2019: Trends and statistics. Technical report, Wind Europe, 2020. URL `https://windeurope.org/about-wind/statistics/european/wind-energy-in-europe-in-2019/`.

Joyce Lee and Feng Zhao. Global Wind Report 2019. Technical report, Global Wind Energy Council, 2020. URL `https://gwec.net/global-wind-report-2019/`.

Mareike Leimeister and Athanasios Kolios. A review of reliability-based methods for risk analysis and their application in the offshore wind industry. *Renewable and Sustainable Energy Reviews*, 91:1065–1076, 8 2018. ISSN 18790690. doi: 10.1016/j.rser.2018.04.004.

Frederick Letson, Rebecca J. Barthelmie, and Sara C. Pryor. Radar-derived precipitation climatology for wind turbine blade leading edge erosion. *Wind Energy Science*, 5(1): 331–347, 3 2020. ISSN 23667451. doi: 10.5194/wes-5-331-2020.

Dongsheng Li, Siu Chun M. Ho, Gangbing Song, Liang Ren, and Hongnan Li. A review of damage detection methods for wind turbine blades. *Smart Materials and Structures*, 24 (3), 3 2015. ISSN 1361665X. doi: 10.1088/0964-1726/24/3/033001.

Casey Martinez, Festus Asare Yeboah, Scott Herford, Matt Brzezinski, Viswanath Puttagunta, Asare Yeboah, Casey R Martinez, Festus A Yeboah, and James S Herford. Predicting Wind Turbine Blade Erosion using Machine Learning. *SMU Data Science Review*, 2(2):17, 2019. URL `https://scholar.smu.edu/datasciencereviewhttp://digitalrepository.smu.edu.Availableat:https://scholar.smu.edu/datasciencereview/vol2/iss2/17`.

Douglas C. Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, 2020, 8 edition, 2020.

Jannie S. Nielsen, Dmitri Tcherniak, and Martin D. Ulriksen. A case study on risk-based maintenance of wind turbine blades with structural health monitoring. *Structure and Infrastructure Engineering*, 2020. ISSN 17448980. doi: 10.1080/15732479.2020.1743326.

Jannie Sønderkær Nielsen and John Dalsgaard Sørensen. Computational framework for risk-based planning of inspections, maintenance and condition monitoring using discrete Bayesian networks. *Structure and Infrastructure Engineering*, 14(8):1082–1094, 8 2018. ISSN 17448980. doi: 10.1080/15732479.2017.1387155.

Jesús María Pinar Pérez, Fausto Pedro García Márquez, Andrew Tobias, and Mayorkinos Papaelias. Wind turbine reliability analysis. *Renewable and Sustainable Energy Reviews*, 23:463–472, 2013. ISSN 13640321. doi: 10.1016/j.rser.2013.03.018.

S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 42 (1):97–101, 2000. ISSN 15372723. doi: 10.1080/00401706.2000.10485986.

D. P. Rommel, D. Di Maio, and T. Tinga. Calculating wind turbine component loads for improved life prediction. *Renewable Energy*, 146:223–241, 2 2020. ISSN 18790682. doi: 10.1016/j.renene.2019.06.131.

Agrim Sareen, Chinmay A. Sapre, and Michael S. Selig. Effects of leading edge erosion on wind turbine blade performance. *Wind Energy*, 17(10):1531–1542, 2014. ISSN 10991824. doi: 10.1002/we.1649.

Matthias Schramm, Hamid Rahimi, Bernhard Stoevesandt, and Kim Tangager. The influence of eroded blades on wind turbine performance using numerical simulations. *Energies*, 10 (9), 9 2017. ISSN 19961073. doi: 10.3390/en10091420.

Xiao Sheng Si, Wenbin Wang, Chang Hua Hu, and Dong Hua Zhou. Remaining useful life estimation - A review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1):1–14, 8 2011. ISSN 03772217. doi: 10.1016/j.ejor.2010.11.018.

H. M. Slot, E. R.M. Gelinck, C. Rentrop, and E. Van der Heide. Leading edge erosion of coated wind turbine blades: Review of coating life models. *Renewable Energy*, 80:837–848, 8 2015. ISSN 18790682. doi: 10.1016/j.renene.2015.02.036.

Wenxian Yang, Peter J. Tavner, Christopher J. Crabtree, Y. Feng, and Y. Qiu. Wind turbine condition monitoring: Technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2014. ISSN 10991824. doi: 10.1002/we.1508.

A. Zaher, S. D.J. McArthur, D. G. Infield, and Y. Patel. Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy*, 12(6):574–593, 9 2009. ISSN 10954244. doi: 10.1002/we.319.

Shizhong Zhang, Kim Dam-Johansen, Sten Nørkjær, Pablo L. Bernad, and Søren Kiil. Erosion of wind turbine blade coatings - Design and analysis of jet-based laboratory equipment for performance evaluation. *Progress in Organic Coatings*, 78:103–115, 1 2015. ISSN 03009440. doi: 10.1016/j.porgcoat.2014.09.016.

A | Programming solutions

This appendix is an addition to the chapters written in the report for the reader who is interested into the more programmatic challenges that arose with the methods used. Two big challenges that were present during the whole process were keeping the code/analysis; generalised, to be able to use it for different turbines and parks, and also lean and fast since handling the data for the lifetime of a turbine becomes very slow if one does not pay attention.

A.1 Affected time intervals

In order to facilitate the flagging of deviating events such as turbine alarms and curtailment, a time range must be converted to affected 10-minute intervals. A fast solution, by converting all events at once, is presented instead of using slow for-loops to check every measurement.

```python

def affected_intervals(config, df):
    """
    Transform time range of database events (alarms, curtailments, etc.) into affected intervals which match
    with the SCADA data requested.

    Args:
        config (.yaml): Config file with all project specific information.
        df (DataFrame): Dataframe containing from_utc and to_utc timestamps.

    Returns:
        aff_intervals (Series): Series with affected rounded timestamps according to given scada_resolution given in
        data_config.yaml.
    """
    # convert scada resolution (seconds) to nanoseconds
    interval_resolution = config["scada_resolution"] * 1_000_000_000

    # calculate how many intervals (scada resolution) fit into time range of event
    from_utc_rounded = ((df["from_utc"].astype(np.int64) // interval_resolution) * interval_resolution).to_numpy(
        dtype=np.int64
    ).reshape(-1, 1)
    to_utc_rounded = ((df["to_utc"].astype(np.int64) // interval_resolution) * interval_resolution).to_numpy(
        dtype=np.int64
    ).reshape(-1, 1)
    nr_intervals = (((to_utc_rounded - from_utc_rounded) / interval_resolution) + 1).astype(np.uint32)

```

```
    # create affected intervals for each time range (event)
    interval_arr = np.empty((nr_intervals.sum(), 1), dtype=np.int64)
    idx_arr = np.append([0], nr_intervals.cumsum())
    for i in range(len(from_utc_rounded)):
        intervals = np.arange(
            from_utc_rounded[i, 0],
            to_utc_rounded[i, 0] + interval_resolution,
            interval_resolution
        ).reshape(-1, 1)

        interval_arr[idx_arr[i]:idx_arr[i + 1]] = intervals

    interval_arr_unique = np.unique(interval_arr)
    aff_intervals = pd.to_datetime(pd.Series(interval_arr_unique.flatten()
    , dtype=np.int64))

    return aff_intervals
```

A.2 Turbulence correction

The IEC 61400-12-1 standard, Annex M, describes how to correct measurements for Turbulence Intensity (TI). In this process the wind speed distribution within a 10-minute measurement is integrated to compute the so called 'simulated power'. In order to facilitate this operation for a large data set (lifetime of a turbine) a Monte Carlo method is used.

1. 1000 random uniform distributed samples are created for each measurement.
2. These samples are converted with the percent point function (PPF), also known as the inverse of the cumulative distribution function (CDF), from the normal distribution to get normally distributed samples.
3. The wind speed mean values are multiplied with the turbulence intensity to compute the standard deviation which is then multiplied with the normally distributed samples and added to the wind speed mean. Now 1000 normally distributed wind speed samples are obtained for each measurement.
4. Then the power is computed for all simulated wind speed samples where after the average is taken over these 1000 samples per measurement to get the simulated power.

With this method it is assumed that the wind speed will be Gaussian distributed within a 10 minute measurement window.

```

def simulated_power(df, max_cp, rated_power, rated_speed, turbine):
    """
    Compute the simulated power for given inputs.
```

```
    Args:
        df (DataFrame): DataFrame with 10-min aggregated measurements.
        max_cp (float): Maximum pressure coefficient.
        rated_power (float): Rated power of turbine.
        rated_speed (float): Rated wind speed of turbine.
        turbine (object): Turbine object with information about turbine.

    Returns:
        power_sim (array): Array with simulated power for each measurement.
    """
    rho_ref = 1.225
    rotor_area = (1 / 4) * np.pi * (turbine.rotor_diameter**2)
    nr_samples = 1000

    # create windspeed samples with Monte Carlo sampling
    samples = np.random.rand(nr_samples, len(df))
    ppf_samples = norm.ppf(samples)
    windspeed_samples = (
            df["WindSpeed", "mean"].to_numpy().reshape(-1, 1) * df["TI", "mean"].to_numpy().reshape(-1, 1)
    ).T * ppf_samples + df["WindSpeed", "mean"].to_numpy().reshape(1, -1)

    # multiply with power
    power_zt = np.zeros((nr_samples, len(df)))
    rated_bool = windspeed_samples < rated_speed
    power_zt = power_zt + rated_bool.astype(np.int) * (0.5 * rho_ref * max_cp * rotor_area * (windspeed_samples ** 3))
    power_zt = power_zt + (~rated_bool).astype(np.int) * rated_power

    # sum and divide by number of samples
    power_sim = (power_zt.sum(axis=0) / nr_samples).T

    return power_sim
```

B | Sensitivity analysis

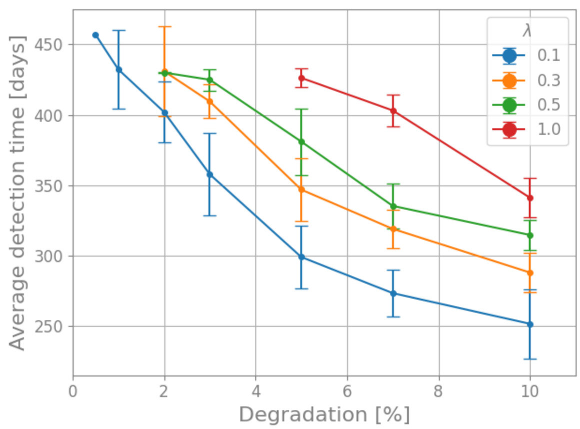

Figure B.1: Lambda sensitivity for the pitch signal.

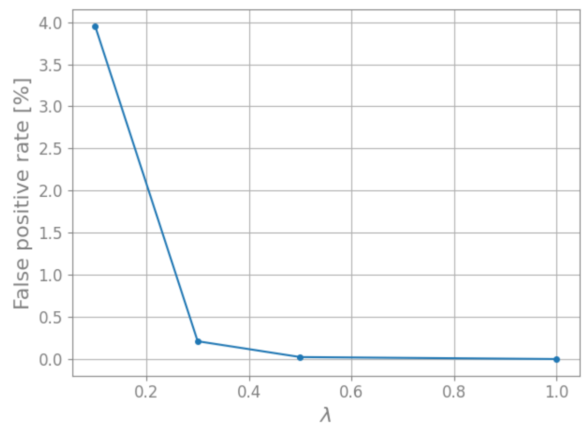

Figure B.2: False positive rate (Type I errors) for the pitch signal with varying lambda values.

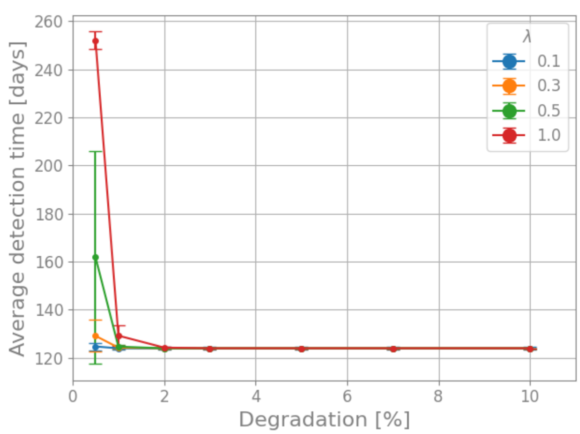

Figure B.3: Lambda sensitivity for the rotor speed signal.

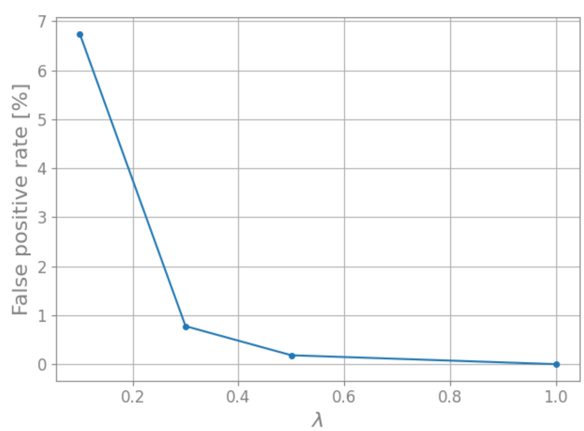

Figure B.4: False positive rate (Type I errors) for the rotor speed signal with varying lambda values.

C | Data evaluation

C.1 Turbine 1

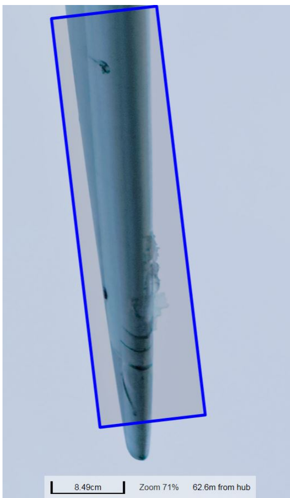

Figure C.1: Inspection October 2017, turbine 1, blade 1.

Figure C.2: Inspection October 2017, turbine 1, blade 2.

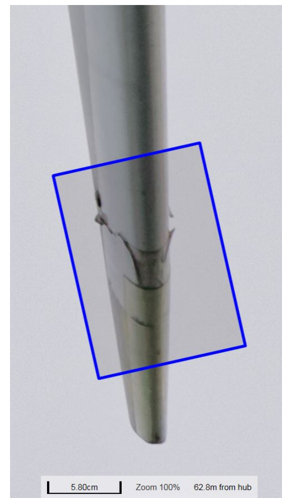

Figure C.3: Inspection October 2017, turbine 1, blade 3.

Figure C.4: Inspection April 2021, turbine 1, blade 2.

Figure C.5: Inspection April 2021, turbine 1, blade 3.

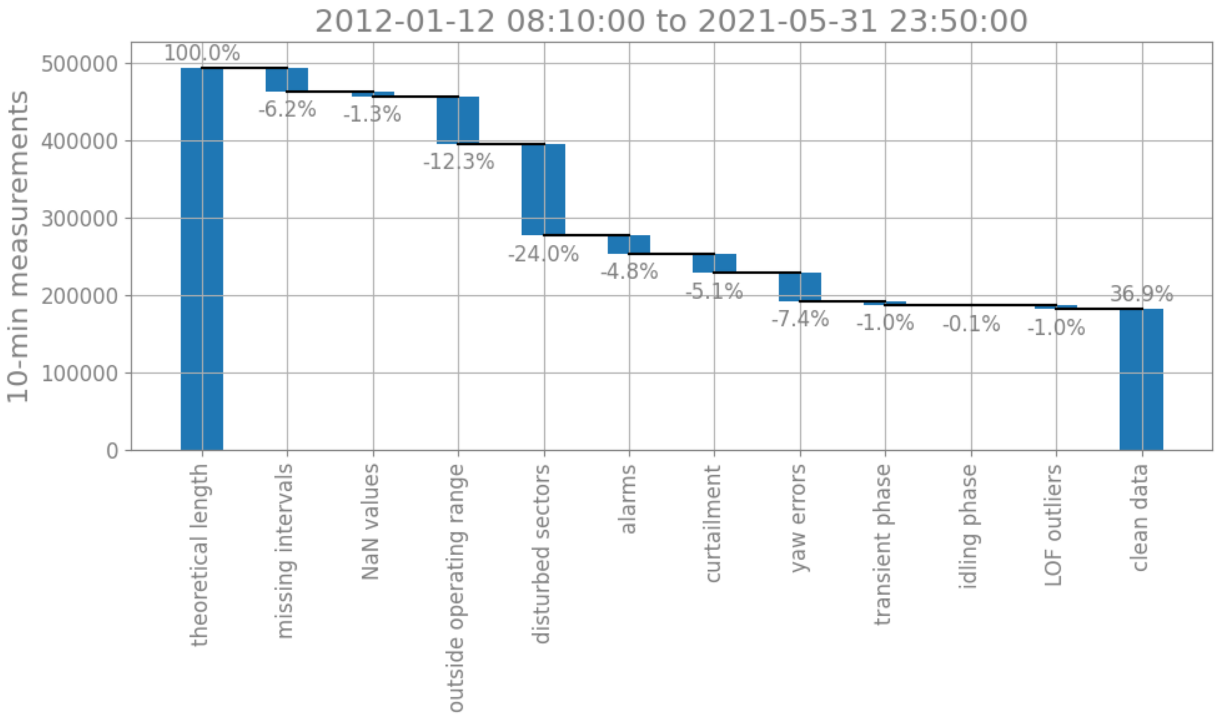

Figure C.6: Overview of filtering statistics, turbine 1.

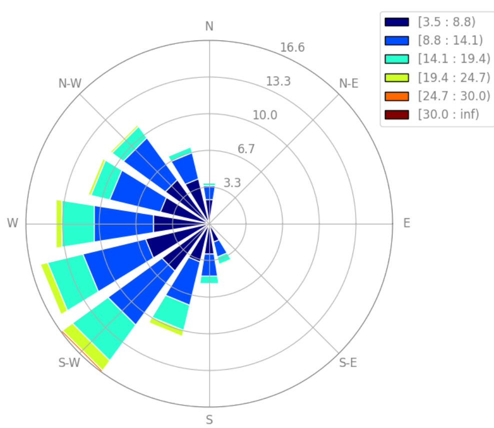

Figure C.7: Free wind sector for turbine 1.

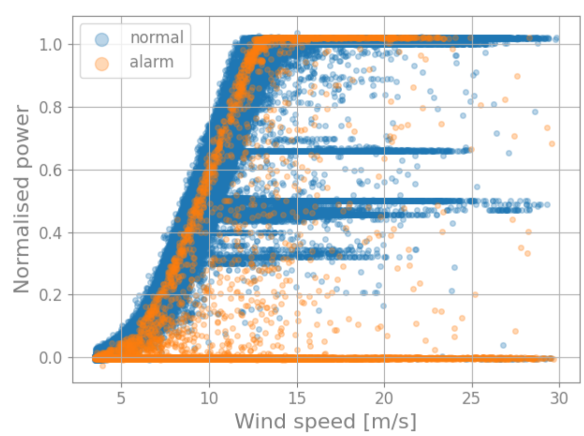

Figure C.8: Flagged alarms for turbine 1.

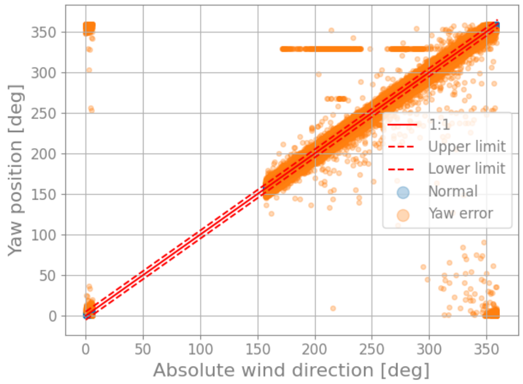

Figure C.9: Flagged yaw errors for turbine 1.

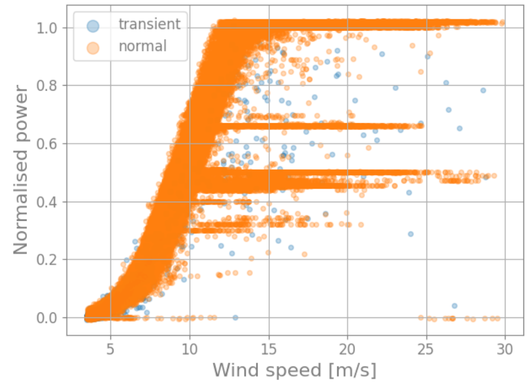

Figure C.10: Flagged transient phase measurements for turbine 1.

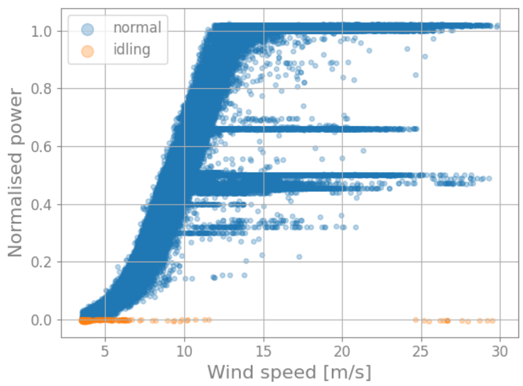

Figure C.11: Flagged idling phase measurements for turbine 1.

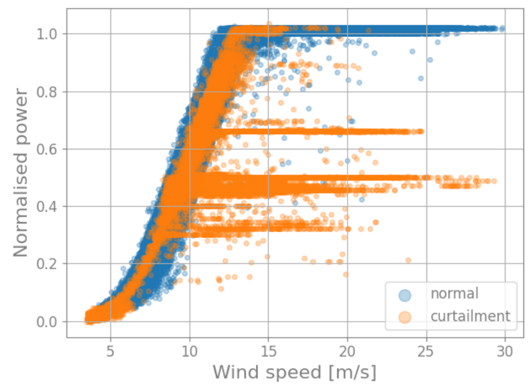

Figure C.12: Flagged curtailment for turbine 1.

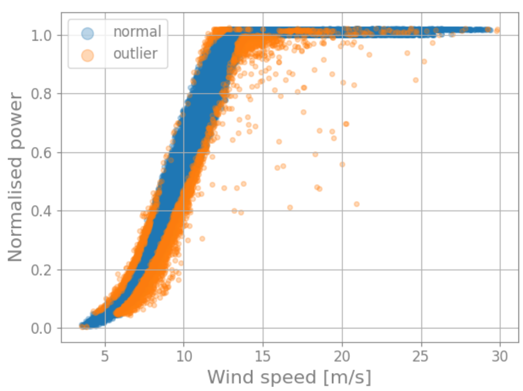

Figure C.13: Flagged outliers for turbine 1.

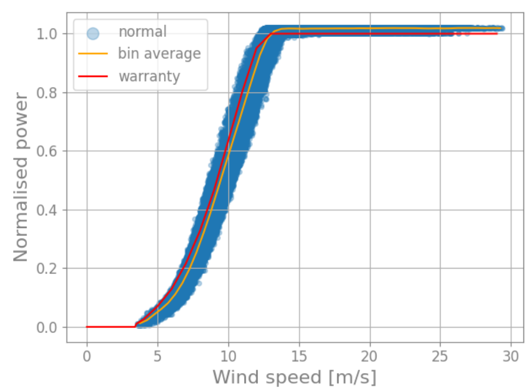

Figure C.14: Final clean power curve for turbine 1.

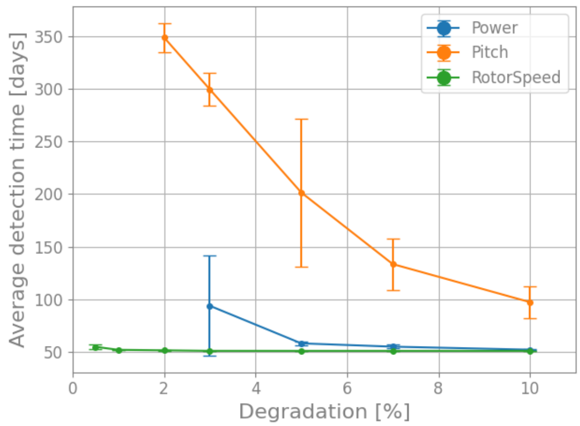

Figure C.15: Average detection time performance of model for step degradation with Monte-Carlo simulation.

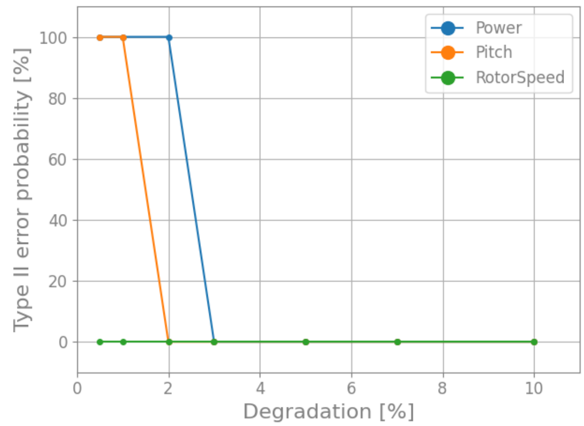

Figure C.16: False negative rate (Type II error) performance of the model for a step degradation with Monte-Carlo simulation.

C.2 Turbine 2

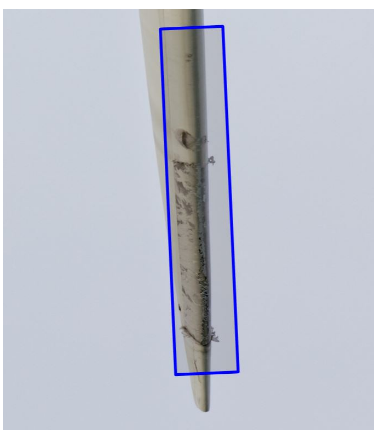

Figure C.17: Inspection November 2017, turbine 2, blade 1.

Figure C.18: Inspection November 2017, turbine 2, blade 2.

Figure C.19: Inspection November 2017, turbine 2, blade 3.

Figure C.20: Inspection April 2021, turbine 2, blade 2.

Figure C.21: Inspection April 2021, turbine 2, blade 3.

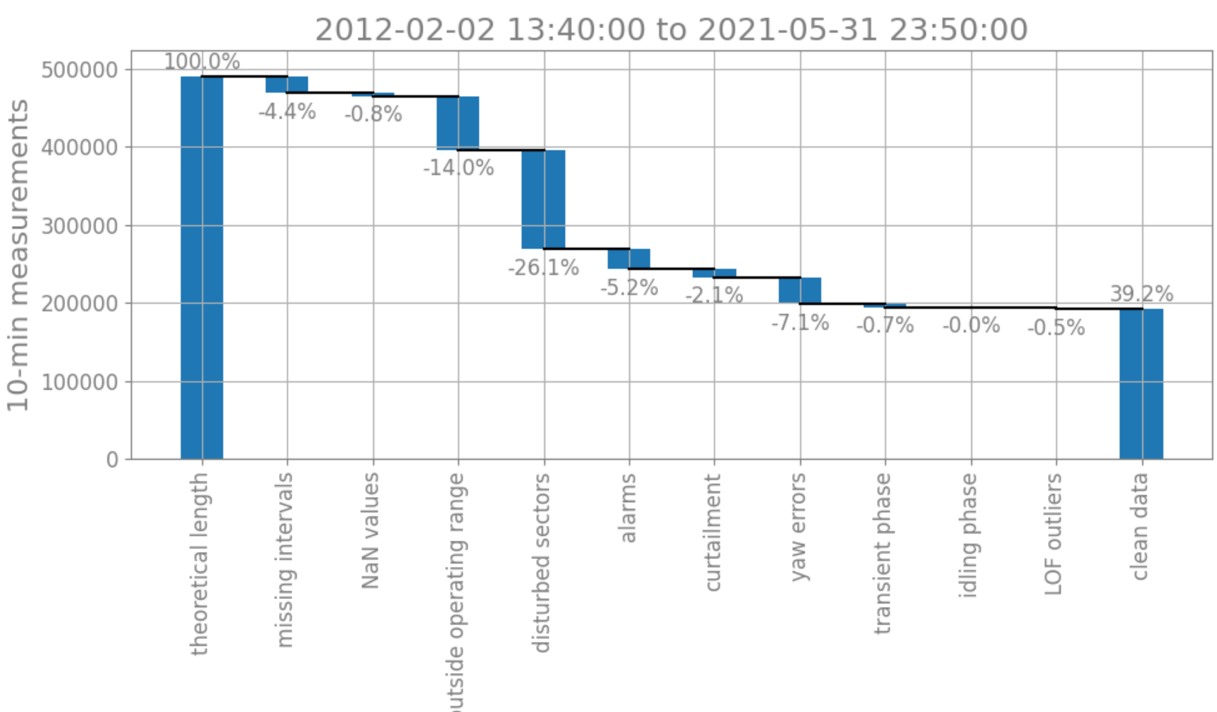

Figure C.22: Overview of filtering statistics, turbine 2.

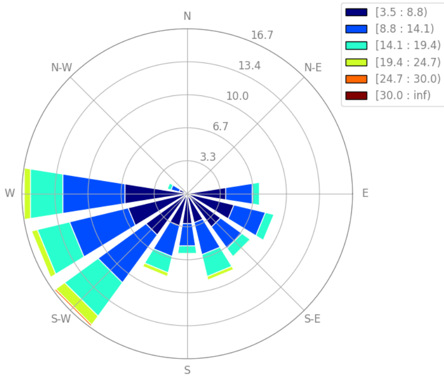

Figure C.23: Free wind sector for turbine 2.

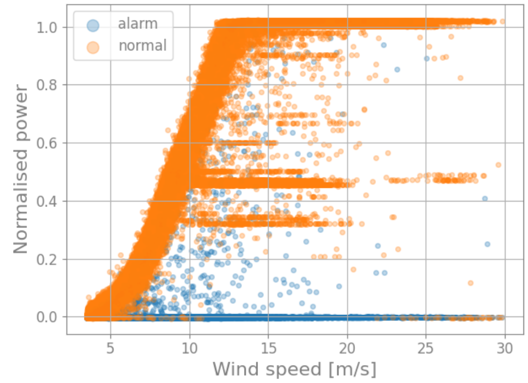

Figure C.24: Flagged alarms for turbine 2.

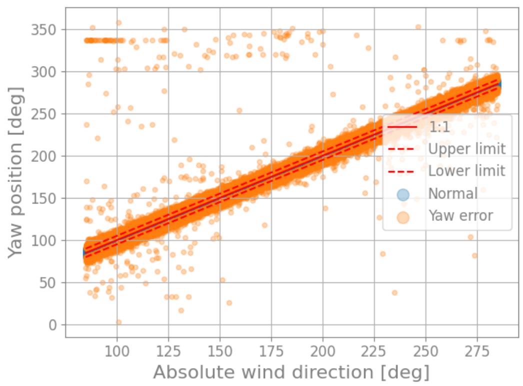

Figure C.25: Flagged yaw errors for turbine 2.

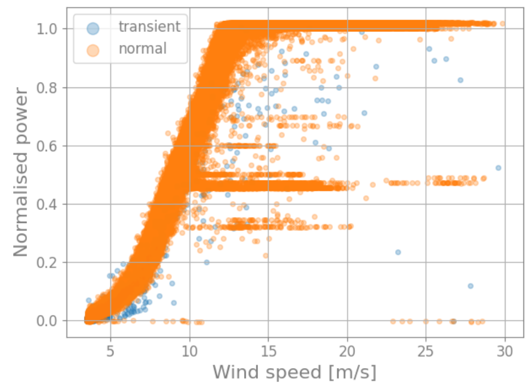

Figure C.26: Flagged transient phase measurements for turbine 2.

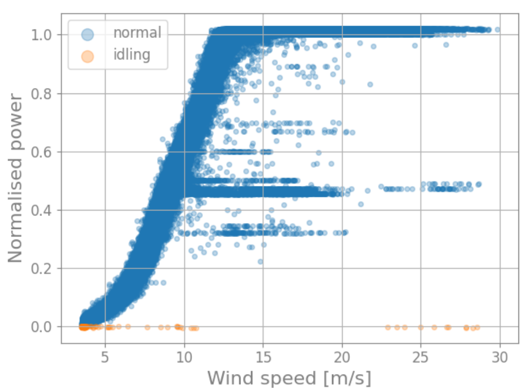

Figure C.27: Flagged idling phase measurements for turbine 2.

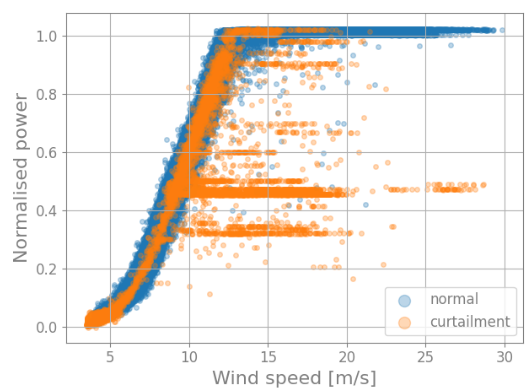

Figure C.28: Flagged curtailment for turbine 2.

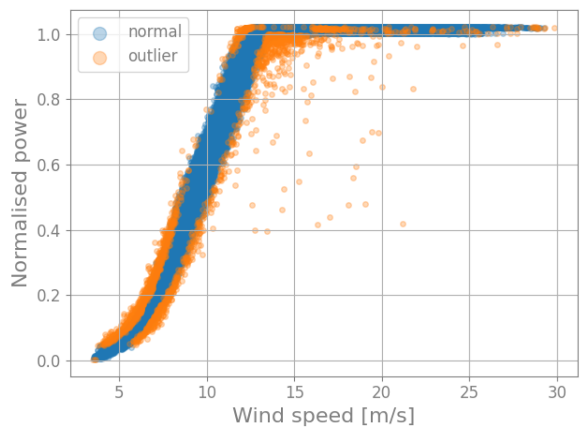

Figure C.29: Flagged outliers for turbine 2.

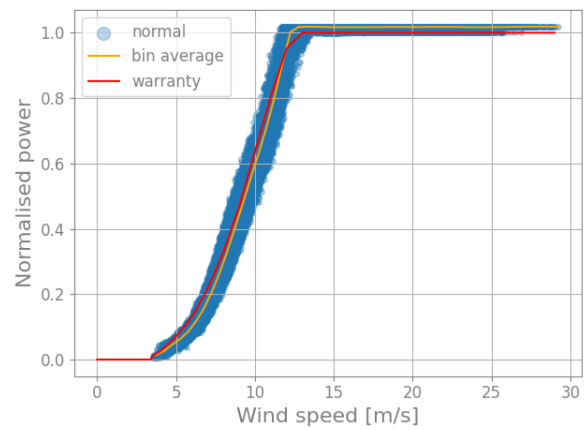

Figure C.30: Final clean power curve for turbine 2.

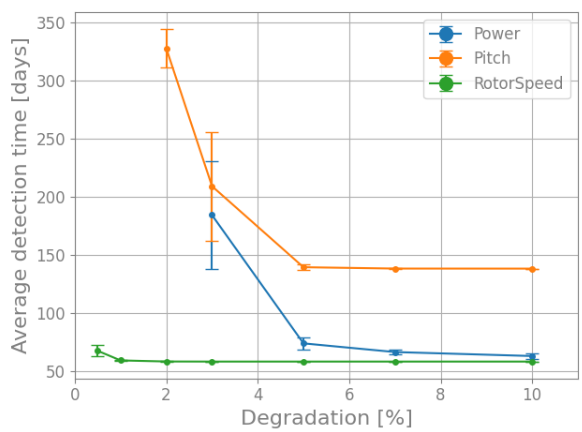

Figure C.31: Average detection time performance of model for step degradation with Monte-Carlo simulation.

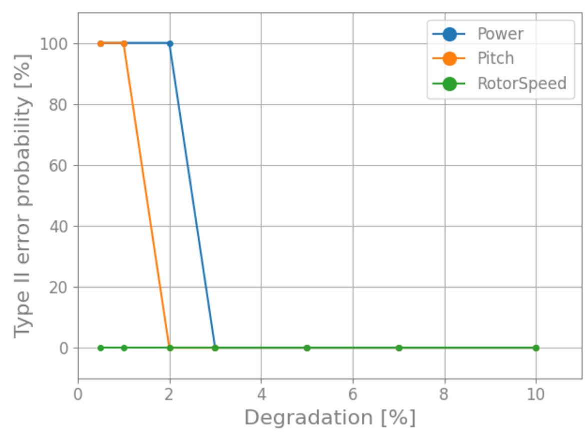

Figure C.32: False negative rate (Type II error) performance of the model for a step degradation with Monte-Carlo simulation.

C.3 Turbine 3

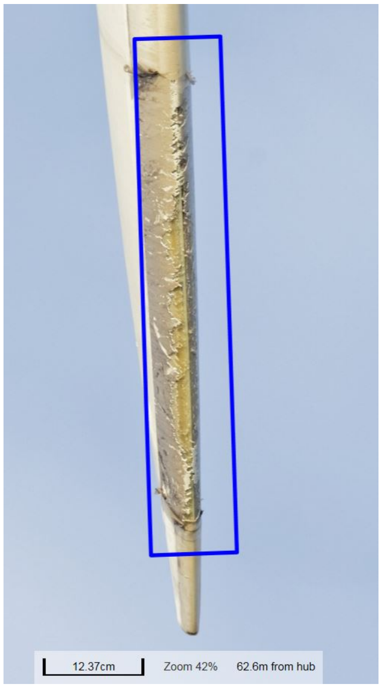

Figure C.33: Inspection November 2017, turbine 3, blade 1.

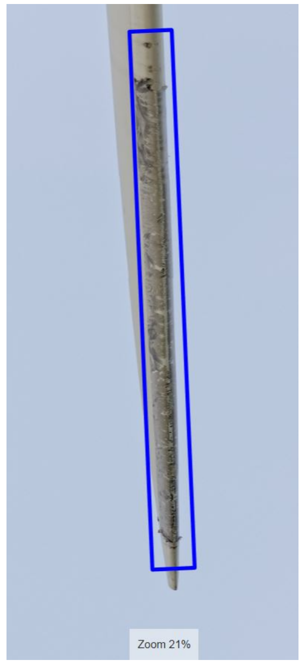

Figure C.34: Inspection November 2017, turbine 3, blade 2.

Figure C.35: Inspection November 2017, turbine 3, blade 3.

Figure C.36: Inspection April 2021, turbine 3, blade 2.

Figure C.37: Inspection April 2021, turbine 3, blade 3.

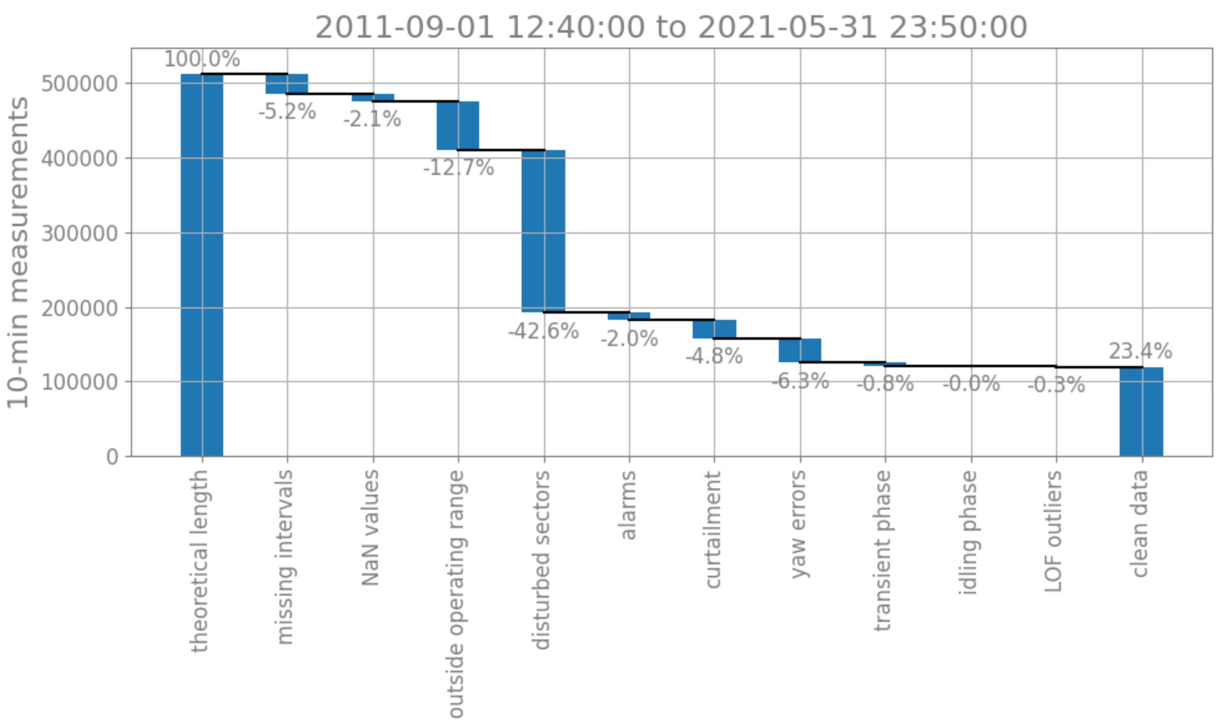

Figure C.38: Overview of filtering statistics , turbine 3.

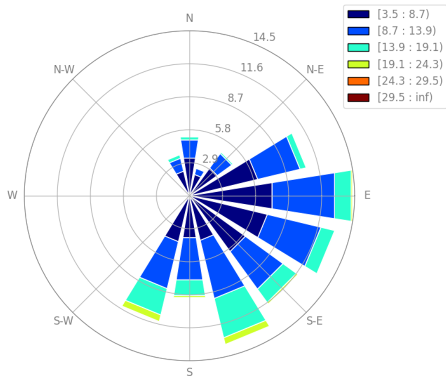

Figure C.39: Free wind sector for turbine 3.

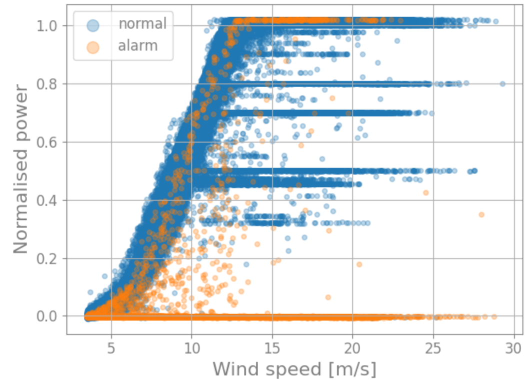

Figure C.40: Flagged alarms for turbine 3.

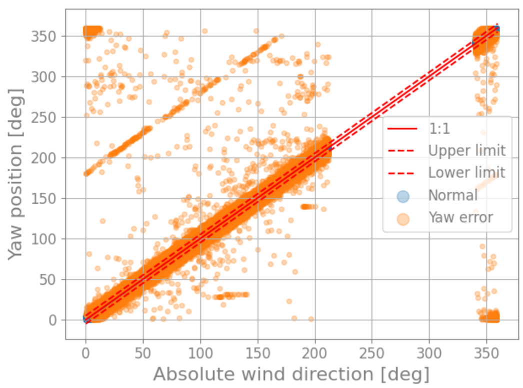

Figure C.41: Flagged yaw errors for turbine 3.

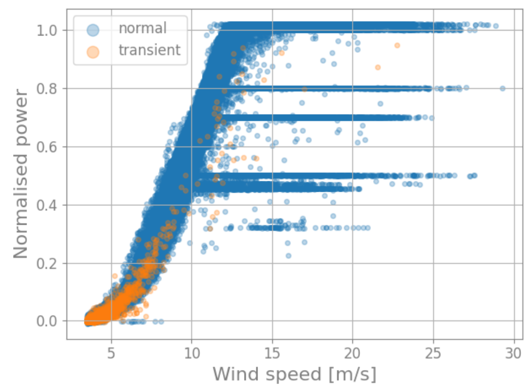

Figure C.42: Flagged transient phase measurements for turbine 3.

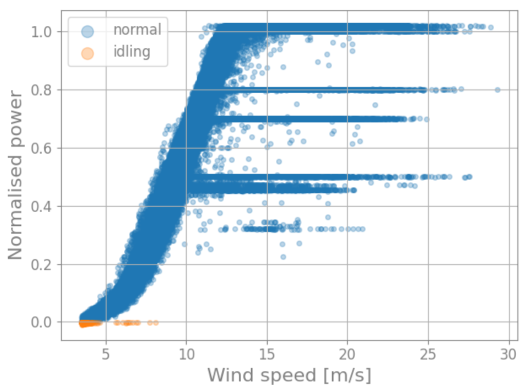

Figure C.43: Flagged idling phase measurements for turbine 3.

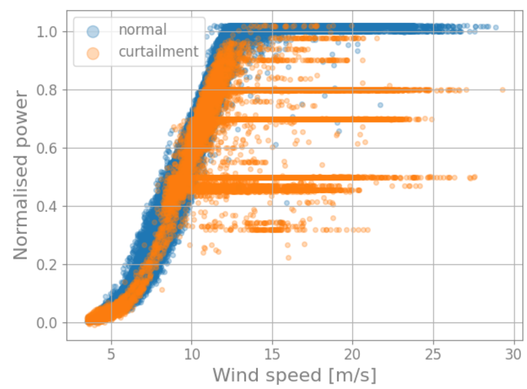

Figure C.44: Flagged curtailment for turbine 3.

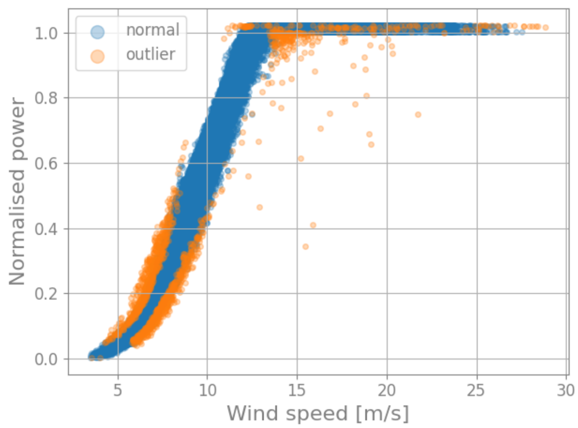

Figure C.45: Flagged outliers for turbine 3.

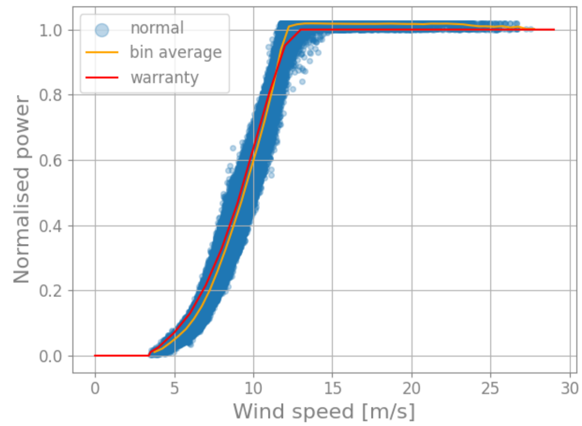

Figure C.46: Final clean power curve for turbine 3.

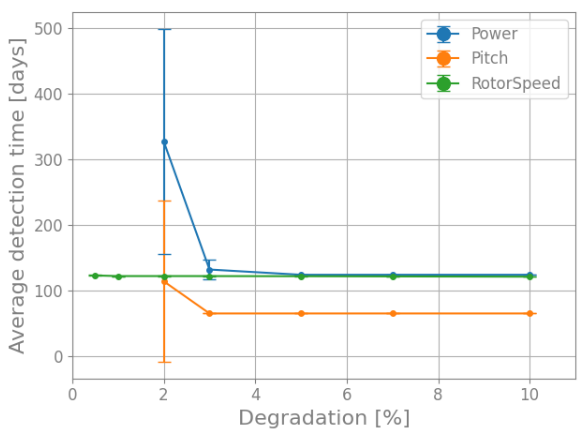

Figure C.47: Average detection time performance of model for step degradation with Monte-Carlo simulation.

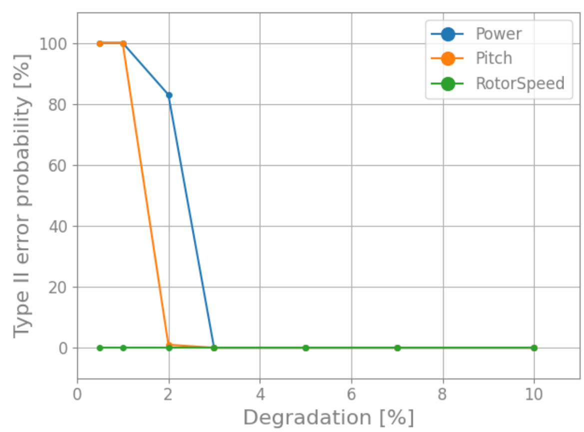

Figure C.48: False negative rate (Type II error) performance of the model for a step degradation with Monte-Carlo simulation.