Maik Dennis Reder

Reliability Models and Failure Detection Algorithms for Wind Turbines

Departamento

Centro de Investigación de Recursos y Consumos Energéticos (CIRCE)

Director/es

MELERO ESTELA, JULIO JAVIER

Tesis Doctoral

RELIABILITY MODELS AND FAILURE DETECTION ALGORITHMS FOR WIND TURBINES

Autor

Maik Dennis Reder
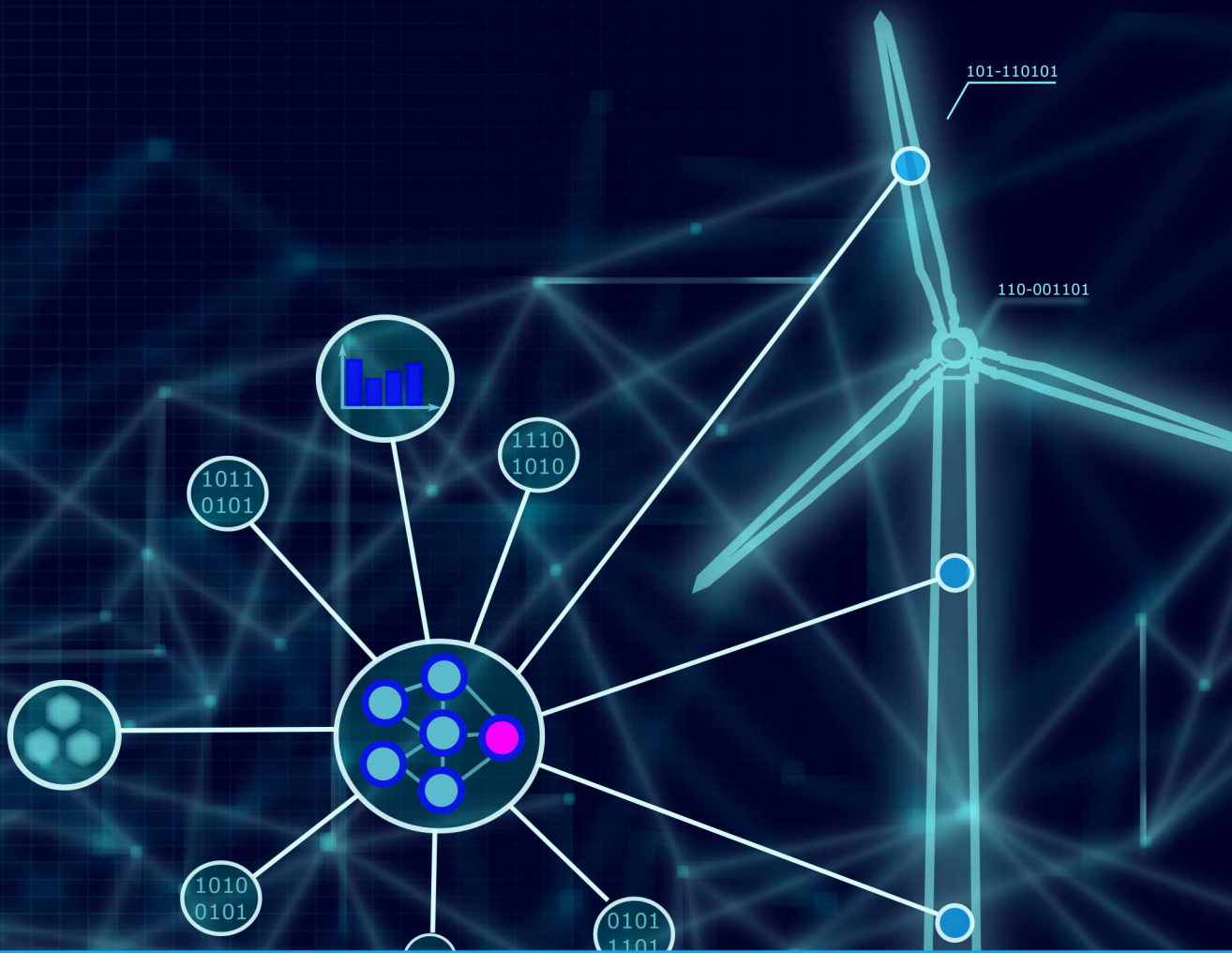
Director/es

MELERO ESTELA, JULIO JAVIER

UNIVERSIDAD DE ZARAGOZA

Centro de Investigación de Recursos y Consumos Energéticos (CIRCE)

2018

Reliability Models and Failure Detection Algorithms for Wind Turbines

Maik D. Reder

101-110101

110-001101

1110
1010

1011
0101

1010
0101

0101
1101

Reliability Models and Failure Detection Algorithms for Wind Turbines

Maik D. Reder

Doctoral thesis written at CIRCE Research Institute - University of Zaragoza within the 'Advanced Wind Energy Operation and Maintenance' (AWESOME) framework.

Memoria presentada a la Universidad de Zaragoza para la obtención del título de Doctor con Mención de Doctorado Internacional.

Zaragoza, 31-May-2018

Director:
Prof. Julio J. Melero Estela, *CIRCE Research Institute - University of Zaragoza*

In collaboration with:

List of Publications

Peer-reviewed publications included in this thesis (I–VII):

(I) Reder M, Gonzalez E, Melero J J. Wind Turbine Failures - Tackling Current Problems in Failure Data Analysis. Journal of Physics: Conference Series (2016), 753 (072027). DOI: `http://doi.org/10.1088/1742-6596/753/7/072027`.

(II) Reder M, Yürüşen N, Melero J J. Data-Driven Learning Framework for Associating Weather Conditions and Wind Turbine Failures. Reliability Engineering & System Safety (2018), 169, 554 – 569. DOI: `http://doi.org/10.1016/j.ress.2017.10.004`.

(III) Reder M, Melero J J. Time Series Data Mining for Analysing the Effects of Wind Speed on Wind Turbine Reliability. Safety and Reliability – Theory and Applications (2017), CRC Press Taylor & Francis. DOI: `http://doi.org/10.1201/9781315210469-93`.

(IV) Reder M, Melero J J. Modelling Wind Turbine Failures based on Weather Conditions, Journal of Physics: Conference Series (2017), 926 (012012). DOI: `http://doi.org/10.1088/1742-6596/926/1/012012`.

(V) Reder M, Melero J J. Modelling the Effects of Environmental Conditions on Wind Turbine Failures. Wind Energy (2018), 1–16. DOI: `http://doi.org/10.1002/we.2201`.

(VI) Reder M, Melero J J. A Bayesian Approach for Predicting Wind Turbine Failures based on Meteorological Conditions. *Accepted for publication in:* Journal of

Physics: Conference Series (2018).

(VII) Reder M, Tautz-Weinert J, Colone L, Watson S J, Melero J. Automated Fault Detection Algorithms for Wind Turbines using CMS and SCADA Data. *Under Review in:* Wind Energy (2018).

Collaborations and other publications (VIII–XIV):

These publications have been published during the time of this PhD but are not included directly in this thesis. Papers (VIII) to (XIII) are peer reviewed scientific publications. Paper (XIV) is a technical report published within the AWESOME project.

(VIII) Gonzalez E, Reder M, Melero J J. SCADA Alarms Processing for Wind Turbine Component Failure Detection. Journal of Physics: Conference Series (2016), 753 (072019). DOI: `http://doi.org/10.1088/1742-6596/753/7/072019`.

(IX) Reder M, Melero J J. Assessing Wind Speed Effects on Wind Turbine Reliability. WindEurope Summit (2016), Hamburg. DOI: `http://doi.org/10.13140/RG.2.2.11134.59200`.

(X) Colone L, Reder M, Tautz-Weinert J, Melero J J, Natarajana A, Watson S J. Optimisation of Data Acquisition in Wind Turbines with Data-Driven Conversion Functions for Sensor Measurements. Energy Procedia (2017), 137, 571 − 578. DOI: `http://doi.org/10.1016/j.egypro.2017.10.386`.

(XI) Yürüşen N Y, Reder M, Melero J J. Failure Event Definitions & their Effects on Survival and Risk Analysis of Wind Turbines, Safety and Reliability − Theory and Applications (2017), CRC Press Taylor & Francis. DOI: `http://doi.org/10.1201/9781315210469-93`.

(XII) Yürüşen N Y, Reder M, Melero J J. Evaluation of different Wind Turbine Technologies in terms of Maintainability and Reliability. *Under preparation for:* Wind Energy (2018).

(XIII) Colone L, Reder M, Dimitrov N, Straub D. Assessing the Utility of Early Warning Systems for Detecting Failures in Major Wind Turbine Components. *Accepted for publication in:* Journal of Physics: Conference Series (2018).

(XIV) Artigao E, Colone L, Pandit R, Reder M, Weinert J, Ziegler L. Optimisation of data acquisition in wind turbines with data-driven conversion functions for measurements. Tech. Rep. (2016). Melero J J, Muskulus M, Smolka U, editors. In: 1st Joint Industry Workshop Scientific Report. Zaragoza, Spain. URL: `http://awesome-h2020.eu/1st-joint-industry-workshop-scientific-report/`. Last Accessed: 14/03/2018.

Contribution to Papers with Multiple Authors:

In the following, the contribution to each of the multi-author papers is stated.

Paper (I): This paper was a joint study with Elena Gonzalez (CIRCE - University of Zaragoza). I was responsible for developing the taxonomy, gathering and cleaning the historical failure and downtime data and carrying out the failure analysis. My co-author Elena was gathering and cleaning the SCADA alarm data for the subsequent alarm analysis. In this thesis, however, only the parts, which I developed are presented.

Paper (II): This was a joint study with Nurseda Y. Yürüşen (CIRCE - University of Zaragoza). We both contributed equally to all parts of this work including the development of the idea and carrying out the analysis. Nurseda an I wrote two JCR journal papers together (Papers (II) and (XII)), to which we both contributed equally to. Due to the similarity to our respective PhD topics, I will be presenting this paper in my thesis and Nurseda will be using Paper(XII).

Paper (VII): This paper was written in collaboration with Jannis Tautz-Weinert (University of Loughborough, UK) and Lorenzo Colone (Technical University of Denmark). We all contributed to the development of the general idea of the paper.

As first author, I was responsible for the major part of the work. This included: gathering the CMS data, developing the binning method and merging CMS and SCADA records, writing the machine learning algorithms for the predictions of CMS alarm counts and the probabilities over time, and developing the codes for the distance based automated vibration evaluation (DAVE). Jannis focused on writing the code for analysing the relationship between the data source and helped out in the DAVE framework. In the final part of the paper (alarm time shifting), I wrote the codes for the two algorithms RF and ANN. Lorenzo and I developed the codes for the receiver operator characteristic curves together.

Paper (VIII):   This paper was a joint study with Elena Gonzalez (CIRCE - University of Zaragoza) using failure and alarm data. Elena focused on gathering and analysis the SCADA alarm data, while I was responsible for the failure data to which we associated the alarms.

Paper (X):   This paper was written in collaboration with Lorenzo Colone (Technical University of Denmark) and Jannis Tautz-Weinert (University of Loughborough, UK). All three of us contributed equally to developing the idea and writing the final manuscript of the paper. Lorenzo was responsible for the data preparation. I was responsible for developing the codes for the Random Forest, Generalised Linear Models and Gradient Boosting Machines. Jannis developed the codes for the Artificial Neural Networks.

Paper (XI):   In this paper Nurseda Y. Yürüşen (CIRCE - University of Zaragoza) is first author and carried out the major part of the work. I contributed as second author with to the wind turbine reliability modelling and helped writing the final paper.

Paper (XII):   This was a joint study with Nurseda Y. Yürüşen (CIRCE - University of Zaragoza). We both contributed equally to the idea and the development of the paper. Nurseda focused on the maintainability of the turbines using virtual age models. While, I was responsible for the reliability models (Andersen-Gill recurrent event model) as well as the analysis of the failure data.

Paper (XIII):   This paper was carried out in collaboration with Lorenzo Colone (Technical University of Denmark). As first author, Lorenzo was responsible for the largest part of this work and the general idea of the paper. I contributed to the paper by working on the variable selection and sensitivity study based on the LASSO regularisation and helped writing the manuscript.

Publication (XIV):   This is a scientific report based on the 1st Joint Industry Workshop held within the AWESOME project in Berlin, Germany. All authors contributed equally to this work.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Julio J. Melero for the guidance, constructive criticism and friendly advice during the planning and development of this thesis. This work would not have been possible without his aid and support.

I am very thankful to all colleagues at the wind energy department of CIRCE for the valuable input and help. Here, special thanks go to Elena Gonzalez and Nurseda Yürüşen, with whom I worked closely together throughout these years.

Furthermore, I would like to thank the other PhD students within the 'AWESOME' project: Estefania Artigao, Emmanouil Nanos, Helene Seyr, Jannis Tautz-Weinert, Lorenzo Colone, Lisa Ziegler, and Ravi Pandit. I am very glad that I had the opportunity to be part of this group.

I am grateful to the colleagues at ENEL Green Power for hosting me during my industrial secondment in Rome and providing me with an important part of the data that I used in my thesis. Particularly, I would like to thank Giannis Loizos, for giving me access to these data and for sharing his knowledge on wind turbine operation & maintenance.

I am thankful to all colleagues of the Structures and Components (SAC) group at the wind energy department of the Technical University of Denmark (DTU) for making my academic secondment at DTU possible and giving me valuable feedback regarding my research.

Last but not least, I would like to thank my whole family, especially my mum, dad and sisters for supporting me throughout my entire life, and I am very grateful to Teresa for the patience she had with me and the encouragement she offered during the time of my PhD studies.

Abstract

Over the past decades, the wind industry has been growing significantly, and wind energy has become one of the most important renewable energy sources in Europe. Nonetheless, especially when considering economic aspects, the wind energy sector is not yet fully competitive with conventional energy generation.

The main cost drivers of modern wind farms are activities related to Operation and Maintenance (O&M). This is due to the fact that current O&M practice is still primarily based on corrective or preventive actions. Applying predictive techniques, however, could significantly decrease the maintenance related costs and with this enhance the wind farms' overall revenue.

Although, the benefits of predictive O&M are increasingly being recognized, there is still a remarkable lack of research regarding these techniques. Advanced reliability models and failure prediction algorithms can enable operators to anticipate wind turbine (WT) component failures and to adapt their maintenance strategies accordingly.

To date, WT reliability models are almost exclusively based on the turbine age, as they were originally developed for machinery operating in fairly steady surroundings, e.g. indoors. However, since wind turbines are exposed to highly varying environmental conditions, these models are not reflecting the reality sufficiently well. New reliability models need to be developed that are capable of representing the failure behaviour of wind turbines and their components taking into account the meteorological and operational conditions at site.

Moreover, failure prediction is often carried out using either data obtained from Supervisory Control and Data Acquisition (SCADA) or Condition Monitoring Systems (CMS). Yet, in most modern wind turbines both systems are installed, and merging these sources can enhance failure detection.

This thesis aims at enhancing current O&M practice by: (1) developing ad-

vanced data-driven WT reliability and failure models including the environmental and operational conditions measured within the respective wind parks, and (2) establishing novel failure detection algorithms using environmental, operational and CMS data. These two objectives are split into four tasks.

In the first task, a thorough wind turbine failure and downtime analysis is carried out, providing an important understanding of the most commonly failing components of different WT technologies and their failure modes. This information is vital for further reliability and maintenance modelling.

Secondly, the meteorological conditions before failure events of main WT components are investigated. For this, a data-driven learning framework based on k-means clustering and apriori rule mining is developed, which is capable of handling large amount of data and deriving useful, human readable results. Additionally, anomaly detection and motif discovery algorithms are applied to find abrupt changes and recurrent patterns in wind speed time series before the failure events of main WT components.

Thirdly, a novel reliability model that directly incorporates the meteorological conditions recorded throughout two months prior to failure is proposed. The model uses two separate statistical processes - one is generating the failure events as well as occasional zeros, and the other process generates the structural zeros. Additionally, possible unobserved effects (heterogeneity) at the wind farm site are taken into account and sophisticated regularisation techniques are applied to avoid problems due to over-fitting and multicollinearity. The model performance is tested using historical failure data and weather related parameters obtained at the wind farms' meteorological measurement masts.

In the final task, failure prediction algorithms based on meteorological conditions or vibrational and operational data are developed. A Bayesian network is trained in order to predict component failures within a wind farm mainly based on the on-site weather conditions. Then, an approach to merge CMS vibrations and SCADA data with the aim of analysing the relationships between both sources is introduced. Main bearing failure events are predicted using several artificial intelligence algorithms, such as random forests, gradient boosting machines, generalised linear models and artificial neural networks. Furthermore, a generic tool (Distance Based Automated Vibration Evaluation - DAVE) is developed, which is an on-line wind farm level failure detection framework based on CMS vibrations.

The results of this thesis show that the failure behaviour of wind turbine components is highly influenced by the site specific meteorological conditions. The presented data-driven learning framework is capable of identifying the general and time specific conditions before component failures. Furthermore, it is shown that with the herein proposed reliability models and failure detection algorithms, wind turbine O&M can be significantly enhanced. These reliability and failure detection models are the first ones to give a realistic and site specific representation, by considering complex combinations of on-site environmental conditions as well as health and operational indicators obtained from merged CMS vibrations and SCADA data, respectively.

Hence, this work provides practicable and effective frameworks, models and algorithms for further application in the field of predictive wind turbine O&M.

Resumen

Durante las pasadas décadas, la industria eólica ha sufrido un crecimiento muy significativo en Europa llevando a la generación eólica al puesto más relevante en cuanto a producción energética mediante fuentes renovables. Sin embargo, si consideramos los aspectos económicos, el sector eólico todavía no ha alcanzado el nivel competitivo necesario para batir a los sistemas de generación de energía convencionales.

Los costes principales en la explotación de parques eólicos se asignan a las actividades relacionadas con la Operación y Mantenimiento (O&M). Esto se debe al hecho de que, en la actualidad, la Operación y Mantenimiento está basada principalmente en acciones correctivas o preventivas. Por tanto, el uso de técnicas predictivas podría reducir de forma significativa los costes relacionados con las actividades de mantenimiento mejorando así los beneficios globales de la explotación de los parques eólicos.

Aunque los beneficios del mantenimiento predictivo se consideran cada día más importantes, existen todavía la necesidad de investigar y explorar dichas técnicas. Modelos de fiabilidad avanzados y algoritmos de predicción de fallos pueden facilitar a los operadores la detección anticipada de fallos de componentes en los aerogeneradores y, en base a ello, adaptar sus estrategias de mantenimiento.

Hasta la fecha, los modelos de fiabilidad de turbinas eólicas se basan, casi exclusivamente, en la edad de la turbina. Esto es así porque fueron desarrollados originalmente para máquinas que trabajan en entornos 'amigables', por ejemplo, en el interior de naves industriales. Los aerogeneradores, al contrario, están expuestos a condiciones ambientales altamente variables y, por tanto, los modelos clásicos de fiabilidad no reflejan la realidad con suficiente precisión. Es necesario, por tanto, desarrollar nuevos modelos de fiabilidad que sean capaces de reproducir el

comportamiento de los fallos de las turbinas eólicas y sus componentes, teniendo en cuenta las condiciones meteorológicas y operacionales en su emplazamiento.

La predicción de fallos se realiza habitualmente utilizando datos que se obtienen del sistema de Supervisión Control y Adquisición de Datos (SCADA) o de Sistemas de Monitorización de Condición (CMS). Cabe destacar que en turbinas eólicas modernas conviven ambos tipos de sistemas y la fusión de ambas fuentes de datos puede mejorar significativamente la detección de fallos. Esta tesis pretende mejorar las prácticas actuales de Operación y Mantenimiento mediante: (1) el desarrollo de modelos avanzados de fiabilidad y detección de fallos basados en datos que incluyan las condiciones ambientales y operacionales existentes en los parques eólicos y (2) la aplicación de nuevos algoritmos de detección de fallos que usen las condiciones ambientales y operacionales del emplazamiento, así como datos procedentes tanto de sistemas SCADA como CMS. Estos dos objetivos se han dividido en cuatro tareas.

En la primera tarea se ha realizado un análisis exhaustivo tanto de los fallos producidos en un amplio conjunto de aerogeneradores (amplio en número de turbinas y en longitud de los registros) como de sus tiempos de parada asociados. De esta forma, se han visualizado los componentes que más fallan en función de la tecnología del aerogenerador, así como sus modos de fallo. Esta información es vital para el desarrollo posterior de modelos de fiabilidad y mantenimiento.

En segundo lugar, se han investigado las condiciones meteorológicas previas a sucesos con fallos de los principales componentes de los aerogeneradores. Se ha desarrollado un entorno de aprendizaje basado en datos utilizando técnicas de agrupamiento 'k-means clustering' y reglas de asociación 'a priori'. Este entorno es capaz de manejar grandes cantidades de datos proporcionando resultados útiles y fácilmente visualizables. Adicionalmente, se han aplicado algoritmos de detección de anomalías y patrones para encontrar cambios abruptos y patrones recurrentes en la serie temporal de la velocidad del viento en momentos previos a los fallos de los componentes principales de los aerogeneradores. En la tercera tarea, se propone un nuevo modelo de fiabilidad que incorpora directamente las condiciones meteorológicas registradas durante los dos meses previos al fallo. El modelo usa dos procesos estadísticos separados, uno genera los sucesos de fallos, así como ceros ocasionales mientras que el otro genera los ceros estructurales necesarios para los algoritmos de cálculo. Los posibles efectos no observados (heterogeneidad) en

el parque eólico se tienen en cuenta de forma adicional. Para evitar problemas de 'over-fitting' y multicolinearidades, se utilizan sofisticadas técnicas de regularización. Finalmente, la capacidad del modelo se verifica usando datos históricos de fallos y lecturas meteorológicas obtenidas en los mástiles meteorológicos de los parques eólicos.

En la última tarea se han desarrollado algoritmos de predicción basados en condiciones meteorológicas y en datos operacionales y de vibraciones. Se ha 'entrenado' una red de Bayes, para predecir los fallos de componentes en un parque eólico, basada fundamentalmente en las condiciones meteorológicas del emplazamiento. Posteriormente, se introduce una metodología para fusionar datos de vibraciones obtenidos del CMS con datos obtenidos del sistema SCADA, con el objetivo de analizar las relaciones entre ambas fuentes. Estos datos se han utilizado para la predicción de fallos en el eje principal utilizando varios algoritmos de inteligencia artificial, 'random forests', 'gradient boosting machines', modelos generalizados lineales y redes neuronales artificiales. Además, se ha desarrollado una herramienta para la evaluación on-line de los datos de vibraciones (CMS) denominada DAVE ('Distance Based Automated Vibration Evaluation').

Los resultados de esta tesis demuestran que el comportamiento de los fallos de los componentes de aerogeneradores está altamente influenciado por las condiciones meteorológicas del emplazamiento. El entorno de aprendizaje basado en datos es capaz de identificar las condiciones generales y temporales específicas previas a los fallos de componentes. Además, se ha demostrado que, con los modelos de fiabilidad y algoritmos de detección propuestos, la Operación y Mantenimiento de las turbinas eólicas puede mejorarse significativamente. Estos modelos de fiabilidad y de detección de fallos son los primeros que proporcionan una representación realística y específica del emplazamiento, al considerar combinaciones complejas de las condiciones ambientales, así como indicadores operacionales y de estado de operación obtenidos a partir de la fusión de datos de vibraciones CMS y datos del SCADA. Por tanto, este trabajo proporciona entornos prácticos, modelos y algoritmos que se podrán aplicar en el campo del mantenimiento predictivo de turbinas eólicas.

List of Abbreviations

1-D One-Dimensional.

ACC prediction accuracy.

AIC Akaike information criterion.

AICc Akaike information criterion with correction factor.

ANN Artificial Neural Network.

ARM Association rule mining.

AWESOME Advanced Wind Energy Systems Operation and Maintenance.

BBN Bayesian Belief Network.

CDF cumulative density function.

CMS Condition Monitoring Systems.

CTMC continuous-time Markov Chains.

DAVE Distance-based automated vibration evaluation.

DB Data Base.

DD direct drive.

DFIG doubly fed induction generators.

DTMC discrete-time Markov Chains.

DTU Danish Technical University.

DTW Dynamic Time Warping.

EGP ENEL Green Power.

EM Expectation Maximisation algorithm.

Enet elastic net.

Env Envelope.

ESD Extreme Studentised Deviate Test.

EU European Union.

FAR false alarm rate.

FFT Fast-Fourier-Transform.

FN false negative.

FP false positive.

G $\geq$ 1 MW geared WTs with rated capacities above or equal to 1 MW.

G $<$ 1 MW geared WTs with rated capacities below 1 MW.

GBM Gradient Boosting Machines.

GDE Generator Drive End.

GLM Generalised Linear Model.

GNDE Generator Non-Drive End.

HC Hierarchical Clustering.

HPP Homogeneous Poisson Process.

HSS high speed shaft.

IMS intermediate speed shaft.

LASSO least absolute shrinkage and selection operator.

LSS low speed shaft.

MAE mean absolute error.

MaxWS maximum wind speed.

MB Main Bearing.

MCC Mathews correlation coefficient.

MCMC Markov Chain Monte Carlo.

MCP minimax concave penalty.

MDA Mean Decrease Accuracy.

MERRA Modern-Era Retrospective Analysis for Research and Applications.

met mast measurement tower.

MLE maximum likelihood estimation.

MSE Mean Squared Error.

MTTF mean time to failure.

Multi-D Multi-Dimensional.

NCAR National Center for Atmospheric Research.

NCEP National Centers for Environmental Prediction.

NegBin negative-binomial.

NHPP Non-Homogeneous Poisson Process.

NREL National Renewable Energy Laboratory.

O&M Operation and Maintenance.

PAA Piecewise Aggregate Approximation.

PDF probability density function.

PEVS parameter estimation and variable selection.

PH proportional hazard models.

PMG permanent magnet generator.

POD probability of detection.

PS Planetary Stage.

PWR average monthly active power in percent of the rated capacity.

RF Random Forests.

RH relative humidity.

RMS root mean square amplitude.

RMSE root mean square error.

ROC receiver operator characteristic.

ROCOF rate of occurrence of failures.

RP Renewal Process.

S-H-ESD Seasonal Hybrid Extreme Studentised Deviate Test.

SAX Symbolic Aggregate approXimation.

SCAD smoothly clipped absolute deviation penalty.

SCADA Supervisory Control And Data Acquisition.

STL Seasonal and Trend decomposition using LOESS.

SubDB Data Base sub-set.

Temp ambient temperature.

TI turbulence intensity.

TN true negative.

TNR true negative rate.

TP true positive.

TPR true positive rate.

WF wind farm.

WS wind speed.

WT wind turbine.

WTs wind turbines.

ZINB zero-inflated negative-binomial.

ZIP zero-inflated Poisson.

List of Symbols

Symbol	Description
B	bias of the ANN nodes
C	A set of cluster centroids
H_0	the null-hypothesis
H_a	the alternative hypothesis
I	Item set in association rule mining
L	the logarithm of the likelihood function $\mathcal{L}$
$M(\beta)$	mean number of successes on a time interval
$M(t)$	expected number of failures
MAD	median of the absolute deviation
$P_{i,j}$	transition probability from state i to state j
Pr	probability
R_T	Residual of a time series
R_q	test statistics ESD
R	vector of r possible candidates for outliers.
S_T	Standard deviation of a time series
S_j	The set of data points assigned to the jth cluster
TR	Transaction in association rule mining
T	A time series
$W(t)$	mean number of failure events on a time interval in a counting process

Symbol	Description
$W_T(t)$	non-overlapping window in time series
$X(t)$	state of a Markov process at time t
Γ	Gamma function
$\mathcal{L}$	likelihood function
Φ	penalty function
Ψ	cost function
α	significance level
$\bar{T}$	Mean of a time series
$\bar{a}$	alphabet in Motif discovery
β_i	estimation coefficients
β	model parameter
χ	state space in Markov process
δ	error at a neuron of an ANN
η	model parameter
$\gamma(z_k^l)$	activation function for a neuron with weighted input value z_k^l
γ	model parameter
$\widehat{\mathcal{L}}$	maximum value of the likelihood function
κ	learning rate
$\lambda(t)$	failure rate function
λ_q	critical values for each outlier candidate
λ	shrinkage parameter controlling the penalty
μ_i	mean number of successes on a time interval
ν	degrees of freedom
φ	parameter controlling the share of the ℓ_1 and ℓ_2 penalties
ρ	repair rate in Markov process
σ_i	zero inflated probability

Symbol	Description
σ	median of the absolute deviation
τ_i	unobserved effects
ϑ	dispersion parameter
$\widetilde{T}$	The median of a time series
$_bC_m$	total number of models trained per component
a	output of an ANN
b	types of different variables in the data set
c_j	The cluster centroid of the jth cluster
d	number of estimated parameters
$f(t)$	probability density function
i	Item in association rule mining
k	The optimal number of clusters
l	layer of an ANN
$m(t)$	repair rate of a counting process
m	members per combination in sensitivity study
n	number of observations
o_k^l	output of neuron k on layer l
r	possible number of outliers
$t_{\alpha/(2n-q+1)}$	upper critical value of the t-distribution
t	operational time
$w(t)$	rate of failure occurence in a counting process
$w_{j,k}^l$	weight of the connection between neurons j and k
x_i	model covariates
y	model response variable
z_i	zero inflated regressors
z_k^l	weighted input value of neuron k of layer l

List of Figures

List of Tables

Contents

1

Introduction

Wind power is currently the second largest form of electricity generation in the European Union (EU) [1] and is expected to grow significantly in order to reach the EU's greenhouse gas emission targets by 2020, [2]. Nonetheless, the cost related to wind energy generation needs to be further reduced to ensure economic feasibility and for it to be competitive with conventional energy sources in the future, [3].

A large share of the levelised cost of wind energy is directly related to Operation and Maintenance (O&M) actions, [4, 5]. In order to decrease these costs and to increase the wind farms' revenue, advanced reliability and failure prediction models are needed. With these, wind farm (WF) operators can anticipate faulty turbine states before serious damage occurs and can react accordingly.

In recent years, O&M has become an emerging field of research and current practice is shifting continuously from corrective towards predictive maintenance. WFs are often located in remote areas, e.g. offshore, and maintenance actions have to be planned with enough time in advance, taking into account for example the availability of spare parts and the accessibility to the wind farms. Unexpected wind turbine (WT) stops, due to e.g. component failures, can lead to long downtimes as well as higher costs and energy losses. Sophisticated O&M models can lower the risk of having unexpected downtimes and enable the operators to react adequately to the given situations. Here, anticipating WT component failures is a core element of predictive O&M practice and can be carried out using probabilistic reliability and failure models as well as condition based failure prediction techniques.

Nowadays, as vast amounts of data are produced by the turbines, data-driven solutions have shown to be very promising for these purposes. The information used for these approaches include failure data taken from maintenance logbooks, operational data obtained from the Supervisory Control And Data Acquisition (SCADA) system, information on the health conditions of various WT components provided by the Condition Monitoring Systems (CMS), as well as meteorological data from the WF's measurement tower (met mast).

1.1 Research Objectives

The global objective of this thesis is to develop advanced data-driven WT reliability and failure models, which take into account the environmental and operational conditions to which the turbines are exposed to. Furthermore, novel failure detection techniques are developed using a combination of environmental, operational data and information on the health conditions of various WT components. This PhD thesis was carried out at CIRCE – University of Zaragoza, ENEL Green Power (EGP) in Rome and the Danish Technical University (DTU) in RISØ. In order to achieve the given overall aim of the thesis, four research tasks have been defined. A summary of the research objectives included in each task is given in the following including a short description of the proposed solutions. In most of these tasks the open source programming language R, [6], was used for setting up the algorithms and carrying out the calculations.

1.1.1 Task 1 – Wind Turbine Failure Analysis

The core elements for any data-driven reliability model are, well structured and analysed, failure and maintenance data. The raw data are mostly collected by operators and need to be carefully pre-processed in order to be used in any further modelling approach. Therefore, firstly, each failure information needs to be assigned to the corresponding failed component using a so called taxonomy. The latter is a component-break down that unravels the WT system into a clearly structured scheme of sub-systems, assemblies and sub-assemblies, according to their functionality. For this thesis, a WT taxonomy has been developed, which can be applied to a wide range of data including different WT technologies, drive train set-ups, etc. Subsequently, an extensive failure data analysis has been carried out using

historical failure data obtained from wind turbines (WTs) of different drive train concepts, ages and rated capacities. The taxonomy and the failure analysis have been published in Paper (I). Additionally, a failure mode analysis has been carried out to identify the cause of various component failures. This has not been published elsewhere. The research objectives of this task are:

- Develop a modern WT taxonomy based on manufacturer information of different WT technologies, such as drive train concepts, power regulation etc. This is expected to eliminate one of the major issues in wind turbine failure analysis, which is related to the non-uniform data treatment throughout literature.
- Determine the yearly failure rates and downtime for each WT sub-system and assembly of different WT technologies. This information is fundamental for further reliability and maintenance modelling and helps to identify the most critical WT components.
- Investigate on the failure modes of the main WT components. With this, the primary causes of these failures can be identified and further research can focus on optimising the components respectively.

1.1.2 Task 2 – Analysing Meteorological Conditions and Component Failures

The failure behaviour of WTs and their components is influenced by their age, their health conditions, as well as the site specific meteorological and operational conditions. These can vary highly throughout the year and can affect the various WT components differently. As stated in [7], understanding the complex combinations of environmental conditions that lead to failures is vital for the development of advanced reliability models and there is a serious lack of research concerning these effects. This task aims at:

- Establishing a methodology that serves to identify the critical meteorological conditions right before WT component failures. This shall be applied to a data base containing historical failures and weather data. Furthermore, it should be extendible and adaptable to similar problems in other areas.
- Analysing the short term changes in the wind speed conditions before failure. These variations are expected to have a high impact on the reliability of WT components. The wind speed variations obtained during 'healthy' WT operation

shall be compared to the ones detected towards the time of failure in order to investigate if higher variations in wind speeds can be correlated to the failures.

For this task, a framework is developed based on machine learning techniques such as k-means clustering and Association rule mining (ARM), that allows to analyse combinations of several environmental variables immediately before failure in big data bases. The analysed meteorological variables are wind speed, relative humidity and ambient temperature. Furthermore, the active power production is included as indicator for the turbines' operational performance. It is shown which component is affected mostly by certain environmental and operational conditions. The results of this analysis were published in Paper (II). Then, the short term changes in wind speed conditions are investigated using time series data mining techniques. These results have been published in Paper (III).

1.1.3 Task 3 – Advanced Reliability Models based on Meteorological and Operational Conditions:

To date, existing reliability models are almost exclusively based on the age of the components or turbines. These models were originally developed for machinery that is mostly operating in stationary surroundings, such as warehouses or machine shops. However, as WTs are exposed to highly variable weather conditions, which have shown to affect the failure behaviour of WTs and their components, these models are not sufficiently representing the reality. After having analysed the meteorological condition before failure, the objectives of this task are to:

- Develop an advanced WT reliability model taking into account the highly variable meteorological and operational conditions the turbines are exposed to.
- The performance of this model shall be compared to other modelling techniques. Furthermore, it has to be ensured that the model can be used to represent the failure behaviour of the whole wind turbine system as well as one single component.

A probabilistic WT failure model has been developed, based on a zero inflated negative binomial distribution using penalised parameter estimation techniques. Its performance is compared to several other probabilistic models, including conventional methods used in WT reliability modelling. The model is based on six environmental variables, including ambient temperature, wind speed, wind gusts,

relative humidity, precipitation and turbulence intensity. Furthermore, turbine and site specific variables, such as WT age, hub height, rated capacity, terrain complexity, and active power production are accounted for in the model. As the weather often has a delayed or cumulative effect on the failure behaviour, a model memory is included accounting for the meteorological conditions observed some time before the failure occurred. The results were published in Papers (IV) and (V).

1.1.4 Task 4 – Failure Prediction based on Operational and Condition Monitoring Data:

The forth and final objective of this thesis is to develop fault detection algorithms based on different data sources. This task is divided into three research objectives, which include:

- Establishing a method for failure detection on wind farm level based on environmental conditions. This method should be able to detect whether one or more failures of several main WT components occur during a specific month.
- Exploring the possibilities of combining information on operational WT conditions and internal component specific health conditions by merging SCADA and CMS data, with the aim of enhancing predictive O&M modelling.
- Development of a generic tool for failure detection on wind farm level using only vibration data.

For the first objective, a Bayesian method will be introduced to predict WT faults based only on information of the external conditions to which the turbines are exposed to.

Subsequently, for the second objective, the possibilities of using merged SCADA and CMS data for fault detection will be explored. Fault detection in wind turbines is usually carried out by either analysing SCADA or CMS data with the aim of finding any abnormal behaviour. To the author's knowledge no studies have been carried out yet combining SCADA and CMS data for WT fault detection. As, however, these two sources can contain substantially different information on external and internal conditions, which might be important for fault detection, it is expected that joining these data sources leads to significantly better results. For this, the relationship between these two data sources needs to be carefully analysed and

suitable techniques and processes for fault detection need to be found. The task explores the performance of several classification and regression algorithms, such as for example artificial neural networks, random forests, gradient boosting machines, and generalised linear models. Finally, a generic tool for failure detection on wind farm level, based on dynamic time warping and using only vibrational data is presented. The approaches and results of task 4 were published in Papers (VI) and (VII).

1.2 Related Projects

This PhD thesis was carried out within the Advanced Wind Energy Systems Operation and Maintenance (AWESOME) project, an H2020 ITN research and innovation programme funded by the EU under the Marie Skłodowska-Curie grant agreement No. 642108. The AWESOME project is coordinated by Prof. Dr. Julio J. Melero at CIRCE – University of Zaragoza and involves eleven PhD students working in the field of wind farm O&M at different institutions throughout Europe. More information can be found on the project website [8].

1.3 Structure of the Thesis

This thesis is divided into seven chapters. In Chapter 1 a brief introduction to the topic is given and the research objectives of this thesis are outlined. Chapters 2 to 7 represent the main part of the thesis and each of these chapters will provide a literature review, an introduction to the topic, as well as the methodology and results for the respective research task. Thus, although to some extend the chapters are thematically building on each other, they can also be treated as stand-alone and informative in themselves. Nonetheless, a broad introduction on wind energy will not be given in this thesis, but the interested reader can find extensive summaries e.g. in [9, 10]. Chapter 2 is concerned with wind turbine failures and failure mode analysis (Task 1). In Chapter 3 the meteorological conditions leading to WT failures are investigated (Task 2). Subsequently, in Chapter 5, a failure model is developed based on several environmental, operational and site specific conditions, as discussed in Task 3. Chapters 6 and 7 discuss several methods for failure detection on wind farm level, based on environmental conditions as well as on merged SCADA and

CMS data. Chapters 8 and 9 conclude the thesis with a discussion of the obtained results and an outlook on further studies in English and Spanish, respectively.

2

Wind Turbine Failure Analysis

In this chapter, at first, a review on existing literature in WT failure analyses is given. Then, the methodologies and results of the herein carried out failure analysis are discussed. Furthermore, the failure modes that led to the respective component damage are presented. The main results of this chapter have been published in Paper (I), however, un-published results including the analysis of pitch and stall regulated turbines as well as a failure mode analysis are included.

2.1 Background and Motivation

The core element of data-driven reliability models are comprehensive failure and maintenance data. Several failure analyses have been carried out investigating on the failure behaviour of different WT technologies, various components and sub-assemblies. However, as stated in [7], the comparison of these studies is rather difficult due to non-uniform data treatment. This can be avoided by using a clear and uniform taxonomy throughout the whole field of wind energy. A taxonomy is a component break down used to assign each reported failure to the corresponding sub-system, assembly and sub-assembly. These are usually classified regarding their functionality and/or physical location within the WT system.

2.1.1 Wind Turbine Taxonomies

Many different WT taxonomies have been developed over the last decade. These include for example a component break down presented by the SANDIA Laboratories [11], a taxonomy developed by the VTT Technical Research Centre of Finland [12], the RDS-PP® taxonomy, published by VGB PowerTech e.V. [13], and the ReliaWind taxonomy [14]. The latter two represent the most recent approaches. As the ReliaWind taxonomy was applied to an extensive wind turbine failure data analysis, it was chosen as base for this study.

Nonetheless, certain drawbacks have been identified, of which the most severe one is the fact that none of these taxonomies are publicly available. Additionally, as the ReliaWind study ended in 2011, most of the 350 WTs used to develop the taxonomy represent older WT technologies built before 2008. However, over the past decade the WT configurations have changed significantly including new drive-train concepts, larger rotor diameters and hub heights, as well as higher rated capacities, [15]. Hence, there is a significant need for developing a modern WT taxonomy taking into account several different turbine technologies and configurations.

2.1.2 WT Failure Statistics

A series of wind turbine failure data analyses have been published in the past, yet, most are based on data of very few WTs and/or outdated technologies. The most important contributions in literature are discussed in the following. For example in [16] a failure analysis of 72 WTs located in Finland is carried out. The data were obtained between the years 1996 and 2008.

Extensive failure data from operating WTs in Denmark, Sweden and Germany are collected within the WindStats data base, [17–19]. Around 4500 Danish, 2500 German and 1200 Swedish WTs are under observation. However, more information on the wind turbine size, age and type is not available. Some results of the WMEP project ('250 MW Wind') containing failure data obtained between the years 1989 and 2008 of more than 1500 German WTs with 350 MW rated power installed were published in [20–23]. Furthermore, for the previously mentioned ReliaWind project, failure information of 350 wind turbines over a varying period of time between 2008 and 2010 were analysed, [14, 24]. More recent approaches were published in [4, 25], where the failure data of approximately 350 offshore turbines with nominal power

between 2 and 4 MW are analysed. In [26], the failure rates of WTs equipped with doubly fed induction generators (DFIG) and permanent magnet generator (PMG) technologies were compared. This study was based on around 2200 onshore WTs with nominal power between 1.5 and 2.5 MW .

A summary of failure statistics from Swedish, Finish and German data bases is given in [27]. An extensive review on failure data bases is presented in [28], they further recommend that the results presented in Paper (I) of this thesis should be used for upcoming applications in the field, as they represent the latest and most relevant results in failure data analysis.

Table 2.1. Wind turbine failure data collection initiatives and publications. Adapted from [28] and extended.

Name	Country	Start-End	Number of WTs	Source
This thesis	**Europe, North and South America**	**01.01.2013 – 31.12.2015**	**~4400 geared; 215 direct-drive**	**[29]**
CREW - Data Base	USA	2011 – ongoing	~900	[30, 31]
CWEA - Data Base	China	2010–2012	unknown	[32]
Elforsk, Vindstat	Sweden	1989–2005	786	[27, 33]
EPRI	USA	1986–1987	290	[34]
EUROWIN	Europe	1986–1995	~3500	[35, 36]
Garrad Hassan	Worldwide	1992–2007	unknown	[37]
Huadian	China	01.2012–05.2012	1313	[38]
LWK	Germany	1993–2006	643	[19]
Lynette	USA	1981–1989	unknown	[39]
Muppandal	India	2000-2004	15	[40]
NEDO	Japan	2004–2005	924	[41]
Reliawind	Europe	2008–2010	350	[14, 24]
Robert Gordon University	UK	1997–2006	77	[42]
Round 1 offshore WF	UK	2004–2007	120	[43]
University of Nanjing	China	2009–2013	108	[44]
SPARTA	UK	2013–ongoing	1045	[45]
Strathclyde	UK	2005-2010	350	[4, 25, 26]
VTT	Finnland	1991–ongoing	72–96	[16, 27, 46]
Windstats	Germany, Denmark	1994–2004	7000	[19, 47, 48]
WInD-Pool	Europe, Germany	2013–ongoing	456	[49–52]
WMEP	Germany	1989–2008	1593	[22, 23]

Table 2.1 is adapted and extended from [28] and summarises some of the most relevant failure data collection initiatives and publications. All initiatives presented in Table 2.1 consist of data taken from onshore turbines, except for "Round 1 offshore" and "Strathclyde", which contained offshore data and "WInD-Pool", which collected data from both, onshore and offshore turbines.

2.2 Data Base

The historical data used in this task consist of failure logbooks obtained during three years of operation (from January 2013 to December 2015). The analysed turbines are three bladed onshore WTs with rated capacities between 300 kW and 3 MW. Different drive train technologies, such as geared and direct drive WTs, from 14 different manufacturers are considered and all wind farms are located at different sites in Europe, North and South America. The direct drive turbines have rated capacities between 600 kW and 2 MW.

In this thesis a WT component failure is defined as an event that leads to a WT stop requiring intervention such as repair or replacement of the faulty component. This excludes, (bi-)annual inspections, cleaning activities, downtime due to grid or meteorological restrictions, etc.

Table 2.2. Composition of the data base.

Total failure Events	~7000
Avg. yearly number of wind farms	230
Containing:	
Avg. yearly number of WTs under 1 MW	2130
Avg. yearly number of WTs equal or over 1 MW	2270
Number of direct drive turbines	215
Mean yearly installed capacity (MW)	5818

Table 2.2 shows the composition of the data base used in this study. The number of turbines in operation changed slightly throughout the observation period, thus, the average yearly number of wind farms, wind turbines, failure events and installed capacity are displayed. In total around 7000 failure events were recorded over the period of three years. In Table 2.3 the minimum, maximum, mean and median age of the turbines (at the beginning of the observation period: January 2013) represented in the data base are summarised for each WT technology.

Table 2.3. Age of the turbines in the data base at the start of the observation period (January 2013).

Technology	Age (years)			
	Min	Median	Mean	Max
All Turbines	0.08	5.19	6.49	22.42
Geared WTs $< 1\,\text{MW}$	0.66	9.94	9.99	22.41
Geared WTs $\geq 1\,\text{MW}$	0.08	1.96	2.75	11.19
Direct Drive WTs	0.67	5.92	6.08	11

In comparison to most of the previous studies, reviewed in Section 2.1, a significantly higher amount of operating turbines is examined. Furthermore, newer WT technologies are presented in this data base, as well as a high diversity of different manufacturers and technology set ups. With this, this study is expected to contribute significantly to the field of wind turbine failure analysis and reliability modelling.

2.3 Methodology

In this chapter, firstly, a taxonomy is developed based on detailed manufacturer information on recent WT technologies and failure data of wind turbines located in different countries throughout Europe, North and South America. The objective was to create a modern taxonomy that can be applied to data of several distinct WT technologies. For this, the ReliaWind taxonomy, which is one of the most sophisticated classification approaches, has been thoroughly reviewed. Furthermore, the latter has been rearranged and extended for modern turbine concepts, while considering the components' functionality and physical location. In order to enable the comparison of this study to previous ones, the taxonomy is in some aspects intentionally similar to the ReliaWind taxonomy. The herein developed modern taxonomy has been verified using a big data base, including a variety of different WT technologies and ages.

Subsequently, an extensive failure rate and downtime analysis is carried out using the failure data presented in Section 2.2. Therefore, the data base had to be cleaned carefully assigning the failure events to the affected sub-system, assembly or sub-assembly with the developed taxonomy. The failure rates as well as the hours of downtime per turbine and year have been calculated for each component. For

the analysis, the wind turbines are divided into geared WTs with rated capacities below 1 MW (G < 1 MW), geared WTs with rated capacities above or equal to 1 MW (G ≥ 1 MW) and direct drive (DD) technologies, as well as stall and pitch regulated turbines.

2.4 Failure Analysis Results

In this section the results of the failure data analysis are presented. At first the taxonomy will be shown, then, the failure rates and downtimes for each sub-system and assembly are presented. Finally, the results of the failure mode analysis are given.

2.4.1 Taxonomy

The WT system was divided into seven sub-systems, 45 assemblies (components) and 199 sub-assemblies. Figure 2.1 shows the sub-systems and assemblies, the sub-assemblies are not displayed at this point.

Subsystem	Assembly	Subsystem	Assembly
Power Module		**Auxiliary System**	
	Frequency Converter		Cooling system
	Generator		Electrical Protection and Safety
	Switch Gear		Human Safety
	Soft Starter		Hydraulic Group
	MV/LV Transformer		WTG Meteorological Station
	Power Feeder Cables		Lightning Protection
	Power Cabinet		Firefighting System
	Power Module Other		Cabinets
	Power Protection Unit		Service Crane
Rotor & Blades			Lift
	Pitch System		Grounding
	Other Blade Brake		Beacon/Lights
	Rotor		Power Supply Auxiliary Systems
	Blades		Electrical Auxiliary Cabling
	Hub	**Drive train**	
	Blade Bearing		Gearbox
Control & Communications			Main Bearing
	Sensors		Bearings
	Controller		Mechanical Brake
	Communication System		High Speed Shaft
	Emergency Control & Communication Series		Silent Blocks
	Data Aquisition System		Low Speed (Main) Shaft
Nacelle		**Structure**	
	Yaw System		Tower
	Nacelle Cover		Foundations
	Nacelle Bed plate		

Figure 2.1. Taxonomy used for the Failure Data Analysis, adapted from Paper (I).

As stated before, the taxonomy is based on the ReliaWind taxonomy, nevertheless, components of modern WTs have been included and several assemblies were re-arranged. The aim was to establish a taxonomy that can be applied to both, older and modern WT technologies. Whenever possible, new sub-assemblies were assigned to existing categories and only if the latter was not possible new categories were established. Additionally, this taxonomy gives higher priority to the functional similarity of components, rather than their physical location. This is of great advantage when e.g. analysing different WT technologies or the effects of environmental conditions on the failure behaviour of certain components, as e.g. mechanical, hydraulic and electric components behave differently.

Several components related to the operational control and safety as well as several sensors, including the condition monitoring system, have been added to the control & communication system. Unlike in the ReliaWind taxonomy, the sub-assemblies responsible for the control of the components, e.g. the converter control unit or the pitch and yaw control, were assigned to the control & communication system not to the component they are controlling. The same applies to the sub-assemblies of the cooling systems, which are classified as auxiliary systems, due to their functional proximity.

2.4.2 Failures and Downtime: Geared and Direct Drive Turbines

Table 2.4 shows the failures and downtime per wind turbine and year, as well as the yearly average downtime per failure for the three analysed categories: $G < 1\,MW$, $G \geq 1\,MW$ and DD turbines. In the further the term 'failure rate' will be used for failures per WT and per year.

Table 2.4. Total failure rates and downtimes per year and turbine for the different WT technologies, adapted from Paper (I).

WT technology	Failures per WT and year	Downtime per WT and year	Yearly avg. downtime per failure
$G < 1\,MW$	0.46	78.46 h	151.46 h
$G \geq 1\,MW$	0.52	44.51 h	112.67 h
DD	0.19	20.50 h	34.98 h

It can be seen that DD turbines showed significantly lower failure rates and

downtime than geared turbines. Comparing the failure rates presented in Table 2.4 to previous studies with similar objectives shown in Table 2.5, the here presented values are slightly lower. This is due to the failure definition used in this study, which focuses exclusively on internal component failures. Other WT stops due to grid problems, wind farm tests, vandalism and similar causes were not considered. Additionally, previous studies frequently consider failures due to 'unknown' or 'other' reasons, which were not included in this study. A full comparison of the detailed component related failure rates for each study including the findings of the present thesis, can be found in [28].

Table 2.5. Failure rates of several previous studies compared.

Initiative	Source	Failures per WT and year
CREW-Database	[30, 31]	7.167
Elforsk/Vindstat	[27, 33]	0.403
EPRI	[34]	10.195
Huandian	[38]	0.846
LWK	[19]	1.855
Muppandal	[40]	1.013
NEDO	[41]	0.171
University of Nanjing	[44]	46.856
VTT	[16, 27, 46]	1.45
Windstats Germany	[19, 47, 48]	1.796
Windstats Denmark	[19, 47, 48]	0.434
WMEP	[22, 23]	2.606

In Figure 2.2 the failure rates and downtime for each sub-system of the $G < 1\,MW$, $G \geq 1\,MW$ and DD turbines are shown. All values are displayed in percent of the respective total values presented in Table 2.4.

The three WT technologies ($G < 1\,MW$, $G \geq 1\,MW$ and DD) showed significant different shares of failure rates and downtime for the respective sub-systems. For $G < 1\,MW$, for instance, the highest contributor to the overall downtime is the drive train sub-system, whereas for $G \geq 1\,MW$ and DD WTs the power module had the highest share. Regarding the failure rates of $G < 1\,MW$ turbines, the sub-systems drive train and rotor & blades were the highest contributors to the overall failure rate. Whereas for $G \geq 1\,MW$ and DD the power module and control & communication system were the biggest contributors to the WT failure rates.

The DD turbines hardly suffered from drive train failures, except for some torque limiter problems. Furthermore, for this turbine technology rather low failure rates of the Rotor & Blades sub-system were registered, these however, were the second highest contributor to the overall downtime of this WT technology. Other structures (apart from the blades) did not significantly contribute to the overall failure rate or downtime of the three technologies. This can be due to the fact that over many years a considerable effort has been made to optimise and enhance the reliability of these structures.

Figure 2.3 shows the failure rates and downtimes for each assembly of the $G < 1\,\text{MW}$, $G \geq 1\,\text{MW}$ and DD turbines. Again the values are normalised to the overall values displayed in Table 2.4 for each technology. It can be highlighted that for both, $G < 1\,\text{MW}$ and $G \geq 1\,\text{MW}$, the gearbox was the highest contributor to the overall failure rate and downtime. For DD turbines, the highest failure rates were recorded for the controller, while the generator and blades contributed highly to their overall downtime. By virtue of the energy conversion principle of DD turbines and the 'missing' gearbox, these components seem to experience higher stresses, and thus, are more likely to fail.

The difference in failure behaviour between older and newer WT technologies can be seen when comparing the results of $G < 1\,\text{MW}$ and $G \geq 1\,\text{MW}$ turbines. Older technologies usually have lower rated capacities, are often not equipped with a pitch system and generally do not use complex control & communication systems. Hence, higher failure rates for these components were registered for the $G \geq 1\,\text{MW}$ technology.
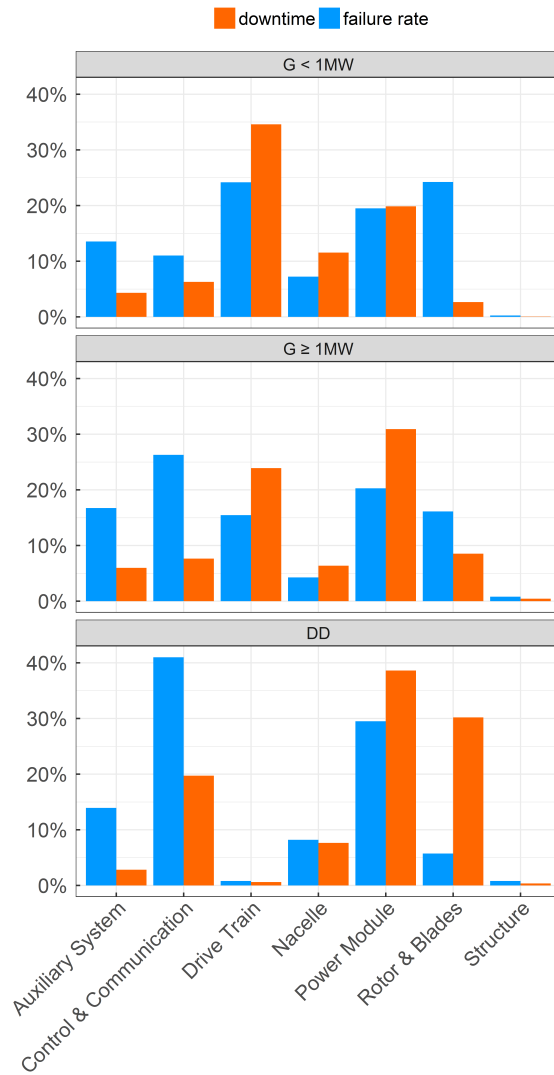
Figure 2.2. Normalised failure rates and downtimes for sub-systems of G < 1 MW, G ≥ 1 MW and DD turbines.
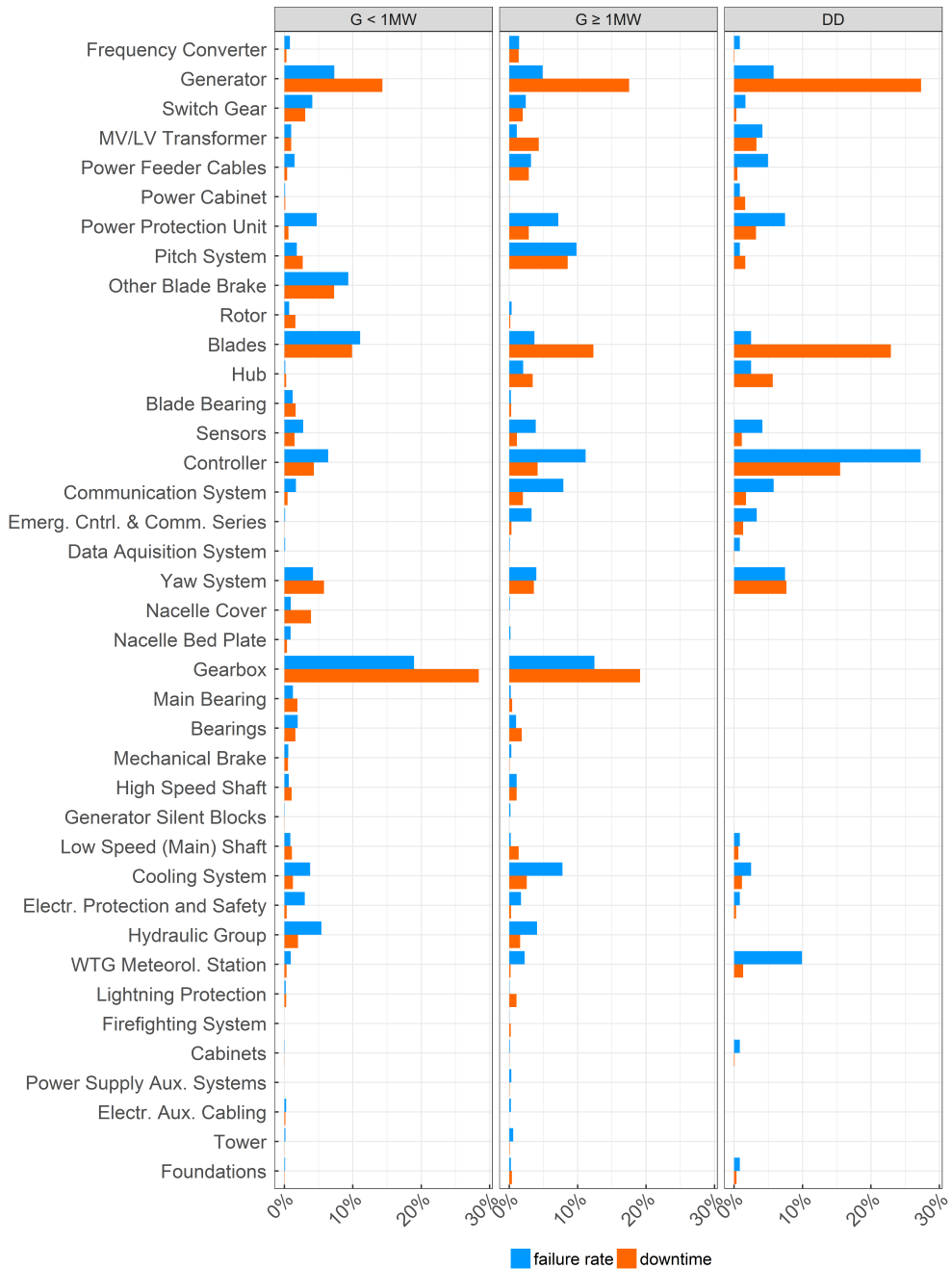
Figure 2.3. Normalised failure rates and downtimes for assemblies of geared turbines.

2.4.3 Failures and Downtime: Pitch and Stall regulated Turbines

The overall failure rates and downtime per WT and year for stall and pitch regulated turbines are displayed in Table 2.6. The data base consists of around 23% stall regulated and 77% pitch regulated machines.

Table 2.6. Total failure rates and downtimes per year and turbine for stall and pitch regulated WTs.

WT technology	Failures per WT and year	Downtime per WT and year
Stall regulated	0.47	86.8 h
Pitch regulated	0.51	54.7 h

In Figure 2.4 the normalised failure rates and downtime of the sub-systems of pitch and stall regulated turbines are compared. This is especially interesting as many stall regulated turbines are nowadays reaching the end of their planned life-time. Operators can either decide to keep on operating the latter (e.g. 'run to failure' - if legally permitted), or replacing them with newer wind turbine technologies (re-powering the wind farm). In either case, to facilitate maintenance planning, it is of great advantage to know which components cause the highest failure rates and downtime for the two technologies.

While pitch regulated turbines were mostly affected by power module and control & communication system failures, the stall regulated WTs showed higher relative failure rates for the drive train components as well as the rotor and blades. Moreover, the highest downtime in pitch regulated WTs was caused by the power module followed by drive train components. For stall regulated turbines, the biggest contributors to the turbine downtime were the drive train components and the rotor and blades.

Figure 2.5 shows the normalised failure rates and downtime for the assemblies of stall and pitch regulated WTs. The downtime and failure rates due to gearbox failure are the most remarkable difference between the two turbine technologies, which were significantly higher for stall regulated turbines. The generator was the second largest contributor to the overall downtime of pitch regulated turbines.
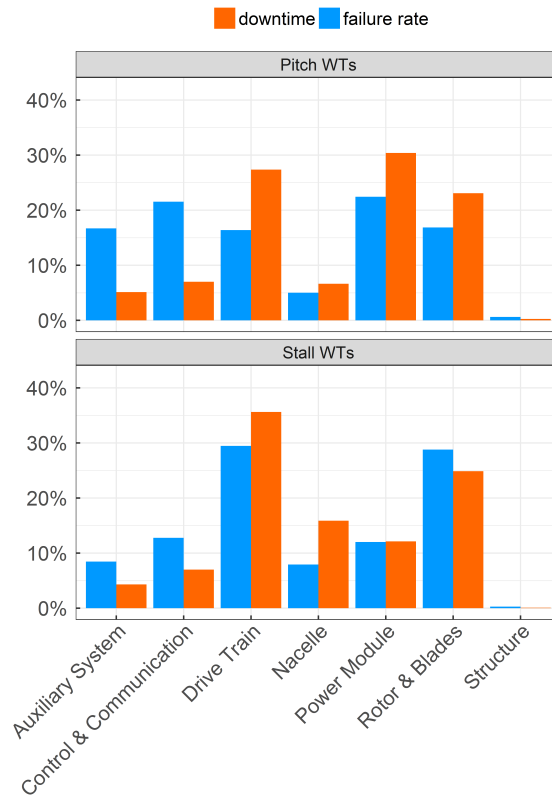
Figure 2.4. Normalised failure rates and downtimes for sub-systems of pitch and stall regulated turbines.
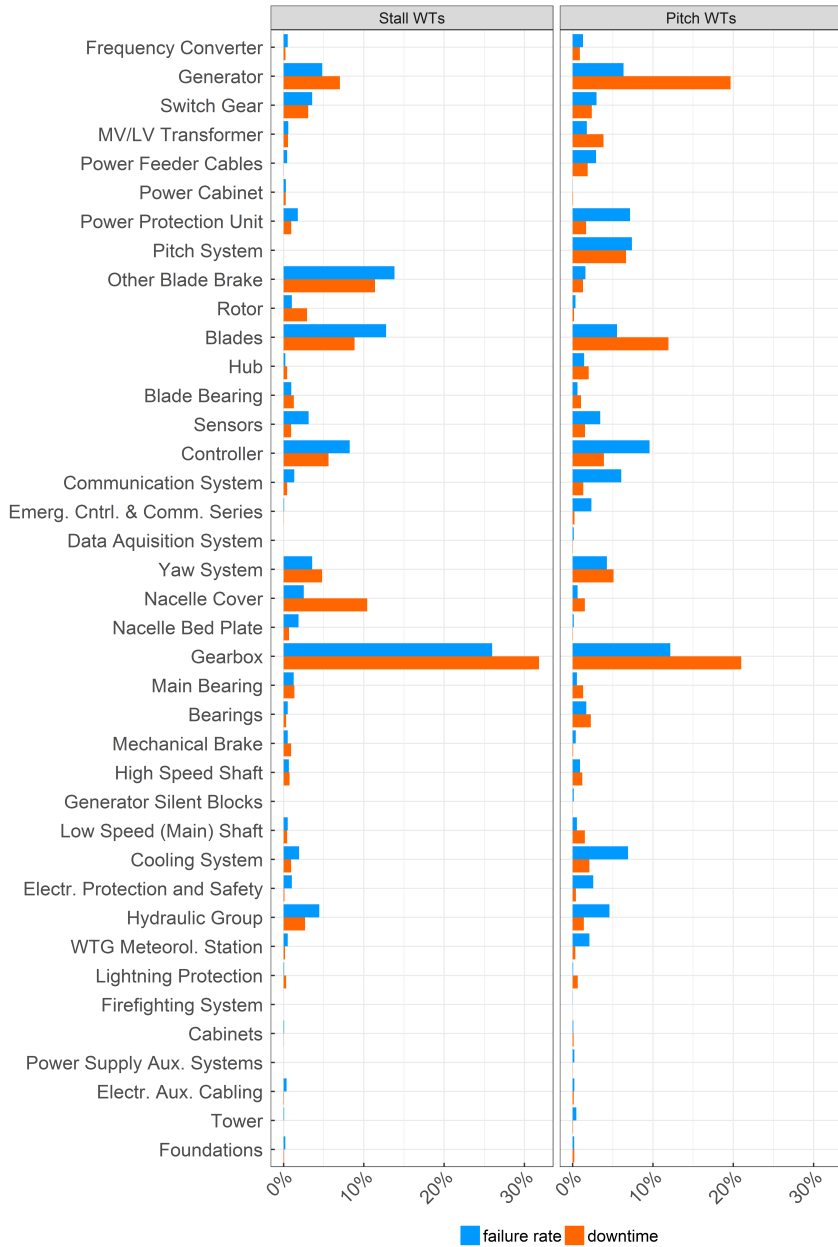
Figure 2.5. Normalised failure rates and downtimes for assemblies of pitch and stall regulated turbines.

2.4.4 Failure Modes

In this section the failure modes obtained for five main components: the blades, gearbox, generator, transformer and yaw-system will be presented. For this part of the analysis only a sub-set of the whole data base, containing 1855 turbines located in several Spanish wind farms, was used. This was due to the significantly better quality of the maintenance logbooks for these wind farms, which gave an insight on the exact cause of the failure. This subset of the data base contained 1432 $G < 1\,\mathrm{MW}$ and 423 $G \geq 1\,\mathrm{MW}$ WTs (972 pitch and 883 stall regulated WTs). Figures 2.6 to 2.7 show the percentages of each failure mode to the respective overall of the component failures. This is shown without distinguishing between the various WT technologies. The corresponding figures for $G \geq 1\,\mathrm{MW}$ and $G < 1\,\mathrm{MW}$ technologies, as well as stall and pitch regulated turbines can be found in Appendix B.
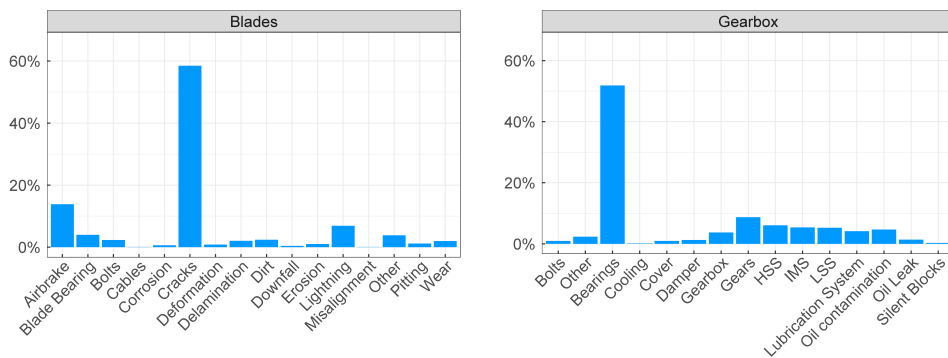
Figure 2.6. Failure modes for the blades and gearbox.

Blades: The blades mainly experienced repair due to propagating cracks on the blade surface. Additionally, for older WT technologies the airbrake had a significantly large share on the causes for blade repair. Damages due to lightning contributed with around 10% to the blade failures.

Gearbox: The gearbox was highly affected by gear bearing problems. This is consistent with literature, e.g. the National Renewable Energy Laboratory (NREL) gearbox failure data base [47, 53, 54]. Additionally, the gears, as well as the high

speed shaft (HSS), intermediate speed shaft (IMS) and low speed shaft (LSS) contributed often to gearbox problems.

Transformer: The transformer most frequently failed due to problems with the windings and bushings, as well as the transformer control system.

Yaw System: The gear rim and the yaw motor were the most frequent causes of yaw system failures.

Generator: The generator failures were mainly caused by generator bearing problems. Furthermore, the coils, phases, rotor and stator caused turbine stops due to generator problems.
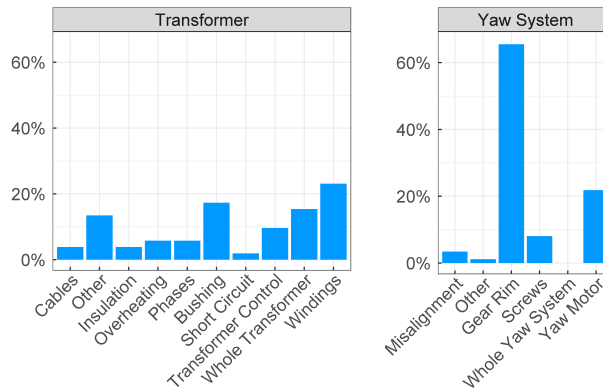
Figure 2.7. Failure modes for the transformer and yaw system.
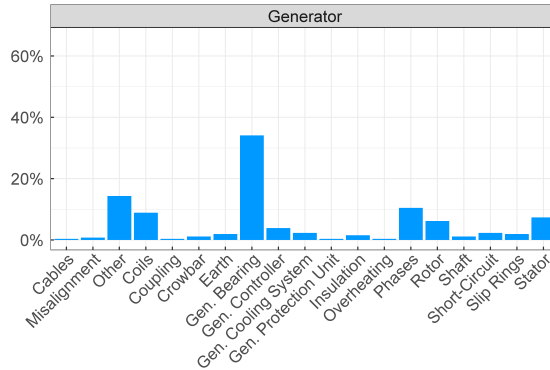
Figure 2.8. Failure modes for the generator.

2.5 Conclusions for the Failure Data Analysis

For this chapter a modern WT taxonomy was developed that divides the wind turbine system into seven sub-systems, 45 assemblies and 199 sub-assemblies. There were several existing taxonomies, however, in most cases they are not publicly available or based on old and outdated WT technologies. The herein presented taxonomy, extends existing ones (mainly the ReliaWind taxonomy) by including information on a higher number of turbines and combining different WT technologies (e.g. age, drive train concepts, power regulation, etc.). This taxonomy has been applied to extensive historical failure data and the failure rates and downtime of geared and direct drive WTs as well as stall and pitch regulated turbines were compared. It was shown that geared turbines have higher failure rates and downtime than direct drive ones. Furthermore, pitch regulated turbines showed more failures per WT and year than stall regulated ones, whereas the downtime per WT and year was lower for pitch regulated WT than for stall regulated machines. In general, the most failure prone components were the blades, controller, gearbox, generator, pitch and yaw system. However, the failure behaviour of the different components varies highly among the turbine technologies and they should be treated separately as shown in this chapter. Finally, the failure modes of five main components were analysed and the primary causes for their failures were identified. The information presented in this chapter is of high importance for further research in the field of O&M, and can serve as input for reliability and maintenance models. It, furthermore, builds the base for the subsequent chapters of this thesis.

3

Meteorological Conditions and Wind Turbine Failures

This chapter is concerned with understanding the meteorological conditions that frequently lead to WT component failures. For this, a machine learning framework was developed, which allows to analyse very big data bases and to obtain useful results regarding the failure behaviour of WT components under complex combinations of environmental conditions. These include ambient temperature, relative humidity and wind speed. The results of this study have been presented in Paper (II).

3.1 Introduction

Previous studies have shown that not only the turbine age, but also certain weather conditions affect the failure behaviour of WT components significantly.

The wind speed conditions leading to WT failures have been subject to several publications. One of the first approaches analysing the effects of wind speed on WT reliability was carried out in [55]. It was shown that with rising average daily wind speeds the failure rates of certain components increased as well. In [56] the effect of monthly averaged wind speed conditions on component failure rates was analysed, using the previously mentioned Danish reliability data base WindStats. They concluded that there is a certain annual periodicity in failure occurrences

due to seasonal variation in the wind speed conditions. However, no site specific meteorological data were analysed, as the wind speed records consisted of monthly average measurements for the entire country of Denmark. An analysis on the effects of wind speed on WT downtimes, was carried out in [57] taking into account energy- and time-based availability.

Other environmental variables, such as temperature and relative humidity, have also shown to affect the failure behaviour of WTs and their components, causing corrosion and other highly dangerous degradation processes. In [58] WT component failures were cross-correlated with average monthly maximum and mean wind speeds, maximum and minimum air temperatures as well as average daily relative humidity. Three wind farms, which are operating quite old 300 kW wind turbine technologies, were analysed using yearly and monthly average weather conditions taken from close-by located meteorological stations. But, neither short-time weather events nor the meteorological conditions at the time of failure occurrence were considered. They, nonetheless, suggested that subsequent work should ensure to include modern WTs as well as short-time weather variations. Wilkinson *et al.* [59] show the impact of temperature, turbulence intensity and wind speed on failure rates, downtimes and availability of WTs. The environmental data were taken from the WT's SCADA systems and the Modern-Era Retrospective Analysis for Research and Applications (MERRA) data base.

As demonstrated, existing studies on the meteorological conditions leading to WT failures are mostly based on monthly or yearly averaged data obtained from close-by located weather stations. Short-term changes in these parameters, as well as the conditions directly in the wind farm are not taken into account. Furthermore, most of these studies only consider WT system failures and do not distinguish between the different components.

According to Kuik *et al.* [7], the complex combinations of external conditions that are affecting WT reliability degradation are still not fully understood. They state that the critical external conditions for the main WT components need to be determined, enabling a more accurate reliability forecasting.

SCADA systems are nowadays among the standard equipment of wind turbines and provide a vast amount of information on many operational parameters, component conditions and external variables. These systems can give a realistic representation of the on site conditions and seem to be a very promising source

for understanding the impact of certain conditions on the failure behaviour of WT components, as well as for setting up data-driven reliability and failure models. However, in order to analyse such extensive data, sophisticated analysis techniques and high computational effort are needed.

Task 2 of this thesis is concerned with analysing the short term effects of environmental conditions on the failure behaviour of the main WT components. The objective is to extend earlier studies by implementing new techniques to catch especially short term changes and unexpected patterns in the recorded meteorological data. As this analysis will be based on significantly bigger data bases than previous studies, methodologies capable of handling big data and deriving meaningful results had to be developed.

In this chapter, a framework based on k-means clustering and association rule mining is introduced for processing big data bases with the aim of correlating the different input variables with failure events. This framework is applied to historical failure data of modern wind turbines and meteorological data taken from the WT SCADA systems and close-by located weather stations.

3.2 K-Means Clustering

Clustering is the problem of assigning a collection of objects to groups based on feature similarity among them. The various clustering algorithms can be roughly classified as partitional, hierarchical, density-based and grid based clustering, [60]. K-means clustering, [61–63], is one of the simplest and most widely used unsupervised learning algorithms for partitional clustering. The algorithm works iteratively by grouping the input data based on the distance between the single data points. The k-means clustering algorithm used in this study consists of two steps:

1. *Finding the optimal number of clusters:*
   At first the optimal number of clusters k, which suitably fits the specific data set, must be calculated. For non-hierarchical clustering procedures this is a fundamental input, which needs to be defined for each given data set prior to running the actual clustering algorithms. Several different methods have been developed to achieve this objective, including the elbow technique, silhouette and gap statistic methods. A technique presented in [64, 65], which is based on

the Euclidean distance and an agglomeration method presented in [66], will be used. This technique is capable of indicating the optimal number of clusters for a given data set by varying all possible combinations of the latter. For further information, see [64].

2. *Clustering:*

This step represents the actual k-means clustering algorithm, which is aiming at minimizing the within-class sum of squares for a given number of clusters, defined in step (1). The algorithm is divided into two sub-steps:

2.a. *Cluster assignment:* Based on the squared Euclidean distance, each data point is assigned to the cluster centroid, to which it is located closest to. This is achieved by assigning each data point x with:

$$\underset{c_j \in C}{\mathrm{argmin}} \left\{ d(c_j, x)^2 \right\} \quad , \tag{3.1}$$

to a cluster centroid c_j in a given set of cluster centroids C. Here, $d(c_j, x)$ denotes the squared Euclidean distance, defined as:

$$d(c_j, x) = \left(\sum_j (c_j - x)^2 \right)^{\frac{1}{2}} \quad . \tag{3.2}$$

2.b. *Moving the centroids:* Subsequently, the positions of the centroids are changed by taking the mean location of all assigned data points:

$$c_j = \frac{1}{|S_j|} \sum_{x_j \in S_j} x_j \quad , \tag{3.3}$$

with S_j being a set of data points assigned to the jth cluster.

Steps 2.a and 2.b are carried out alternately until the clustering algorithm converges. This is usually achieved when the sum of the distances is minimized and the data points do not change clusters any more.

3.3 Association Rule Mining (ARM)

Association Rule Mining (ARM) [67] is a method that is often used in data mining for finding correlations, frequent patterns, or interesting relationships among the

different variables of a data set. Several algorithms employing the concept of ARM have been developed, e.g. [68–70]. One of the most frequently used ARM algorithms is called apriori rule mining, [70], which will be used in this chapter. In the following, the concept of ARM will be explained in the context of the apriori rule mining algorithm.

Table 3.1. Explanation of the terminology used in ARM.

Name	Explanation
Group	all data related to one specific variable.
Item (i)	an attribute or member of a group.
Item set (I)	a collection of one or more items, e.g. $I = \{i_1, i_2, ...\}$
Frequent item set	an item set, which has a support value higher than the defined minimum support.
Transaction (TR)	a combination of items of different groups that will define at least one rule.
Association rule	an implication of the form *Item set X → Item set Y*.
Data Base (DB)	a set of transactions, e.g. $DB = \{TR_1, TR_2, ...\}$
Data Base sub-set (SubDB)	a part of a set of transactions.

In Table 3.1 the terminology used in this explanation of the ARM methodology is introduced. A set of items is given by $I = \{i_1, i_2, ...\}$ and contains several attributes (items). A data base (DB) is a set of transactions DB$= \{TR_1, TR_2, ...\}$, of which each transaction contains a number of items. A rule is defined as a combination of an antecedent and a consequent transaction as $X \rightarrow Y$, whereas $X, Y \subset I$,

The ARM procedure can be split into four steps:

1. Generating the item sets, transactions and data bases from the given input data.
2. Finding all combinations of item sets in the data that have a support value higher than the defined minimum support (frequent item sets).
3. Generating rules among the frequent item sets
4. Evaluating the strength of the generated rules.

In order to evaluate the strength of the discovered rules and to find the most interesting ones, three measures of interest are applied: *support*, *confidence* and *lift*. Support is defined as:

$$supp(X \rightarrow Y) = \frac{Freq(X \rightarrow Y)}{N} = Pr(X \wedge Y) \quad , \qquad (3.4)$$

and can be interpreted as the frequency of occurrence for each association rule in relation to the total number of transactions N. Here, X indicates an item set on the left hand side of the rule, Y is an item set on the right hand side of the rule and $Freq$ denotes the absolute frequency of appearance. The logical conjunction *AND* ($\wedge$) is used to indicate the relation between the item sets X and Y. Consequently, the fraction of transactions containing only X would be denoted as $supp(X) = Freq(x)/N = Pr(X)$.

The support is calculated for each rule and must be greater than a pre-defined value (between 0 and 1) for the rule to be considered in the evaluation process. The confidence metric is given by:

$$conf(X \rightarrow Y) = \frac{supp(X \rightarrow Y)}{supp(X)} = Pr(X \mid Y) \quad , \qquad (3.5)$$

and can be seen as a measure of how often a specific rule was true. It shall be underlined that there is a notation difference between the probability of an association $Pr\,(Y \wedge X)$ given as support, and the conditional probability $Pr(X \mid Y)$ representing the confidence metric. This is explained in detail in [71]. The lift metric:

$$lift(Y \rightarrow X) = \frac{supp(X \rightarrow Y)}{supp(X)supp(Y)} \quad , \qquad (3.6)$$

is often referred to as "interestingness" and shows how different rules are correlated within the DB. While, *lift* > 1 and *lift* < 1 indicate positive and negative correlation between the rules respectively, *lift* $= 1$ means that the rules are independent of each other. The last case allows the other metrics, support and confidence, to be more representative.

By comparing the support, confidence and lift values of all of the generated rules, their strengths are evaluated. If the same support values are obtained for two or more rules, the confidence value determines which one is more important. Likewise, if the same support and confidence values are calculated for more than one rule, the lift metric decides over which one is more important, [72].

For a better understanding, in the following this concept is illustrated in a simple example. Further explanation can be found e.g. in [72].

Example: In a supermarket, a set of products (items) is given by $I=\{$apples, bananas, cherries, milk, cheesecake$\}$. Table 3.2 exemplifies the transactions recorded based on the customers buying behaviour over a certain period. Each transaction ID indicates a set of products that were bought by one costumer, where 1 indicates that the item was bought, and 0 indicates that it was not bought. An example of

Table 3.2. Example for a data base DB containing transactions in ARM.

ID	apples	bananas	cherries	milk	cheesecake
1	1	1	0	0	1
2	0	0	1	1	0
3	1	1	0	0	1
4	0	1	0	1	1
5	1	0	0	0	1
6	1	0	0	0	1

a rule derived from this data base would be $\{apples, bananas\} \rightarrow \{cheesecake\}$, which means that if apples and bananas were bought, cheesecake was bought as well.

The support of this rule is:

$$supp\{\{apples, bananas\} \rightarrow \{cheesecake\}\} = 2/6 = 0.33 \quad , \tag{3.7}$$

as this rule appears twice in the given DB containing six transactions. The confidence of the rule is then obtained with:

$$conf\{\{apples, bananas\} \rightarrow \{cheesecake\}\} = 0.33/0.33 = 1 \quad , \tag{3.8}$$

which indicates that in 100% of the cases when apples and bananas are bought, cheesecake is also bought. Finally, the lift results in:

$$lift\{\{apples, bananas\} \rightarrow \{cheesecake\}\} = 0.33/(0.33 \cdot 0.83) = 1.2 \quad . \tag{3.9}$$

3.4 Data-Driven Learning Framework

Based on the previously introduced concepts of k-means clustering and ARM, a framework was developed that is capable of quantifying the impact of meteorological conditions on WT failures by combining environmental and historical failure data. The framework was presented in Paper (II) and its procedure is displayed in the flowchart in Figure 3.1. It consists of four steps: (1) data pre-processing, (2) data processing, (3) unsupervised rule mining and, finally, (4) ranking and interpretation of the rules. These steps will be discussed in detail in the following. At first, however, the data used in this chapter are introduced.

1) Pre processing

Failure , SCADA, RNCEP

Cleaning

··

2) Data Processing

Supervised Labelling

Unsupervised K Means Clustering

Multi D Input

1 D Input

··

3) Unsupervised Rule Mining

Apriori Rule Mining

Apriori Rule Mining

Apriori Rule Mining

Rules of Literature Based Process

Rules of Multi D

Rules of 1 D

··

4) Rule Ranking and Interpretaion

Combination of Factors

%

Temp → Failure ← Power

Humidity — WS

Process Duration
Number of generated Rules
Significance of Rules
Evaluation of Rules

Figure 3.1. Flowchart of the developed framework and its sub-processes.

3.4.1 Historical Failures, Meteorological and Operational Data

The presented framework is applied to a case study including historical failures as well as meteorological and operational data taken from SCADA systems and reanalysis data archives. Table 3.3 summarises all the input parameters for the data-driven learning framework. The sources for each of the used parameters are discussed below.

Table 3.3. Summary of the input parameters obtained from the various data sources.

Name	Explanation
Wind speed (WS)	Series of 10-minute mean values 80 minutes before failure (WS_{80}, WS_{70}, ... , WS_{10}), and one wind speed measurement taken during last 10 minutes before failure occurrence (WS_{atF}).
Power Production (P)	Two measurements: the power production at failure time P_{atF} and the power production before failure P_{bF} (the previous 10-minute measurement). Both values are given in relation to the measured monthly mean power curve, as explained in the text.
Relative Humidity (RH)	Eleven measurements are considered, including the hourly RH values of 10 hours ahead of failure, as well as at the time of failure. These are indicated by RH_{atF}, RH_1, RH_2,...,RH_{10}.
Ambient Temperature (Temp)	The monthly mean $Temp_m$, maximum $Temp_{max}$ and minimum temperatures $Temp_{min}$ recorded during 30 days before failure, as well as the temperature at the time of failure $Temp_{atF}$.
Maintenance Strategy	The availability of the maintenance personnel affects the WT downtime. It was classified considering typical working hours of close-by located personnel, into day shift (8:00 to 18:00 o'clock) and night shift (the remaining hours).
Downtime/Severity (S)	The downtime or un-availability caused by each failure is an indicator of its severity, and is obtained from the historical failure data and cross-checked with the SCADA data.

Failure Data. The historical failure data used in this chapter represents a sub-set of the data presented in Chapter 2. It is comprised of 146 failure events obtained for 448 WTs with rated capacities between 660 kW and 2000 kW. All turbines were

produced by the same WT manufacturer and are three bladed, horizontal axis turbines, equipped with DFIGs. Five WT components are analysed: the gearbox, generator, frequency converter, pitch and yaw system. These were identified in the previous chapter as some of the most important components, regarding their contribution to the overall failure rates and downtime of the whole wind turbine system. The failure data contain around 30 failures for each component and the exact time and duration of the failure events. Table 3.4 summarises the failure data used in this chapter and compares them to studies with similar objectives.

Table 3.4. Historical failure data used in this chapter compared to previous studies.

	This study	Wilson et al. [73]	Tavner et al. [58]
Number of WTs	448	140	32
Total WT Years	972	381.7	130
WT Technology	47 WTs with 0.66 MW,	Variable speed,	Enercon
	289 WTs with 0.85 MW,	pitch regulated,	E32/33 (300 kW)
		2.3 MW	
	112 WTs with 2 MW		

Meteorological and Operational Data. The wind velocity, the downtime as well as the active power production were obtained directly from the failed turbines' SCADA systems. The measurements are available as 10-minute average values. In order to ensure properly functioning SCADA systems and correct data collections, the measurements were compared to those obtained at close-by located turbines.

The active power is used to calculate the relative performance of the WTs, which is considered a measure of efficiency regarding the turbines power production. For this the observed active power SCADA measurements P_o are divided by the monthly mean value P_m taken at the same wind speed and the specific WF location within the month of failure occurrence. This results in a measure for the turbines' relative performance, or efficiency: $P_e = P_o \ / \ P_m$ with values between 0 and 1. The manufacturers power curves are not site and season specific and cannot be used for this purpose.

The humidity and temperature measurements at the wind farm location were obtained from the NCEP/NCAR reanalysis data set [74, 75]. This is a frequently updated meteorological data base developed by the National Centers for Envir-

onmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), exploiting observations and numerical weather predictions. Again, to ensure good data quality, the downloaded reanalysis data were compared to real data taken at three wind farm sites and very acceptable results were obtained (mean absolute error under 3 %) for both variables, humidity and temperature.

Expert Judgement. As the availability of maintenance personnel has an effect on the component's repair time, the opinions of two big European wind farm operators were consulted. This helped to define the maintenance strategy and the classification of failure severity according to the WT downtime caused by each failure event.

3.4.2 Data Pre-Processing

Firstly, the historical failure data were cleaned according to the failure definition and taxonomy used in this thesis. The meteorological conditions were obtained from the WT SCADA systems as well as the NCEP/NCAR reanalysis data set. As the framework is very sensitive to missing observations ("NAs" in the data), only wind farms with complete SCADA histories over the whole observation period were chosen.

3.4.3 Data Processing

The objective of the data processing is to group the data into different categories regarding their properties. For this, two distinct approaches are used in parallel: (a) unsupervised k-means clustering and (b) supervised data labelling. These two approaches will lead to different results that can be used for various analysis objectives.

Supervised Labelling. The supervised labelling is a manual process that is based on expert judgements, literature or experience in the field, which are often a result of extensive experiments. This technique is used to assign thresholds and labels manually to each input variable. It requires profound knowledge about the behaviour of the respective variable and can in some cases be very time consuming. The labelling process will here be exemplified for wind speed and relative humidity.

For classifying the wind speed the typical cut-in and cut-out wind speeds for the respective WT technology were chosen as lower and upper limits. The remaining wind velocities were divided into calm, low, high and stormy wind conditions.

Labelling the relative humidity (RH) conditions required an extensive literature research. As indicated in [76–78], this variable can be labelled according to the effect it has on metallic components – the resulting corrosiveness. However, one has to bear in mind that this is only valid for one certain type of steel and as WTs consist of a high variety of different materials, very expensive experiments would be needed to quantify the RH in terms of its corrosiveness for the entire WT system. Nevertheless, it was considered a good indicator and the values for RH will be grouped into categories reaching from 'dry air' to 'highly corrosive' and 'precipitation'.

As demonstrated, the manual labelling approach has certain limitations, which will be discussed in detail in the results section below.

Unsupervised Clustering. While the supervised labelling is based on experiments, expert judgements and literature findings, unsupervised k-means clustering relies exclusively on the input data. As the layout of the input data highly affects the outcome of the k-means clustering method, two distinct layouts of the input data are chosen. Both lead to different results for the entire framework and will be compared. Multi-Dimensional (Multi-D) input is the standard format used in k-means clustering. This layout considers all time steps of the input parameters as one observation before a specific failure. In addition to the Multi-D approach, a One-Dimensional (1-D) input is used. This layout does not distinguish between the failures but rather separates the input parameters based on the type of input variable. This results in a one dimensional input vector for each input variable, similar to the results of the supervised labelling, however, without prior knowledge. In literature both input layouts have been found, although multi-dimensional data are the most frequently used input format.

In Figure 3.2, the input and output formats for Multi-D and 1-D are explained for wind speed measurements taken from a specific WT before failure. For the Multi-D input, all wind speed observations before a certain failure are gathered and an overall cluster is defined. Thus, this cluster represents the general wind speed conditions over the observation period.
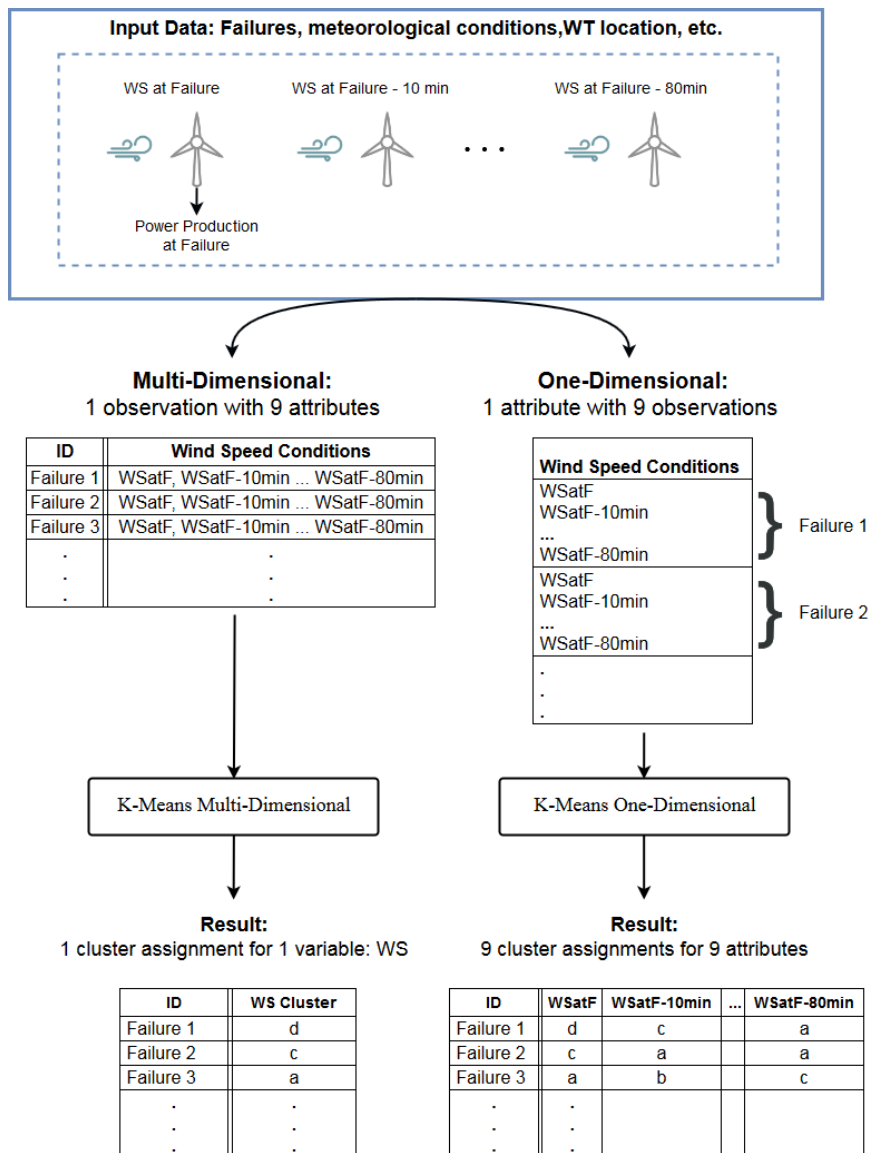
Figure 3.2. Example for Multi-D and 1-D input and results.

The one-dimensional inputs consider the wind speed records with 9 observations for each failure as one array for all analysed turbines. Hence, the measurements at each time step are assigned to an individual cluster, which allows to analyse how the environmental conditions change over time before the failure occurs.

3.4.4 Rule Mining

The output of each classification approach serves as input to an apriori rule mining algorithm, which derives rules from the input data by interconnecting logically the environmental parameters and the historical component failures. The resulting rules, not only determine the weather variables that have the highest impact on the WT components' failure behaviour; but also allow to evaluate for which input parameters which machine learning technique, and input format are more appropriate. The framework has been developed, so that it can be applied to a variety of problems and parameters can be changed easily.

Set-up of the Apriori Algorithm: The minimal support value in the present application was set to 0.3, as this has shown to be the best compromise between the size of the input data and the amount and quality of the derived rules. Furthermore, the pre-defined minimum of the confidence value was set to 0.5. Additionally, the minimum rule length needed to be defined, which controls how many items have to be included in a rule. In this study, the minimal rule length was set to 2, meaning that each rule needs to consist at least of one input and one output item, in order for it to be considered.

By dividing DB into several SubDBs according to the failed component, the influence of the meteorological condition on the failure behaviour can be analysed for each component separately. By sub-setting the data base however, the confidence and lift metric of all rules obtained for one SubDB are equal to 1. In that manner the support metric can be used as a single indicator for the importance of a rule. The evaluation of the resulting apriori rules shall be explained in the following.

3.4.5 Evaluation of the Apriori Rules

The post-processing or evaluation procedure consists of (1) finding the frequency of appearance of an item set in DB (or SubDB); and (2) calculating and comparing the evaluation metrics.

As stated before, however, the evaluation process differs slightly depending whether the whole DB is used or several SubDBs are analysed. When analysing the whole DB all three metrics (support, confidence and lift) need to be compared. Whereas, the analysis of one SubDB at a time gives more weight to the support

metric, by setting confidence and lift to 1. In this manner a fast and very representative evaluation is granted, due to the fact that the support values can directly be interpreted as percentage of appearance. This is exemplified in Tables 3.5 and 3.6 for the cases of using the whole DB or a SubDB containing only gearbox failures as basis for the calculation of the evaluation metrics. The frequencies of appearance of each item set as well as the total number of transactions N within the DB or SubDB are shown in Table 3.5. These serve as input to the calculations of the evaluation metrics for gearbox failures in Table 3.6, where it is shown that when using the SubDB as reference input, the *confidence* and *lift* metrics automatically become equal to 1 and the value for support becomes larger.

Table 3.5. Counting appearances in DB (N=146) and SubDB$_{GB}$ (N=30).

ID	Input	N	Item-set	Freq.
(1)	DB	146	$Freq(Failure = GB)$	30
(2)	DB	146	$Freq(WS = 3 \wedge T = 1)$	21
(3)	DB	146	$Freq(WS = 3 \wedge T = 1 \wedge Failure = GB)$	9
(4)	SubDB$_{GB}$	30	$Freq(Failure = GB)$	30
(5)	SubDB$_{GB}$	30	$Freq(WS = 3 \wedge T = 1)$	9
(6)	SubDB$_{GB}$	30	$Freq(WS = 3 \wedge T = 1 \wedge Failure = GB)$	9

Table 3.6. Calculating metrics for DB (N=146) and SubDB$_{GB}$ (N=30) metrics.

Input	Item	Supp.	Conf.	Lift
DB	$Pr(Failure = GB \wedge WS = 3 \wedge T = 1) = 9/146$	0.06	–	–
DB	$Pr((Failure = GB \mid (WS = 3 \wedge T = 1)) = \frac{(9/146)}{(21/146)}$	–	0.43	–
DB	$\dfrac{Pr(Failure=GB \mid (T=3 \wedge T=1))=\frac{(9/146)}{(21/146)}}{Pr(Failure=GB)=(30/146)}$	–	–	2.09
SubDB$_{GB}$	$Pr(Failure = GB \wedge WS = 3 \wedge T = 1) = 9/30$	0.3	–	–
SubDB$_{GB}$	$Pr(Failure = GB \mid (WS = 3 \wedge T = 1)) = \frac{(9/30)}{(9/30)}$	–	1	–
SubDB$_{GB}$	$\dfrac{Pr((Failure=GB \mid (WS=3 \wedge T=1))=\frac{(9/30)}{(9/30)}}{Pr(Failure=GB)=(30/30)}$	–	–	1

3.5 Results of the Data-Driven Learning Framework

The presented data-driven learning framework was applied to the case study data set introduced in Section 3.4.1. In the following, the results for each of the framework's sub-processes will be presented and discussed separately.

3.5.1 Performance of the Data Processing

In Table 3.7 the results for the three grouping techniques are presented. The table shows the thresholds for the manual labelling process, the minima and maxima for the clusters assigned with 1-D clustering and the cluster centroids for the Multi-D clustering. It is shown that in some cases the number of groups established for the several variables differs highly depending on the used classification technique. For example for the temperature measurements, nine categories were defined manually, while the k-means algorithms only defined three (for Multi-D) and four clusters (for 1-D). Hence, the labels might give too much information, which can impede a simple interpretation of the results. Contrarily, the downtime was split into two categories during the manual labelling: major ($\geq$ 48 hours) and minor ($<$ 48 hours) interventions, whilst three clusters were assigned by both clustering algorithms. Here, k-means can provide more information for the subsequent data processing. The power production efficiency was grouped into three categories by labelling and 1-D clustering and the Multi-D clustering algorithm only assigned two. Compared to manual labelling, the clustering techniques slightly reduced the number of categories that were assigned to wind speed and RH.

The objective of comparing manual labelling and unsupervised clustering is to find the best method to process the input data for further use in the apriori ruling algorithms. K-means clustering was expected to show better results, as the input parameters are all clustered according to their appearance in the data. However, for parameters that have previously been analysed thoroughly and their effect on the components are well known, the manual labelling can provide a better solution. Nonetheless, if there is little to no information on the effect of the respective variable on the components' failure behaviour, manual labelling is likely to introduce a bias.

Besides the variables shown in Table 3.7, also the availability of the maintenance personnel was taken into account and classified manually into day- and night-shift.

During the night-shift a detection of a faulty component is less likely and the availability of the maintenance personnel is limited.

Table 3.7. Results for labelling and clustering with one-dimensional (1-D) and multi-dimensional (Multi-D) input.

Minimum and Maximum Values for the Labelling

Label	WS [m s^{-1}]	Label	RH [%]	Label	P [–]	Label	Downtime [h]	Label	Temp [°C]
calm	< 3	dry air	20 - 40	consumption	$P_o < 0$	Minor	< 48	freezing	−10 - 0
low	3 - 10	moist air	40 - 60	not efficient (ne)	$0 \leq P_o < P_m$	Major	≥ 48	very cold	0 - 5
high	10 - 26	corrosive	60 - 80	efficient (e)	$P_o \geq P_m$			cold	5 - 10
storm	> 26	highly corr.	80 - 98					cool	10 - 15
		precipitation	100					mild	15 - 20
								room temp.	20 - 25
								warm	25 - 30
								hot	30 - 35
								very hot	35 - 40

Minimum and Maximum Values for the Clusters obtained for the 1-D Input

Cluster	WS [m s^{-1}]	Cluster	RH [%]	Cluster	P [–]	Cluster	Downtime [h]	Cluster	Temp [°C]
1	0 - 6.568	1	32.46 - 62.3	1	0 - 0.27	1	27.5 - 144	1	3.34 - 10.65
2	6.62 - 12.98	2	62.84 - 79.19	2	0.3 - 0.7509	2	151.17 - 360	2	10.68 - 14.85
3	13 - 27.08	3	79.2 - 98.46	3	0.759 - 1	3	402 - 980.66	3	14.96 - 20.15
								4	20.44 - 30.13

Centroids of the Clusters obtained for Multi-D Input[1]

Cluster	WS [m s^{-1}]	Cluster	RH [%]	Cluster	P [–]	Cluster	Downtime [h]	Cluster	Temp [°C]
1	3.97	1	64.63	1	0.12	1	72.68	1	10.24
2	9.36	2	80.78	2	0.53	2	226.84	2	17.03
3	15.98					3	563.98	3	22.69

[1] As the multi-dimensional input also results in multi-dimensional clusters, the minima and maxima cannot be displayed. Thus, the cluster centroids are being displayed instead, giving an idea about the location of the clusters.

In Table 3.8 the performance of the three grouping techniques is compared. It can be seen that the techniques performed differently in terms of time consumption, possible results and post-processing effort.

Table 3.8. Characteristics of the three processing techniques and framework performance.

	Labelling	1-D Clustering	Multi-D Clustering
DATA ANALYSIS			
Time consumption data pre-processing	high	medium	low
Time consumption data processing	low	high	medium
Expert judgement needed	✓	✗	✗
POSSIBLE RESULTS			
Behaviour (variations) over time	✓	✓	✗
Punctual analysis	✓	✓	✗
General conditions	✗	✗	✓
POST-PROCESSING			
Interpretation of results	medium	dificult	easy
Number of rules per component	very high	very high	low
Number of obtained rules in the case study	15783	12814	127

Data Analysis: Supervised labelling was the most time consuming and complex pre-processing. This is due to the fact that it involves a high amount of manual work and relies heavily on experience and expert judgements, which are often difficult to obtain. Nevertheless, after having defined the labels for each variable, the data processing is significantly faster than for the other two methods. For 1-D clustering the effort for the data pre-processing was slightly lower than for the manual labelling. However, the time consumption related to the processing was the highest compared to the other techniques, as the data have to be converted to a one-dimensional array and the computational effort increases with the one-dimensional input. Multi-D clustering does not require any time consuming pre-processing and processing.

Possible Results: The three grouping techniques can result in fairly different results. Labelling and 1-D clustering perform in a similar manner. Both can be used to analyse a specific time step (e.g. $Temp_{atF}$, RH_4, etc.) or to detect for example variations by considering subsequent time steps throughout the observation period. Multi-D clustering is especially useful for looking into the general conditions over the whole observation period. Due to the fact that the latter method threats all

the input data before a specific failure as a whole and it is not possible to analyse a specific time step nor alternations over time.

Post-Processing: The effort for post-processing is strongly related to the amount of obtained rules. As shown in Table 3.8, the manual labelling resulted in the highest number of obtained rules, which however, were quite easy to interpret. The 1-D clustering also resulted in a fairly large amount of rules, yet, in contrary to labelling, a high effort has to be made to interpret the latter. Multi-D clustering showed the lowest amount of generated rules and the interpretation of these results is straightforward and fast. Regarding the evaluation of quality of the obtained results, one has to keep in mind the restrictions of each technique, discussed in Section 3.5.1. Multi-D does not need a time consuming data pre-processing and processing as 1-D clustering and labelling do, thus, it is a very efficient technique for quick analyses, which do not require a very high accuracy. Nevertheless, k-means clustering only takes into account data that are fed to the algorithm and these inputs need to be carefully chosen in order to assure consistent results. Contrarily, labelling does not rely on the data, but the results in some cases can be heavily biased.

3.5.2 Results for the whole Data Base (DB)

As stated before, the DB can either be analysed as a whole or split into component related SubDBs. In this section the resulting rules for the application of the framework to the whole DB are briefly discussed. These have been evaluated using the support metric as primary indicator for the strength of the rules, followed by confidence and lift. When analysing the whole DB the minimum support value needed to be lowered to 0.03 as this set contained much more data than each SubDB.

Multi-D clustering was used to analyse the general conditions throughout the whole observation period before each failure. With labelling and 1-D clustering the conditions at one specific point in time, e.g. WS_{20} or WS_{atF}, were obtained. Furthermore, the variations of a parameter over time (alternations) were investigated using these two techniques.

Figure 3.3 shows a grouped matrix including the rules obtained for manual labelling using the whole DB. The corresponding rule matrices for 1-D and Multi-D

clustering can be found in Appendix C. It can be seen that the interpretation of the latter is quite difficult due to the high amount of items represented in each rule, and not all of these items can be presented in the graph. The most important items per set of rules are displayed, these were the ones that were found most frequently. If more items were found, these are displayed without specifying their names (e.g. '+ 2 items') as they can be different within each set of rules.
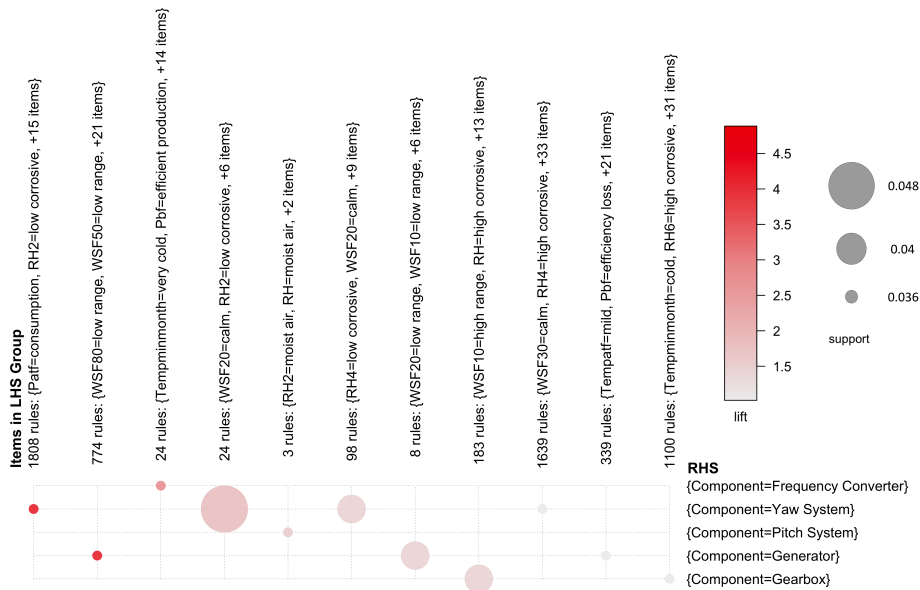
Figure 3.3. Grouped matrix for labelling rules with a minimum support of 0.03. On the right hand side (RHS) of the rules the components are displayed; on the left hand side (LHS) the input items are shown.

The support value is given by the size of the circles, the lift values are defined by the level of transparency of the colour. Given the example presented in Figure 3.3 the following conclusions can be drawn:

- *Frequency converter:* The rules show that this component often fails during months with very low temperatures. Furthermore, there is no sign of inefficient power production right before the failure event. Rather low support and medium lift values were recorded for these rules.

- *Yaw system:* These failures frequently occur under calm wind speed (WS) conditions 20 minutes before failure and low corrosive RH values 2 hours before

the failure event. In two further sets of rules, the RH conditions 4 hours before are once classified as highly corrosive and once as low corrosive. Hence, there is no clear rule for this variable 4 hours before failure. Additionally, the power production efficiency at failure is extremely low ('consumption').

- *Pitch system:* This component seems to be highly influenced by moist air at failure as well as 2 hours before the failure occurrence. the rules show relatively low support and weak lift values.

- *Generator:* The generator failures occurred frequently under low wind speed conditions 10 minutes and 20 minutes before failure. Very week rules were obtained indicating lower temperatures and efficiency loss before the failure event.

- *Gearbox:* The rules for gearbox failures show that this component frequently fails under high wind speed conditions right before the failure and high corrosive relative humidity. These rules showed medium range support but high lift values. Another weaker set of rules indicates that low monthly minimum temperature and highly corrosive RH conditions (RH_6) contribute to higher numbers of gearbox failures.

The figures in Appendix C can be interpreted in a similar way. Nevertheless, it was shown that the results obtained from the whole DB are rather limited and difficult to interpret.

3.5.3 Results obtained for the SubDB for each Component

Sub-setting DB, however, can simplify the interpretation of the rules, as demonstrated in Section 3.4.5. The results for the framework applied to the respective SubDB are shown in Table 3.9. This shows the most frequently obtained rule for each technique in percentage of their appearance in each SubDB. For easier interpretation of the 1-D and Multi-D rules, the maximum number of clusters for each variable are shown in brackets.

These results can be interpreted in three different ways:

1. *Technique-based:* Comparing the results row-wise according to the technique that was used, i.e. labelling, 1-D and Multi-D clustering gives an insight on what kind of results can be obtained from each technique.

2. *Component/Item-based:* For comparing the influence of the various environmental variables on the different components, the table can be analysed column-wise. With this, the general environmental conditions before failure (Multi-D) as well as the condition at a specific time step (Labelling and 1-D) that most frequently appeared in the respective SubDB, can be analysed.

3. *Combinations:* for analysing the effects of combinations of several environmental variables before the respective component failures, all rows related to the this component can be combined.

At this point it shall be mentioned, that further environmental conditions, as for example ice and snow, can also cause certain components to fail. These are not included directly in the analysis, but enter indirectly considering the temperatures and seasons.

The results that can be obtained from Table 3.9 will be discussed in the following sections.

Table 3.9. Results for the three approaches - rules with highest support value (*supp.*).

		Wind Speed	supp.	Relative Humidity	supp.	Power	supp.	Severity	supp.	Temperature	supp.
Gearbox	Label	$WS_{atF} = $ low	46%	$RH_{atF}=$ highly corr.	60%	$P_{atF}=$ne $P_{bF}=$ne	57%-46%	Major	96%	$Temp_m = $ cool	60%
	1-D	$WS_{60} = 2$ (max: 3)	50%	$RH_2=3$ (max: 3)	63%	$P_{atF}=1$ (max: 3) $P_{bF}=3$ (max: 3)	43%-67%	1	46%	$Temp_{atF} = 1$ (max: 4)	53%
	Multi-D	$WS = 2$ (max: 3)	40%	$RH = 2$ (max: 2)	70%	$P = 2$ (max: 2)	67%	1	46%	$Temp = 1$ (max: 3)	66%
Generator	Label	$WS_{atF} = $ low	53%	$RH_8 = $ highly corr.	70%	$P_{atF}=$ne $P_{bF}=$e	67%-50%	Major	96%	$Temp_{min} = $ cold	46%
	1-D	$WS_{30} = 2$ (max: 3)	50%	$RH_8 = 3$ (max: 3)	70%	$P_{atF}=1$ (max: 3) $P_{bF}=3$ (max: 3)	40%-80%	2	50%	$Temp_{atF} = 2$ (max: 4)	30%
	Multi-D	$WS = 2$ (max: 3)	46%	$RH = 2$ (max: 2)	73%	$P = 2$ (max: 2)	90%	2	50%	$Temp = 1$ (max: 3)	46%
Converter	Label	$WS_{30} = $ low	55%	$RH_{atF} = $ highly corr.	48%	$P_{atF}=$ne $P_{bF}=$e	78%-48%	Major	81%	$Temp_{max} = $ mild	48%
	1-D	$WS_{atF} = 1$ (max: 3)	52%	$RH_{atF} = 3$ (max: 3)	48%	$P_{atF}=2$ (max: 3) $P_{bF}=3$ (max: 3)	48%-81%	1	81%	$Temp_{atF} = 1$ (max: 4)	63%
	Multi-D	$WS = 1$ (max: 3)	51%	$RH = 2$ (max: 2)	55%	$P = 2$ (max: 2)	85%	1	81%	$Temp = 1$ (max: 3)	51%
Pitch	Label	$WS_{atF} = $ high	58%	$RH_8 = $ highly corr.	62%	$P_{atF}=$ne $P_{bF}=$e	79%-35%	Major	86%	$Temp_m = $ cool	37.9%
	1-D	$WS_{20} = 2$ (max: 3)	51.7%	$RH_8 = 3$ (max: 3)	62.1%	$P_{atF}=1$ (max: 3) $P_{bF}=3$ (max: 3)	48%-69%	1	48.3%	$Temp_{atF} = 4$ (max: 4)	34%
	Multi-D	$WS = 2$ (max: 3)	41%	$RH = 2$ (max: 2)	58.6%	$P = 2$ (max: 2)	69%	1	48.3%	$Temp = 1$ (max: 3)	48.3%
Yaw	Label	$WS_{atF} = $ low	63%	$RH_2 = $ low corr.	60%	$P_{atF}=$ne $P_{bF}=$e	53%-43%	Major	70%	$Temp_{max} = $ mild	47%
	1-D	$WS_{20} = 1$ (max: 3)	70%	$RH_{atF}=2$ (max: 3) $RH_{10}=3$ (max: 3)	56% 56%	$P_{atF}=2$ (max: 3) $P_{bF}=3$ (max: 3)	33%-63%	1	90%	$Temp_{atF} = 1$ (max: 4)	36%
	Multi-D	$WS = 1$ (max: 3)	60%	$RH = 2$ (max: 2)	70%	$P = 2$ (max: 2)	70%	1	90%	$Temp = 1$ (max: 3)	53%

3.5.4 Environmental Conditions at Time of Failure - Labelling and 1-D Clustering

In the following the meteorological conditions at the exact time of failure are discussed. These were obtained using labelling and 1-D k-means clustering as input to the apriori rule mining algorithm.
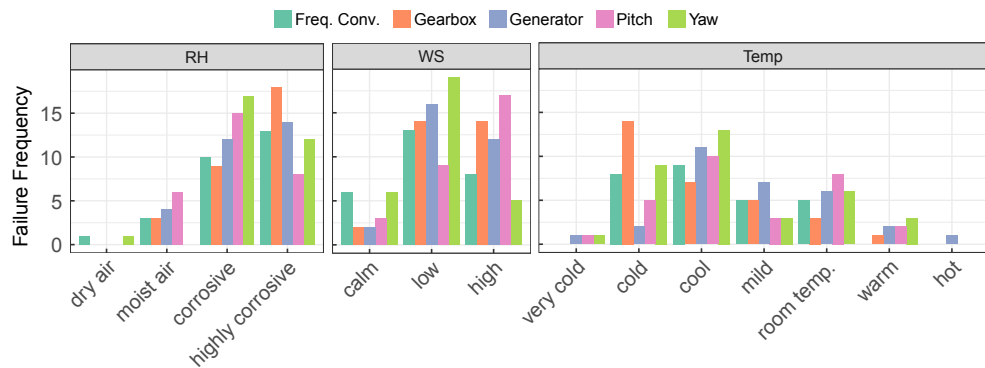
Figure 3.4. Conditions at the time of failure, obtained for labelling.

Figure 3.4 displays the number of failures registered for each label assigned to relative humidity, wind speed and temperature at the time of failure occurrence. It is shown that the relative humidity at the time of failure occurrence is classified for all components as corrosive or highly corrosive. Especially, the gearbox, generator and frequency converter showed the highest failure frequencies under these conditions. As higher wind speeds induce higher loads on most of the wind turbine components, a correlation between high wind speeds and failure occurrences was expected. This was not clearly visible in the results and high wind speeds did not seem to affect all components. Only the results for the pitch system showed higher numbers of failures in the presence of high wind speeds at the time of failure occurrence. Thus, in the presence of high winds the pitch system is likely to fail abruptly. This does not seem to match previous studies, which indicated that most WT components are affected negatively by high wind speeds. However, this section is only concerned with the wind speed conditions during the last 10-minute observation before failure. During this time the wind speeds that cause the the component to fail might not be present.

In Figure 3.5 the results for the 1-D clustering and apriori rule mining are shown.
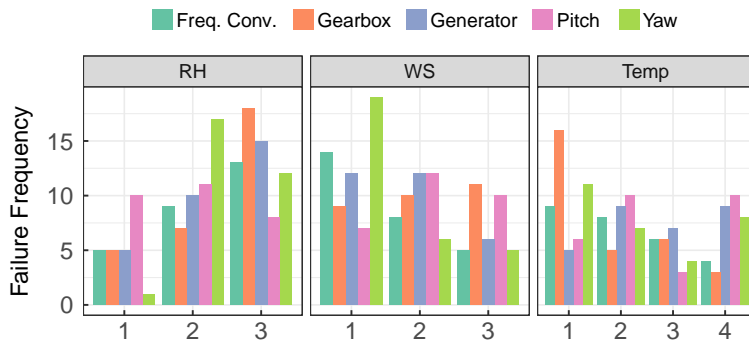
Figure 3.5. Conditions at the time of failure, obtained for 1-D clustering.

It can be seen that the number of clusters for RH and temperature differs from the number of manually assigned labels. The results for RH are quite similar to the ones obtained with the manually labelled data. The distribution of failure occurrences for the different clusters assigned to WS and ambient temperature (Temp) is way more balanced among the different wind speed categories than in Figure 3.4. This could be due to the clustering algorithm, which intents to balance the number of data points assigned to each cluster.

Most component failures occurred under mild and cold temperature conditions at the time of failure occurrence. Both, the apriori results obtained with manual labelling as well as clustering indicated these conditions as critical ones. Especially, the gearbox was subject to many failures in the presence of lower temperatures.

3.5.5 Meteorological Conditions during the whole Observation Period - Multi-D Clustering

The ARM algorithm based on Multi-D clustering can be used to find the general environmental conditions throughout the whole observation period ahead of failure. In Table 3.9 it is shown that all components were affected by corrosive RH conditions. Especially, gearboxes showed high support values for cool mean monthly temperatures $Temp_m$. The algorithm suggests further that generator failures mainly occur at low wind speeds and lower temperatures (cluster 1) at the exact time of failure, as well as mean monthly temperature in the cold range. Nonetheless, right before the failure occurrences, elevated wind speeds were registered (cluster

2). Converter failures mainly occurred in the presence of low wind speeds, mild monthly mean temperatures and in highly corrosive RH conditions at the time of failure. This might be due to the fact that electrical components are very sensitive to high relative humidities. The pitch system frequently failed under higher wind speed conditions at the time of failure as well as cool mean temperatures and highly corrosive RH conditions. This was an expected outcome, as these are the conditions that mostly challenge the pitch system. The yaw system showed to be affected mainly by low wind speeds and cool monthly mean temperatures. This component is used to place the WT rotor in the best position to catch the leading wind direction, and especially in low wind speed conditions this is a rather difficult task. Hence, the yaw system is constantly moving in search of the best direction in order to capture most of the incoming wind and to maximise the energy production.

3.5.6 Changing Environmental Conditions over Time - 1-D Clustering and Labelling

Having discussed the weather conditions directly at the time of failure occurrence in Section 3.5.4 as well as the general conditions during whole observation period in Section 3.5.5, this section will analyse their behaviour throughout the observation period. For this purpose, the apriori framework based on labelling and 1-D will be used. The aim is to test if the conditions are alternating or steady throughout the observation period. This will be carried out for the wind speed and relative humidity. As mentioned earlier, the Multi-D clustering did not perform well for punctual analyses and cannot be used here.

In order to analyse the wind speeds, for which extensive previous information was available, labelling was used for the data processing. For analysing the relative humidity a mixture of labelling and the 1-D-clustering algorithm was used. Here, there was very limited previous knowledge about harmful RH thresholds, which could only be assigned based on findings from other areas.

In Figure 3.6 the percentage of appearance for each characteristic of the wind speed and relative humidity time series before each component failure are summarised. It is distinguished between steady (indicating no significant changes in the thresholds), alternating between different thresholds, as well as descending and ascending behaviour towards the time of failure.

Wind speed variations (alternations) seemed to affect especially the generator

and yaw system. The frequency converter showed higher failure frequencies under steadily low or rising wind speeds. The pitch system failed more often under low or high wind speeds without variations.

Nonetheless, further studies should evaluate observation periods that are longer than 80 minutes before failure and compare these to wind speed conditions measured at healthy WTs, in order to analyse in more detail the effects of wind speed variations.
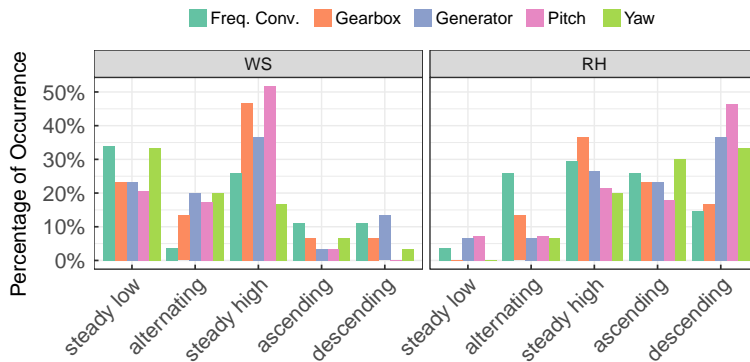
Figure 3.6. RH and WS conditions over time.

The relative humidity conditions before failure were mainly within the steady(high) corrosive region or ascending towards the latter. Descending relative humidities might be misleading, as this still includes the RH being in the (highly) corrosive range at least 10 hours before the failure occurrence.

3.5.7 Seasonal Failure Occurrence

In Figure 3.7 the failure events per season are displayed for each of the five components. During the winter months an increased number of failures was recorded for each component. This is most likely due to lower temperatures, higher wind speeds and other factors such as icing and snow. The gearbox showed the highest failure frequency during the winter months, which could e.g. be due to problems arising with lower lubrication oil temperatures.
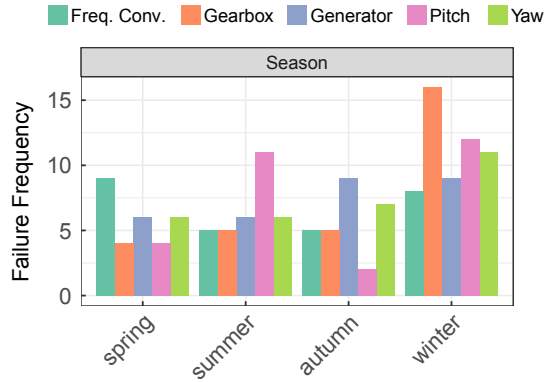
Figure 3.7. Seasonal failure occurrences.

3.5.8 Energy Production Losses

In nearly all cases, the power production shortly before failure decreased below the reference value for efficient production at the specific wind speed. This information can be especially useful for failure detection, as the power production is quite easy to measure in real time.

Using labelled input data for the ARM algorithm, it was possible to determine whether the power production was efficient or not. 1-D clustering additionally served to test to what extend the power production was inefficient. Hence, for analysing the power production efficiency before failure a mixture of labelling and 1-D clustering gives the best results. With Multi-D clustering only the general power production conditions were obtained, which were continuously in cluster 2. As the power production is highly correlated to wind speed, a production efficiency in cluster 2 means that most failures occur around an efficiency of 50 %, which corresponds to a wind speed in low to medium range. In the wind speed distributions this is the threshold with the highest probability of appearance.

In Table 3.9 an inefficient power production is detected for all components during the last 10 minutes before failure. The highest support values for inefficient power production right before failure were recorded for the converter and pitch system. Additionally, the gearbox showed an inefficient production also in P_{bf}. The reduced production efficiency could indicate a degradation before the failure.

3.5.9 Downtimes and Availability of Maintenance Personnel

In this study, the severity of each failure was characterised by the hours of downtime it has caused. The time of failure and the maintenance strategy affect the downtime. Hence, if there is a low availability of maintenance personnel at the time of failure occurrence, e.g. when the failure occurs during the night, the repair will start with a delay. The framework has identified the gearbox and generator failures as the most severe ones. This is consistent with the findings presented in Chapter 2. The pitch system failed in 79% during the day shift (8:00 to 18:00 o'clock). As it was shown that the pitch system mainly fails under high wind speed conditions, which can be related to the diurnal wind speed patterns and their peaks during the day. The converter failures were registered in 65% of the cases during the night shift (18:00 to 8:00 o'clock). This could be due to a rise in relative humidity during the late evening and night time, which highly affects electronic equipment. Additionally, lower temperatures during the night can have a negative effect on these components. The generator and gearbox failures occurred equally distributed between day and night shift.

3.6 Concluding Remarks for the Data-Driven Learning Framework

In this chapter a framework was presented for correlating failure data and environmental conditions. The functionality of the framework was demonstrated using a case study analysing the weather conditions shortly before wind turbine component failures. It was shown that the framework serves to indicate three different forms of behaviours of the meteorological conditions. These are: (1) analysing the general conditions over the whole observation period; (2) a punctual analysis of specific time steps before failure (e.g. the last 10-minute time step before a failure event); and (3) variations over time (alternations) in the measurements of certain variables. For each of these objectives, the data processing techniques: supervised labelling, unsupervised 1-D clustering and unsupervised Multi-D clustering performed differently. Each of these processing techniques was used as input to an apriori rule mining algorithm.

For obtaining rules on the general conditions over the observation period, the

Multi-D clustering provided the best input to the ARM. The interpretation of these rules is straight forward, nonetheless, the accuracy is lower compared to the other two techniques. For analysing the meteorological conditions before failure in more detail, supervised labelling and unsupervised 1-D clustering can be used. This includes analysing the conditions at a specific time step as well as the behaviour (e.g. variations) of each weather parameter over the whole observation period before failure. Nevertheless, while 1-D needs less pre-processing, labelling requires a large amount of information for categorising the variables, this includes time extensive and costly experiments or expert judgements. If profound knowledge on the behaviour of the weather variable is available, e.g. for wind speed, labelling performs best. But, it can introduce a certain degree of bias, especially when the previous information is limited. In the latter case, 1-D clustering works better.

The analysis showed that the failures of all components occurred mainly during winter time. Furthermore, high relative humidities, low temperatures and fairly high wind speeds affected most components, especially the pitch system. Moreover, a loss in production efficiency was recorded for all components at the moment of failure and for the gearbox even during the prior time step.

The framework has shown to perform well and the spectrum of possible results is big. It can be used in subsequent studies for a variety of objectives related for example to O&M and condition monitoring.

As in this chapter, however, only weather data obtained during a relatively short period before the failure event were considered, some effects might not have been observed. Hence, further approaches should consider longer observation periods. Especially, the wind speed conditions shall be investigated in more detail, which will be subject to the following chapter.

4

Anomalies in Wind Speed Conditions before Failures

The last chapter was dedicated to the analysis of the environmental conditions over a relatively short period of time before WT component failures. Unquestionably, failures are also often caused by cumulative stress over a larger period of time, which can be induced by short-time variations in e.g. wind speeds.

Hence, in this chapter the wind conditions before the failures of six main WT components are investigated in more detail over a larger time period. Especially short-time variations are expected to induce significant loads on the components affecting their reliability negatively, [56]. These variations have not been studied in detail previously. Earlier studies tend to average the wind speed conditions over a longer period of time, resulting in a loss of information regarding their short-term variations. For example in [79, 80] the mean and standard deviations of wind speed records before the failures of several components have been studied. It was shown that both increase during the last month and week before the failures. The mean and standard deviation, however, do not describe the wind speed conditions sufficiently well. Even shorter time step anomalies, such as e.g. sudden velocity peaks, can seriously affect the components' degradation processes.

In order to find these previously unknown pattern, algorithms for time series knowledge discovery such as anomaly detection and pattern discovery are used. These are applied to the data taken directly from the failed WTs' SCADA systems,

providing a realistic picture of the on-site conditions. The objectives are (1) to detect unexpected patterns (anomalies) in wind speed time series before failures and (2) to identify previously unknown and frequently recurring patterns (motifs) in the series. An observation period of 140 days before each failure event is used to capture the wind speed conditions throughout several stages of the components' deterioration. During the observation period no other failures besides the final event were registered for the turbines, hence, it can be assumed that the WTs were in a healthy state during the time before the failure event.

The detected short-term wind speed changes will, in the further, be referred to as anomalies. However, it shall be made clear, that they do not indicate an abnormality in the wind speed per se. These variabilities are a perfectly natural behaviour. The term 'anomalies' rather refers to unexpected peaks or changes from the WTs' point of view, as they affect the health of their components negatively.

The motif discovery will be carried out on the trend components of the respective time series with the aim of finding similar trends in the series recorded before each component failure. The results presented in this chapter were published in Paper (III).

4.1 Time-Series Data Mining

Knowledge discovery techniques, such as for example anomaly detection and pattern recognition, are used in many different areas, [81] and have become some of the most important machine learning tasks for time series data.

Similar to other industries, the majority of the information gathered by WT CMS and SCADA systems is available in time series format. Techniques for time series data mining are used extensively for wind speed and power forecasting [82] as well as condition monitoring of wind turbine components, [83]. However, to the author's knowledge detecting anomalies in wind speed time series before WT component failures using advanced time series data mining techniques has not been subject to previous work, yet. The use of CMS and SCADA for condition monitoring is not discussed in this chapter as this will be done in Chapter 7, where it is considered more appropriate.

In Figure 4.1 the two techniques used in this chapter as well as the necessary

time series processing approaches are summarised. They will be explained in detail in the following.
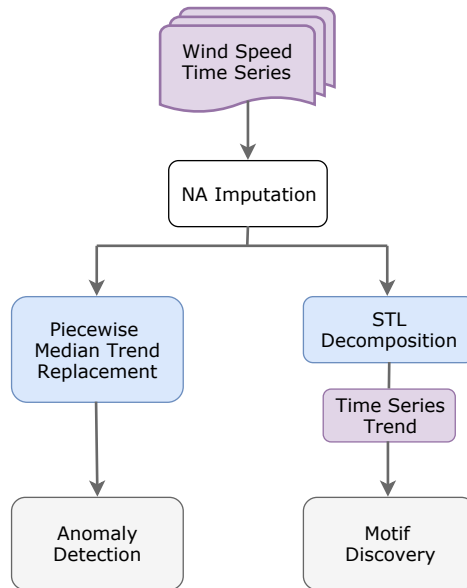
Figure 4.1. Flowchart of the anomaly detection and motif discovery using time series data.

4.1.1 Anomaly Detection

In data science, anomalies are defined as specific patterns that do not conform to expected behaviour, [84]. Frequently, these non-conformities are also referred to as outliers, discords, aberrations, surprising patterns or peculiarities. They can often indicate critical or faulty states. Considering the effects of weather conditions on WT reliability, for example, abrupt changes in wind speed can induce higher loads on components and worsen their health conditions. Anomaly detection could help to identify the changes that contribute to the deterioration of the components. However, in order to avoid problems, the input data have to be carefully pre-processed. It must be considered that anomalies might also be introduced into the data due to e.g. measurement errors, or the pre-processing itself.

In literature, a large amount of different anomaly detection techniques can

be found. These include regression and statistical approaches, as well as classification techniques such as naive Bayes, support vector machines, and clustering algorithms. Extensive reviews of various time series data mining techniques for anomaly detection are given in [81, 85, 86].

For the herein performed anomaly detection in wind speed time series an algorithm based on statistical learning is applied, similar to the Seasonal Hybrid Extreme Studentised Deviate Test (S-H-ESD) presented in [87], which can detect both local and global anomalies. This algorithm was chosen as it has shown to perform very well for long time series. The procedure consists of two steps: (1) a time series decomposition based on sliding windows and (2) the generalised Extreme Studentised Deviate Test (ESD), proposed in [88] and [89].

(1) Decomposition: It is very difficult to detect anomalies in raw wind speed time series, due to the fact that the latter are known to show daily seasonality, as well as an underlying trend. Hence, the time series need to be decomposed into seasonal, trend and random components before being further analysed. The most common technique for time series decomposition is the Seasonal and Trend decomposition using LOESS (STL), [90], where LOESS stands for 'LOcal regrESSion'. The STL, however, is problematic in the presence of anomalies and for long time series, as it is likely to introduce artificial anomalies, [87]. For this reason, a de-trending approach based on sliding windows is proposed in [87]. The time series is divided into relatively short segments using non-overlapping sliding windows. Then, the trend component of each time series segment is replaced with the piecewise median of the same raw segment. In this manner, long time series can be de-trended without introducing false anomalies. The decomposition procedure is summarised in the following steps:

- *Extract seasonality:* Determine the seasonality of the whole time series using STL. This serves to decide the size of the windows for the subsequent step.
- *Piecewise trend replacement:* Split the raw time series T into i non-overlapping windows $W_T(t)$, which contain at least two periods of the seasonality. Then, for all $W_T(t)$:
    a. Extract the seasonal component S_T with STL ,
    b. Calculate the median $\widetilde{T}$,
    c. Obtain the residual with $R_T = T - S_T - \widetilde{T}$.

(2) Test for outliers (ESD): Subsequently, the generalised ESD is applied to the residual (time-series after removing seasonality and trend), whereas the piecewise median serves as baseline for the anomaly test. The ESD is a generalisation of the Grubbs' test [91], and its objective is to test the null hypothesis of not having any outliers against the alternative hypothesis of having r outliers. For this, r separate tests are performed for each possible number of outliers $r = \{1, 2, ..., m\}$, given an upper bound m. Given a number n of observations in the time series, the test statistics are defined in [89] as:

$$R_q = \frac{max|T - \bar{T}|}{\sigma} \quad , \quad \text{with} \quad q = 1, 2, ..., r \quad . \tag{4.1}$$

with the sample mean $\bar{T}$ and the sample standard deviation σ. However, as the mean and the standard deviation are known to be very sensitive to outliers in the input data, [92], more robust measures will be used instead. Hence, in the present case the sample median $\widetilde{T}$ and the median of the absolute deviation MAD from the sample median are used respectively. The latter is defined as:

$$MAD = median(|T_i - median(T)|) \quad , \quad \text{with} \quad i = 1, 2, ..., n \quad . \tag{4.2}$$

With this Eq. 4.1 is modified to:

$$R_q = \frac{max|T - \widetilde{T}|}{MAD} \quad , \quad \text{with} \quad q = 1, 2, ..., r \quad . \tag{4.3}$$

The aim is to find the observation that maximizes $|T_i - \widetilde{T}_i|$. This observation is excluded and the same test is carried out for $n - 1$ observations. This process is repeated until r observations have been removed. The result is a vector R of r possible candidates for outliers, the r test statistics. In order to test the null-hypothesis H_0 (i.e. there are no outliers in the data set) versus the alternative hypothesis H_a (i.e. there are up to r outliers), the critical values for $q = \{1...r\}$ have to be calculated with:

$$\lambda_q = \frac{(n - q)}{\sqrt{n - q + 1}} \sqrt{\frac{t^2_{\alpha/(2n-q+1),\nu}}{(n - q - 1) + t^2_{\alpha/(2n-q+1),\nu}}} \quad , \tag{4.4}$$

where α corresponds to the significance level, and $t_{\alpha/(2n-q+1)}$ is the upper critical

value from the t-distribution with $\nu = (n-q-1)$ degrees of freedom. A candidate is accepted as outlier if $R_q > \lambda_q$ (i.e. the null hypothesis is rejected). By maximizing q for which $R_q > \lambda_q$, the number of detected outliers in the time series is defined. For further information reference is made to [89, 91, 93]. The maximum number of possible outliers r is specified as 15% of the total number of observations n and the statistical significance level is set to $\alpha = 5\%$.

4.1.2 Motif Discovery

Motifs are defined as frequently appearing patterns in sub-sequences of time series, which are very similar to each other, [94, 95]. Extracting these previously unknown recurring patterns from time series data, has been subject to research for many years, e.g. [94–98]. An extensive review on the latter, can be found for example in [86]. Here, the aim of this application is to find recurring patterns only within the trend component of the wind speed time series. Hence, before starting the actual motif discovery, the trend component of each time series are extracted using STL.
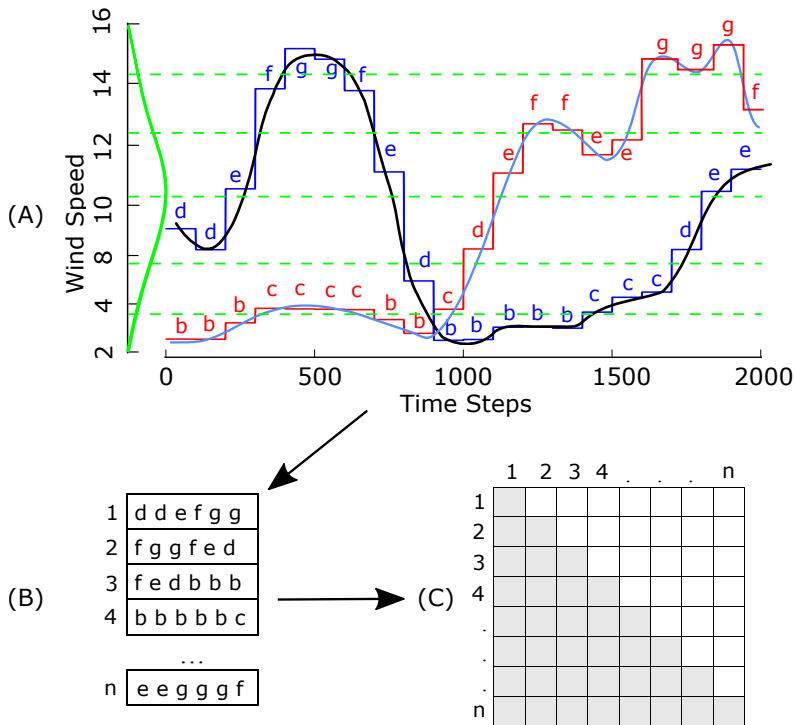
Figure 4.2. Example for the motif detection using SAX representation of two time series.

A widely used algorithm for motif detection in time series was proposed by Chiu *et al.* in [97] and is able to detect recurring patterns within one time series.

In Figure 4.2 the procedure of the motif detection is exemplified for two time series. This procedure consists of three steps:

(A) Dimensionality reduction with Symbolic Aggregate approXimation (SAX),

(B) Extracting sub-sequences using sliding windows.

(C) Finding motifs with random projection

These will be discussed in detail in the following. The reader shall keep in mind that although some of these techniques have similar features as the techniques discussed in the previous Section 4.1.1, this is an independent analysis and does not build upon the latter.

(A) SAX representation   Due to the fact that the time series are very long, firstly, the dimensionality of the time series has to be reduced. A Symbolic Aggregate approXimation based decomposition, as presented in [99], is used to symbolically represent the numerical series. SAX is frequently used in bioinformatics and has shown very good potential when used in combination with anomaly detection and motif discovery in time series, [100]. It uses symbols (often letters) to represent the various aspects of a time series, as shown very simplified in Figure 4.2 (A). A so called 'alphabet' indicates how many different symbols are used to represent the whole series. The alphabet used in this study consists of 6 letters. The SAX algorithm consists of two sub-steps: (A.1) Piecewise Aggregate Approximation (PAA) and (A.2) conversion of the PAA sequence into a series of letters. The PAA divides the time series consisting of n data points into ω equally spaced segments, whereas $\omega << n$. Then, each of the ω segments is replaced by the average value of the data points contained in ω. These piecewise averages are shown in the figure as horizontal blue or red lines. The sequence of these mean values is called the PAA approximation (or transform) of the original time-series. The SAX algorithm assumes that the discretised time-series approximately follow the Normal distribution. This assumption was taken by the developers of SAX after analysing a very high number of time series data sets. With this, q equal-sized areas under the Normal curve are produced. These are given as green dashed cut lines in the Figure 4.2, where as the Normal distribution is displayed as solid green line. Assigning a symbol (e.g. letter) to each area, results in a sort of look up table

for each value in the PAA transform. If we consider an alphabet $\bar{a}$ consisting of the letters $\bar{a} = \{b, c, d, e, f, g\}$, all PAA coefficients below the lowest cut line are marked b, those between the second lowest and the lowest are named c, and so on.

(B) Extracting sub-sequences In step (B), the time series are split into sub-sequences of a predefined length using sliding windows with a size of 100 time steps, corresponding to approximately sixteen hours, with a 50% overlap. With this several sub-sequences are extracted and stored, see Figure 4.2 (B). These sub-sequences are often called 'words' as they are made up of symbols represented in the earlier defined 'alphabet'.

(C) Finding motifs Finally, all sub-sequences are compared among each other using random projections. The results are saved in a collision matrix, as shown in Figure 4.2 (C). The 'planted (l, d)-motif' problem was defined in [101] as finding an unknown pattern of length l (called 'l-mer') in a set of sequences. This unknown patterns should differ at exactly d positions. Buhler and Tompa [102] developed a powerful algorithm to solve this problems via random projections. For this, randomly h out of the l positions are selected (hashed), where $h \leq (l - d)$. In this study, l corresponds to the window size of sixteen hours and d is chosen such that the different sub-sequences have to match at least 70% in order for them to be considered a motif. All l-mers are stored in a hash table and grouped in so called 'buckets' according to their hashed value. This is carried out several times with different hash functions. The number of iterations must be chosen high enough so that the probability of a bucket containing at least a previously defined number of sequences s in w iterations, is ≥ 0.95.

In Figure 4.3 this is exemplified for a length of $l = 10$ and a projection size of $h = 5$. The projection size h must be chosen small enough to guarantee that several possible motifs hash to the same bucket, but at the same time large enough to avoid including spurious l-mers.

If any buckets contain a number of l-mers that is greater or equal to s, these are called enriched buckets and are expected to be motifs. In the subsequent 'motif refinement step' the enriched buckets are used as input to an Expectation Maximisation (EM) algorithm, developed for motif discovery problems in [103, 104],

which maximises the likelihood of each motif candidate being actually a motif. More information on this procedure can be found e.g. in [105].
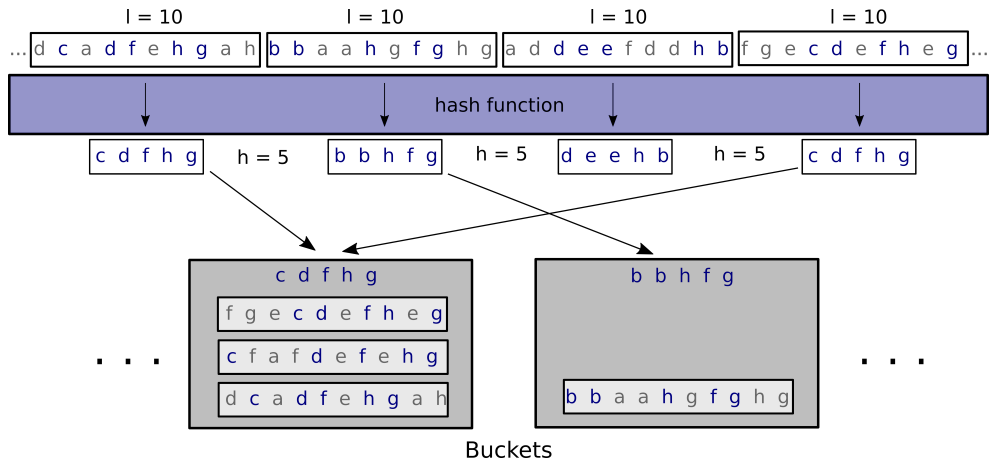
Figure 4.3. Example for the random projection algorithm.

4.2 Data for the Time Series Knowledge Mining

The data used for the anomaly detection is a sub-set of the data presented in Chapter 2. The set consists of data representing 3288 operational years taken from 1096 modern three-bladed and pitch regulated turbines with rated capacities between 0.66 MW and 2 MW. In Table 4.1 the number of failures per component considered in this sub-set are summarised. Six main components are analysed: the blades, converter, gearbox, generator, pitch and yaw system. This data set was chosen according to the availability and quality of the SCADA data, as discussed below.

The wind speed conditions were obtained directly from the SCADA systems of the failed WTs and were available as 10-minute mean measurements. SCADA systems are installed in most operating wind turbines and can provide a huge amount of information on the WTs operational status. Certainly, the SCADA systems are able to provide measurements at higher resolutions (e.g. 1 Hz), however, many operators do not store these huge amounts of data. Thus, this chapter explores the use of 10-minute mean data for the purpose of anomaly detection in wind speed time-series before WT component failures. These data are usually available to

Table 4.1. Data used for the anomaly detection.

Component	Nb. of Failures
Blades	48
Generator	40
Pitch System	16
Gearbox	26
Yaw System	15
Converter	13
Total	**158**

operators at no additional cost (e.g. due to extensive data storage). Furthermore, after analysing the information given in the SCADA systems of several different WT technologies, it was found that the wind speed and power are measured by all SCADA systems. Other variables such as turbulence intensity, etc. are often not recorded. Especially, older wind turbines, which are more likely to suffer from component failures do not provide this information.

The SCADA data needed to be carefully checked for measurement errors as well as missing data points, which could affect the outcome of the study negatively. To ensure correctly functioning measurement systems, the wind speed data obtained at the analysed WT were compared to the measurements of close-by located WTs of the same type. Additionally, only SCADA time series containing less than 5% missing values over the whole observation period of 140 days were used. As anomaly detection is very sensitive to missing values, even these small amounts had to be imputed before applying the algorithm.

In [106] various methods for replacing missing values in time series are summarised. After testing several, a technique presented in [107] was used, which showed the best compromise of computational effort and accuracy. This method is especially useful for time series with strong seasonality. It extracts the seasonal component of the time series and uses linear interpolation on the remaining data. Afterwards the seasonal component is added again. It shall be stressed that this technique only served for the imputation of the missing values and it is not part of the decomposition methods used for anomaly detection.

4.3 Results of the Anomaly Detection

By applying the anomaly detection algorithm to the discussed data set, the global and local anomalies for each time series have been detected. In Figure 4.4 an example of the discovered anomalies (blue circles) before a generator failure are shown. These include unexpected steep drops and sudden wind speed peaks. It can be seen that the majority of the anomalies were detected during the last 7 days before failure.
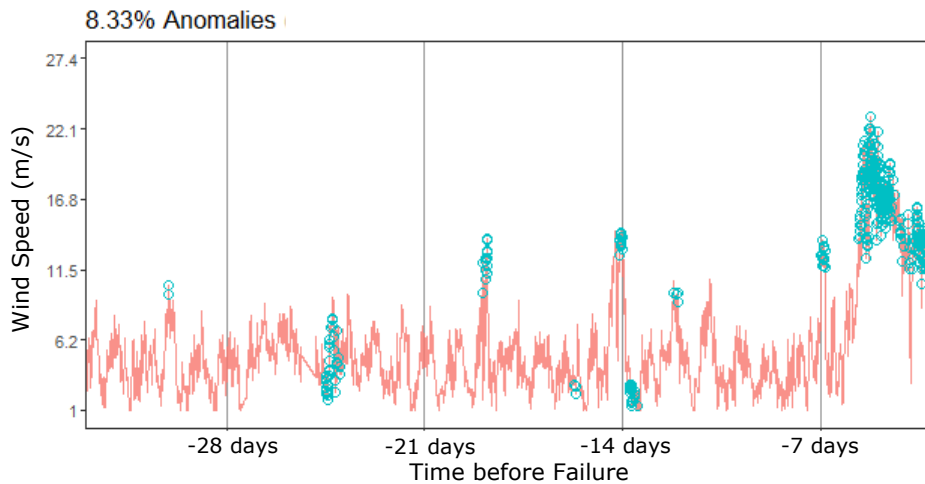
Figure 4.4. Example of detected anomalies (blue) in wind speed time series before a generator failure for an observation period of 1 month.

As displaying the all 158 available time series would not be very handy for the interpretation of the results, the average number of detected anomalies per day in all time series of each component are summarised in Figure 4.5. The total number of time series per component corresponds to the number of failures of the respective component. These are shown over three different observation periods: 140, 70 and 30 days before failure.

Within the last 30 days before failure a significantly higher number of anomalies was detected for all components. Especially during the last month before blade failures the figure shows a steep increase in anomalies per day. The wind speed conditions before gearbox failure contained the lowest amount of anomalies per day throughout the whole observation period. These results underline the hypothesis

69
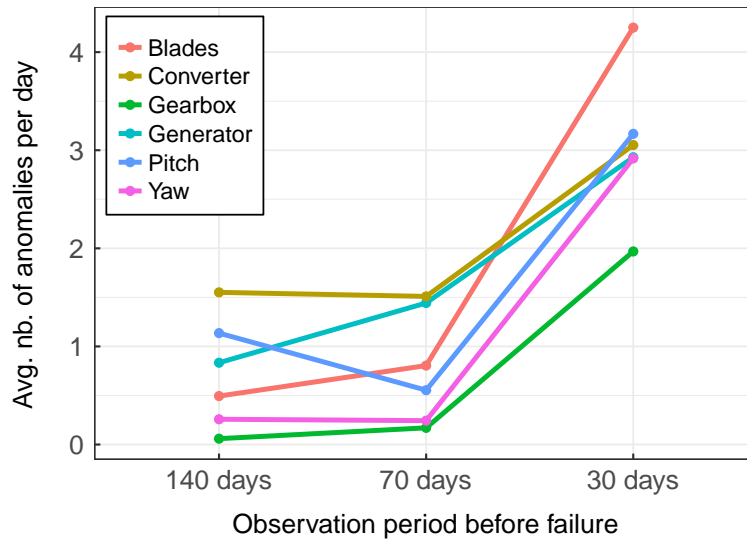
Figure 4.5. Average number of detected anomalies per day in wind speed time series of different lengths before failure occurrence.

that the failure behaviour of these components is affected by short-term changes in wind speed.
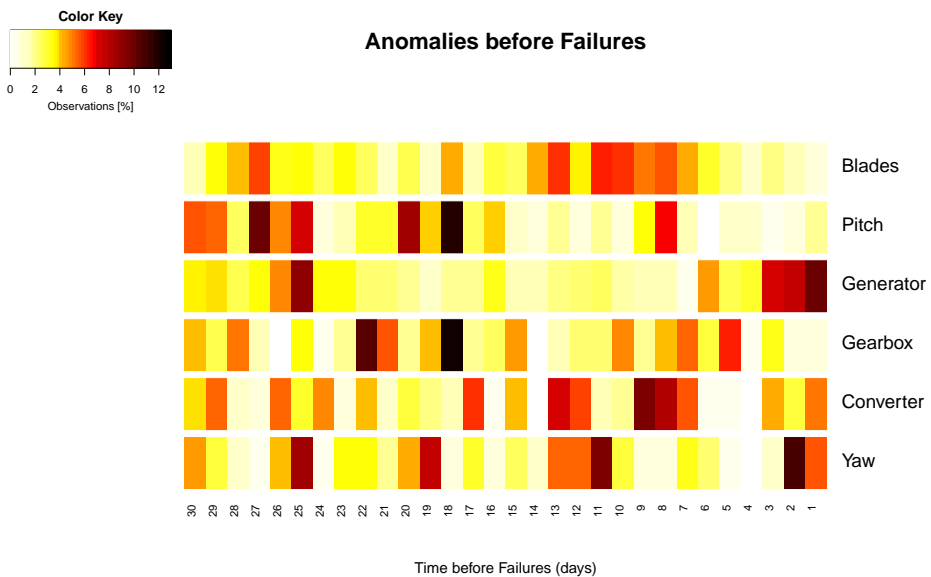
Figure 4.6. Detected anomalies in wind speed time series 1 month before component failures.

For analysing when exactly during the last month before failure the anomalies were detected, in Figure 4.6 a heat-map is presented, showing the obtained anomalies per day. The detections for each time step are presented as their contribution to the total number of anomalies per time series for this component. This enables an easier interpretation of the results, as the total number of anomalies differ for each component. However, the components shown in Figure 4.6 have to be analysed separately.

Blades: A high number of anomalies were detected between 14 and 7 days before the blade failures. During the last 7 days before the event, significantly less anomalies were present. A possible reason for this is that blade problems are in many cases not detected at the exact moment when they occur. For example cracks in the blade surface often occur under extreme wind speed conditions and propagate until they reach a critical state after a certain number of load cycles. Furthermore, manual blade inspections are a rather difficult task and usually require special tools and equipment (e.g. cranes, drones). Thus, these are frequently carried out with a certain delay. Even if the blade is not in a good state it might still be functioning, until the maintenance team decides to shut down the turbine due to the blade degradation.

Pitch system: Between day 30 and 25 a large number of anomalies were detected followed by a period of 3 days with rather regular wind speed time series. Subsequently, during day 20 and 15 again a period with a higher number of anomalies were recorded. The pitch system regulates the rotor speed and protects the turbine from damages, in wind speed conditions that show a large number of irregularities it is especially challenged. Experiencing several periods in which the wind speed suddenly rises or drops, the pitch system has to actuate immediately, which affects the component's reliability.

Generator: The anomalies were almost exclusively detected during the last three days before failure. This indicates that the generator is highly affected by these conditions. The generator converts mechanical to electrical energy and consists of both, electrical and mechanical components. In general, electrical components have shown to fail abruptly, while mechanical ones are subject to degradation

processes over a longer period of time. Hence, the anomalies can lead to short circuits and damages in the electrical parts of the generator, which cause immediate failure of the component. Nonetheless, only taking into consideration the wind speed conditions might, however, not be the best method to detect malfunction in this component. Generator currents, for example, have shown to be quite reliable indicators for the detection of faulty generators, [108, 109]. These two methods could be combined in further applications.

Gearbox: As shown in Figure 4.5, prior to gearbox failures very few anomalies were found. The degradation processes of gearboxes are slow and the failures are discovered with a certain time delay. They could not clearly be related to the occurrence of anomalies.

Yaw System: As shown in Figure 4.5 a lower number of anomalies per day were recorded for this component. The anomalies obtained for the yaw system during 30 days before failure showed a similar behaviour as the ones registered for the pitch system. As discussed in Chapter 3, the yaw system seems to be less affected by wind speed, rather by frequently changing wind directions. Hence, it would be interesting to include the changing wind directions in further work.

Converter: Throughout the observation periods of 140 days and 70 days, the converter showed the highest and almost constant daily average number of anomalies. Furthermore, a higher number of anomalies were detected during the last 30 days. This leads to the assumption that the converter is affected by constantly occurring anomalies over longer period of time. As these anomalies lead to peaks in the produced electrical power, they can cause recurring damage to the converter.

4.4 Results of the Motif Discovery

Having shown to be highly affected by short term changes and recurring anomalies in wind speed over long periods, the trends of the time series recorded before converter failures are used to analyse the performance of the motif detection.

In Figure 4.7 it can be seen that the most frequently detected pattern is characterised by a steep increase in wind speed with a minimum length of approximately

100 time steps (almost 16 hours). This is visualised by the coloured sub-sequences in five randomly selected wind speed time series measured before converter failures. The red lines indicate the regular time series data, while the other colours are the detected patterns in each of the series. The graphs show in total 4320 time steps before the time of failure occurrence, which is equal to 30 days prior to failure. This pattern has been found on average 6.4 times per time series before converter
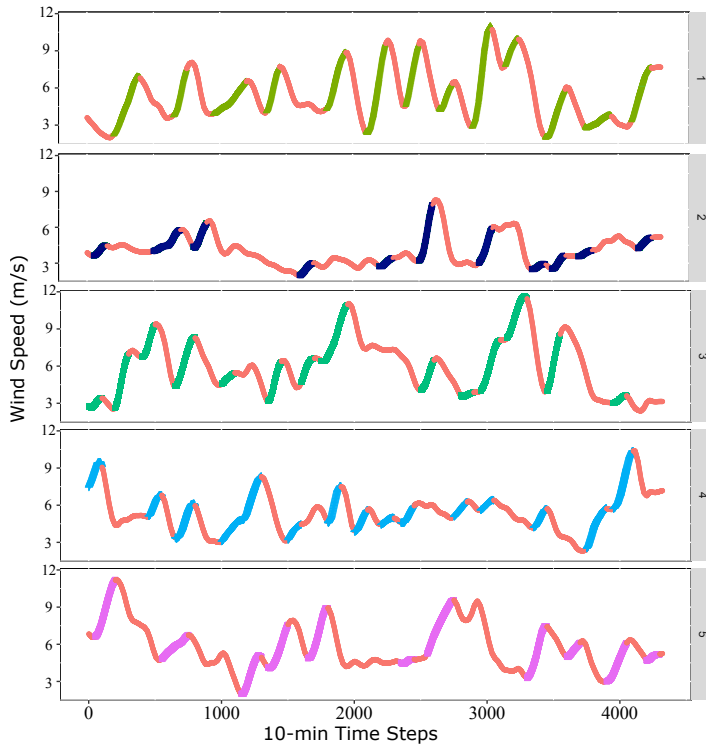
Figure 4.7. Detected patterns in wind speed time series of converter failures.

failures. Nonetheless, the algorithm also detected fairly short parts of the time series as motifs, which can be seen in time series 2 and 4 in Figure 4.7.

Although, the computational effort is quite high, this technique has shown to be able to discover previously unknown and recurring pattern. The results obtained for the converter failures lead to the assumption, that the latter tend to occur after frequent steep rises in the wind speed trend over a longer period of time. This underlines the findings from Section 4.3.

Nonetheless, the application of the motif detection in the case of wind turbine failures did not result in a clear useful outcome as for example the anomaly detection

did and needs to be studied further. The presented findings need to be compared to those of other components. Interpreting these results and deriving statistically relevant conclusions was found to be rather difficult, as also discussed by [110]. This is a common issue, and usually the detected motifs have to be analysed manually. In [110] a method to calculate the expected motif counts with the obtained ones based on Markov Chains is proposed. This could be taken into consideration in future studies. Furthermore, an optimisation of the applied window size and the size of the 'alphabet' should be carried out.

4.5 Conclusions for the Anomaly and Motif Detection

Earlier studies showed that the wind speed measurements during the last month before component failures are frequently characterised by higher means and standard deviations compared to the months where no failures were recorded. It was stated that these metrics give a good insight on the wind speed conditions, however, do not describe its short term changes sufficiently. Hence, the aim of this chapter was to detect the short-term changes in wind speeds and their effect on the failure behaviour of six main components: blades, converter, gearbox, generator, pitch and yaw system. An anomaly detection algorithm based on statistical learning was applied to wind speed time series recorded during 140 days before 158 WT component failures. In order to provide a useful method for wind farm operators, intentionally, 10-minute mean wind speed measurements were used. The herein employed knowledge discovery techniques have shown great potential and could be implemented in future O&M models for wind turbine failure prediction.

All of the analysed components, but especially the blades and the pitch systems, showed to be highly affected by short-term changes and anomalies within the wind speed time series measured before their failures. Hence, a significant higher average number of anomalies per day has been recorded throughout the last two to four weeks before failure. The wind speed measurements before converter failures were characterised by many anomalies over a long observation period up to 140 days. In further analyses of the yaw system, the changing wind directions should be taken into account. Furthermore, analysing the generator currents in combination with the anomaly detection in wind speeds, could lead to further findings regarding the failure behaviour of the generator.

The motif discovery has shown to be able to successfully detect unknown patterns in wind speed time series before failure. It underlined the results obtained from the application of the anomaly detection algorithm. However, this method needs to be tested more extensively and a larger-scale optimisation of this procedure tailored to wind turbine failure analysis needs to be carried out.

Nonetheless, both, the anomaly detection and the motif discovery, provide interesting new insights on the wind speed behaviour before failure and should be implemented in further O&M modelling.

After having shown in this and the previous chapter, that certain environmental conditions affect the failure behaviour of certain wind turbine components, the following chapter will develop reliability models including the complex combinations of these conditions.

5

Advanced Reliability Modelling

In Chapters 3 and 4 it was shown that there is a strong correlation between the meteorological conditions to which the turbines are exposed to and their failure behaviour. In this chapter, at first, an overview over existing methods for reliability modelling is given and specific applications in wind energy are discussed. Then, a WT failure model is proposed taking into account the effects of several environmental parameters on the failure behaviour of WTs and their components. In order to avoid common problems in failure modelling, variable selection and complexity reduction techniques are incorporated. The results of this chapter were published in Papers (IV) and (V).

5.1 Background Reliability Modelling

The objective of reliability modelling is to estimate the failure behaviour of systems and components over time. Reliability models can be categorised into probabilistic models and physical models. Probabilistic models use failure data collected over a considerable amount of the systems' life-time and are often referred to as data-driven reliability models. Physical models can be used to represent the deterioration of the system under given load conditions, however, require a thorough understanding of the physical processes leading to failures. As WTs are very complex systems, building detailed physical models of all components involved in the different WT technologies, would be an extremely difficult task.

This thesis focuses on data-driven (probabilistic) reliability models and in the following some of the main concepts for reliability modelling will be introduced. At first, probability distributions traditionally used in reliability modelling are discussed. Then, stochastic processes, such as counting processes and Markov chains will be explained. Finally, a short summary of other techniques for reliability modelling will be given.

Nonetheless, a complete review containing all possible reliability modelling techniques will not be presented. This can be obtained from literature, e.g. [111–113].

5.1.1 Probability Distributions

Probabilistic reliability models are usually based on so-called life distributions, which are probability density functions defined over an operational parameter, such as time, distance or cycles. Among these distributions are for example the exponential, normal, log-normal and Weibull distributions [111].

The probability density function (PDF) $f(t)$ expresses the relative frequency of failure occurrence and has an expected value (mean), which is commonly called the mean time to failure (MTTF), of:

$$E[T] = MTTF = \int_0^\infty t f(t) dt \quad .$$ (5.1)

Given a specific PDF, the probability of experiencing a failure event on a time interval $[0, t]$ can be obtained from the cumulative density function (CDF), which is defined as:

$$F(t) = \int_0^t f(t) dt \quad .$$ (5.2)

The CDF is often referred to as *unreliability*. With this, the reliability function or the probability of not having any failure event can be derived as:

$$R(t) = 1 - F(t) \quad .$$ (5.3)

The failure rate function $\lambda(t)$, i.e. the number of failures occurring at a specific unit of time, is then given by:

$$\lambda(t) = \frac{f(t)}{R(t)} \quad . \tag{5.4}$$

Weibull distribution. As the Weibull distribution is one of the most widely used life distributions, [113], it will be used in the further to exemplify how to derive the reliability and failure rate functions. More information on other life distributions can be found for example in [111, 113, 114].

The PDF of the three-parameter Weibull distribution is given by:

$$f(t) = \frac{\beta}{\eta} \cdot \left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \cdot e^{-\left(\frac{t-\gamma}{\eta}\right)} \quad , \tag{5.5}$$

with the shape parameter β, the scale parameter η, the location parameter γ and the operational time t. The MTTF, is denoted as:

$$MTTF = \gamma + \eta \cdot \Gamma\left(\frac{1}{\beta} + 1\right) \quad , \tag{5.6}$$

with the gamma function Γ. The CDF, the Weibull reliability and failure rate functions can then be written as:

$$F(t) \quad = \quad 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \quad , \tag{5.7}$$

$$R(t) \quad = \quad e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \quad , \tag{5.8}$$

$$\lambda(t) \quad = \quad \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \quad . \tag{5.9}$$

Due to its flexibility, the Weibull distribution can be used to model several properties of a system's failure behaviour. By adjusting the shape parameter β, the life-time can be divided into three periods: the early life, the useful life-time and the wear-out period. This results in the famous bathtub-curve, of which an example is shown in Figure 5.1.

The early life is characterised by a decreasing failure rate and is modelled with a shape parameter $\beta < 1$. The failure rates during the useful life-time are assumed to be constant and are modelled with $\beta = 1$, which simplifies the Weibull distribution to a two-parametric exponential distribution with $MTTF = \frac{1}{\lambda}$. Towards the end of the system's life the failure rates show an increasing behaviour ($\beta > 1$) due to wear out processes that are affecting the components reliability negatively.
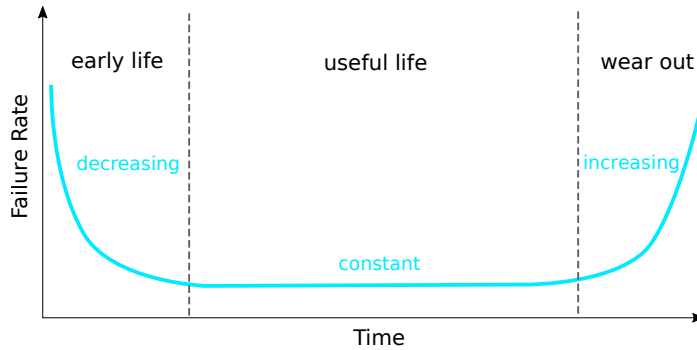
Figure 5.1. Example for the bathtub curve with the three periods.

5.1.2 Counting Processes

Counting processes are stochastic processes that are often used to model the deterioration of repairable systems over time. Three of the most frequently used counting processes in reliability modelling are the Homogeneous Poisson Process (HPP), Non-Homogeneous Poisson Process (NHPP) and the Renewal Process (RP). The HPP and NHPP shall be shortly introduced here, however, as they will not be used in this thesis no complete explanation will be given. Further information can be found e.g. in [115–117]. A counting process is used to model the number of failure events $N(t) \geq 0$ that are occurring over an interval of time $(0, t]$. The rate of failure occurrence $w(t)$ at time t can be derived by taking the derivative of the expected value $W(t) = E(N(t))$:

$$w(t) = W'(t) = \frac{d(E(N(t))}{dt} \quad . \tag{5.10}$$

The expected value denotes the mean number of failure events on the time interval, [111].

The HPP is the simplest form of a stochastic counting process and it assumes that the inter-arrival times of the failure events are independent and identically distributed following an exponential distribution with a constant rate λ, [118]. The inter-arrival time is the time difference between the failure events and its CDF is defined as:

$$F(t) = 1 - e^{-\lambda t} \quad . \tag{5.11}$$

The probability of exactly having n failures over time-interval $(0, t]$ is given by the Poisson distribution:

$$Pr_t\{N(t) = n\} = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad ,$$ (5.12)

with the expected number of failures $M(t)$ at time t:

$$M(t) = \lambda t \quad ,$$ (5.13)

and the repair rate $m(t)$ is denoted as:

$$M'(t) = m(t) = \lambda \quad .$$ (5.14)

The HPP is widely used to model the failure behaviour of repairable systems. This is due to the fact that it is the only model for repairable systems that can be applied to the useful-life period with a constant failure rate in the bathtub curve, [118].

The NHPP is a more flexible approach to model the failure rates of repairable systems. It can be used with a Power Law or exponential intensity function. The most popular approach is the NHPP with a Power Law intensity function and will be discussed in the further. However, more information on the NHPP with both intensity functions can be found e.g. in [115]. The probability of having n failures during $(0, t]$, is given by:

$$Pr\{N(t) = n\} = \frac{M(t)^n e^{-M(t)}}{n!} \quad ,$$ (5.15)

with the expected number of failures:

$$M(t) = at^b \quad , \quad \text{with} \quad a, b > 0 \quad .$$ (5.16)

Unlike the HPP, the NHPP does not have a constant repair rate (intensity function):

$$m(t) = \lambda(t) = \alpha t^{-\beta} \quad .$$ (5.17)

The model can be used to model both, the early life ($\beta < 1$) and the wear out phase ($\beta > 1$) of the bathtub curve. Moreover, for $\beta = 0$ it simplifies to the HPP.

5.1.3 Markov Processes

Markov chains are a special form of stochastic processes, and can be used to model systems with various states as well as transitions between these states, [111]. Markov chains can be classified into discrete-time Markov Chains (DTMC) and continuous-time Markov Chains (CTMC) depending on the definition of time either being discrete or continuous. In literature, CTMC are often referred to as Markov processes, [112, 119], which will be discussed in the further. The state of the process at time t is denoted by $X(t) = \{0, 1, 2, ..., r\}$ and all possible states are gathered in a state space χ. The transition probabilities $P_{i,j}$ from state i to state j of the Markov process at time t are defined as:

$$P_{i,j}(t) = Pr(X(t) = j | X(0) = i) \quad \text{for all} \quad i, j \in \chi \quad . \tag{5.18}$$

This can be written in form of a matrix:

$$P(t) = \begin{bmatrix} P_{0,0} & P_{0,1} & ... & P_{0,r} \\ P_{1,0} & P_{1,1} & ... & P_{1,r} \\ ... & ... & ... & ... \\ P_{r,0} & P_{r,1} & ... & P_{r,r} \end{bmatrix} \quad , \tag{5.19}$$

Figure 5.2 shows an example of a system with two states A and B, with which the use of CTMC for failure modelling will be shortly explained. If it is assumed that state A represents the healthy and state B the faulty state of the system, then the transition rate from state A to state B can be called the failure rate $P_{A,B} = \lambda$. This represents the probability that the process $X(t)$ enters state B after leaving state A.
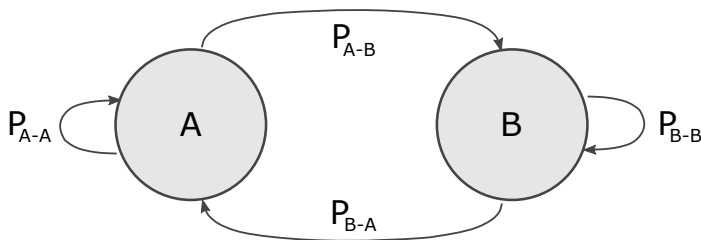
Figure 5.2. Example for a Markov Chain with two possible states A and B.

Likewise, the transition rate from state B to A is called the repair rate $P_{B,A} = \rho$. With this the stochastic transitional probability matrix at a specific point in time can be derives as:

$$P = \begin{bmatrix} P_{A,A} & P_{A,B} \\ P_{B,A} & P_{B,B} \end{bmatrix} = \begin{bmatrix} 1 - \lambda & \lambda \\ \rho & 1 - \rho \end{bmatrix} \quad . \tag{5.20}$$

For further information see e.g. [112, 113].

5.1.4 Other Aproaches

Accelerated life models [120] or proportional hazard models (PH) [121] can be alternatives to the discussed techniques. These account for externally induced stresses, however, they can only be applied when modelling continuous responses such as the time to failure of non-repairable systems. As they are not appropriate when modelling discrete responses as for instance the rate of occurrence of failures (ROCOF) in repairable systems, they will not be discussed further.

5.1.5 Reliability Modelling in Wind Energy

In the previous section some of the main concepts for reliability modelling have been introduced. As shown in Table 5.1, several studies have applied these techniques to wind turbine reliability modelling.

Table 5.1. References applying common reliability modelling techniques to wind energy systems.

Technique	References
Weibull Models	[122–124]
HPP and NHPP	[48, 125–129].
Markov Models	[130–135]
Renewal Process	[136, 137]
Accelerated Life Models	[138]
Proportional Hazard Models	[139–141]

However, being designed for machinery operating indoors or in fairly stable surroundings, these models are mainly driven by the operational life of the system or component. Thus, as it was shown that the WT failure behaviour is highly influenced by the surrounding conditions, they are not reflecting the reality sufficiently well. An exception are Markov processes, which could also be used to represent the different states of the environmental conditions as well as their effect on the failure behaviour. These, however, become extremely complex with an increasing number of input variables and their respective states. Hence, Markov chains are not feasible for achieving the herein defined objectives.

Despite the fact that much research has been dedicated to analysing the meteorological conditions before WT failures, there are very few existing studies that actually develop models to represent the effects of these conditions on the failure behaviour of WTs.

Wilson *et al.* [142] use artificial neural networks to model the effects of rainfall, pressure, relative humidity, temperature, wind direction, wind speed and gust speed at ground level on different WT components. In [73] the relationship between O&M cost and wind speed conditions was modelled via a Markov Chain Monte Carlo (MCMC) model, using data for one wind farm over a relatively short period of time. Furthermore, in [143] the same authors presented a non-parametric mixture model to compare the distributions of weather conditions including relative humidity, temperature and wind speed, obtained during normal WT operation to the ones measured in the presence of WT failures.

An additive Weibull failure rate model for WT rotor blades based on their age, the hours being exposed to full load and the wind speed (overload) was proposed by Faulstich *et al.* in [124].

Slimacek *et al.* [129] proposed a Poisson-Gamma model for modelling the ROCOF of WTs using a time-constant base function following the HPP. The environmental conditions are not modelled directly, but a proxi-covariate indicating the number of stops caused by external natural factors is implemented. They conclude that the proxi-variable representing the environmental conditions was the most important input covariate and that further studies should include these meteorological covariates directly.

Hence, in most cases the existing approaches either do not consider the environmental conditions directly in the modelling process, or the influence of different

weather variables on the WT failure behaviour is modelled separately for each variable. Furthermore, in order to set up effective maintenance strategies, WF operators are mostly interested in having models that can represent failure rates on relatively short time intervals e.g. on a monthly basis. Assuming constant failure rates during the entire useful-life period (e.g. with the HPP) can lead to wrong conclusions and higher O&M expenses.

Thus, there is a significant need for developing advanced reliability models that take into account the effect of the surrounding meteorological conditions.

5.2 Data used in this Chapter

For practical reasons before introducing the proposed models, the data which will be used to evaluate their performance is presented.

The data used in this chapter are comprised of historical failure logbooks, WT SCADA data and the WF's met mast data. Furthermore, weather data obtained from meteorological stations located close to the WFs are used. The failure data represent a sub-set of the historical data presented in Chapter 2. The composition of the latter regarding the failed components is shown in Figure 5.3.
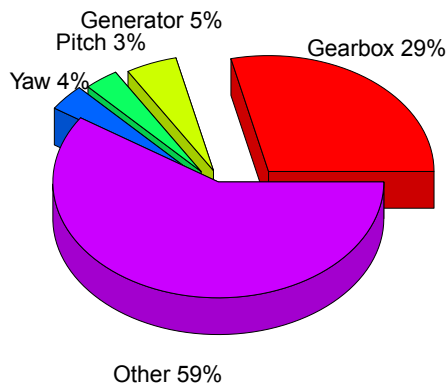
Figure 5.3. Failure data composition in this chapter.

The various WT components react differently to certain combinations of environmental conditions and setting up separate models for each component can be very useful for WF operators. Thus, apart from modelling the whole WT system, the functionality of the models for representing a single component's failure behaviour will be analysed. This will be exemplified using gearbox failures. It can be seen,

that around 29% of all recorded failures in this sub-set were related to the gearbox. Furthermore, in Chapter 2 as well as in [19], the gearbox was identified as one of the most critical WT components in terms of failure rate and downtime, which is highly affected by environmental conditions, [142].

5.2.1 Wind Farm Characteristics and Failure Data

In Table 5.2 the specifications of the wind farms included in the failure data base are displayed. Furthermore, the site specific terrain types, classified according to the IEC 61400-12-2 [144] standard, are shown.

Table 5.2. Wind farm specifications.

Wind Farm	Rated Capacity (kW)	Age (years)	Nb. of Turbines	Hub Height (m)	IEC Terrain Class
WF-A	2000	5	30	67	4
WF-B	1800	5	21	80	3
WF-C	800	8	36	55	5
WF-D	660	12	43	45	5
WF-E	660	14	25	45	2
WF-F	660	9	32	45	3
WF-G	660	11	74	45	3
WF-H	660	14	21	45	4
WF-I	330	4	16	30	2
WF-J	330	16	40	30	2
WF-K	300	15	45	30	4

Table 5.3 displays the number of registered WT system and gearbox failures per turbine and month. The data were recorded over a period of three years (January 2013 to December 2015) at eleven Spanish WFs, operating in total 383 turbines.

In Figures 5.4a and 5.4b the histograms of the monthly WT system and gearbox failures are displayed. It can be seen that the distributions of the failure occurrences are very right skewed with a high number of zeros.

Table 5.3. Summary of the historical failure data.

Wind Farm	WT Failures	Failure/Turb./Month	Gearbox Failures	Failure/Turb./Month Gearbox
WF-A	121	0.112	11	0.010
WF-B	40	0.053	14	0.019
WF-C	86	0.066	16	0.012
WF-D	59	0.038	30	0.019
WF-E	36	0.040	16	0.018
WF-F	36	0.031	14	0.012
WF-G	32	0.012	12	0.005
WF-H	23	0.030	8	0.011
WF-I	35	0.061	14	0.024
WF-J	29	0.020	10	0.007
WF-K	44	0.027	10	0.006
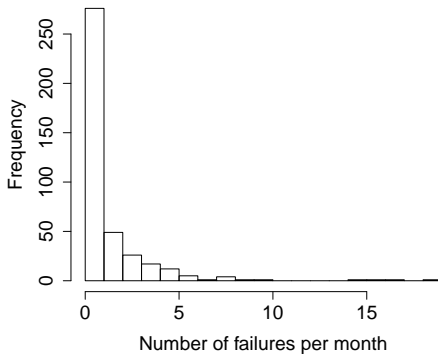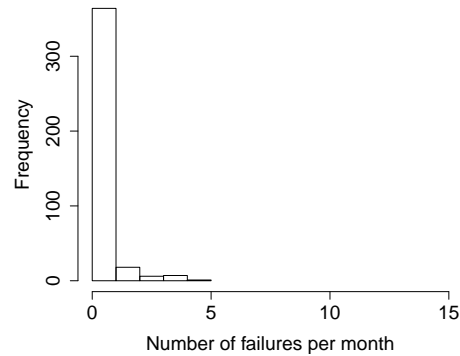
(a) Wind turbine system.   **(b)** Gearbox.

Figure 5.4. Histograms for the recorded failures per month.

5.2.2 Meteorological and Operational Data

Six environmental variables are included into the models, which were identified previously as the most critical ones regarding their effect on the WT failure behaviour, [59, 143]. To avoid issues due to distinct covariate magnitudes, all inputs are centred to a mean of 0 and divided by their standard deviation (scaled). In that manner, the feature importance within the context of the respective model can be compared, [145].

The monthly mean WS, turbulence intensity (TI), and monthly maximum wind speed (MaxWS) were obtained from the wind farms' met-mast at a height of

45 meters. The monthly mean Temp, RH and total monthly precipitation (Rain) are taken from close-by located meteorological stations. In Figures 5.5 to 5.7 the histograms of the measured meteorological data are shown. The TI was only available for two wind farms (*WF-A* and *WF-C*), hence, it will not be considered in the model for all WT technologies. Nonetheless, the influence of TI on the models will be discussed separately for *WF-C*.

To account for the WTs' operational conditions, the covariate PWR is introduced. It is defined as the average monthly active power (taken from the SCADA system) in percent of the turbines' rated capacity. With this, it is possible to include a model covariate that indicates how long and with how much capacity the turbines are operating on average during each month. Additionally, the hub-height, rated capacity and turbine diameter are included in the models, in order to distinguish between the different WT technologies.

(a) Wind speed.　　　　　**(b)** Ambient Temperature.
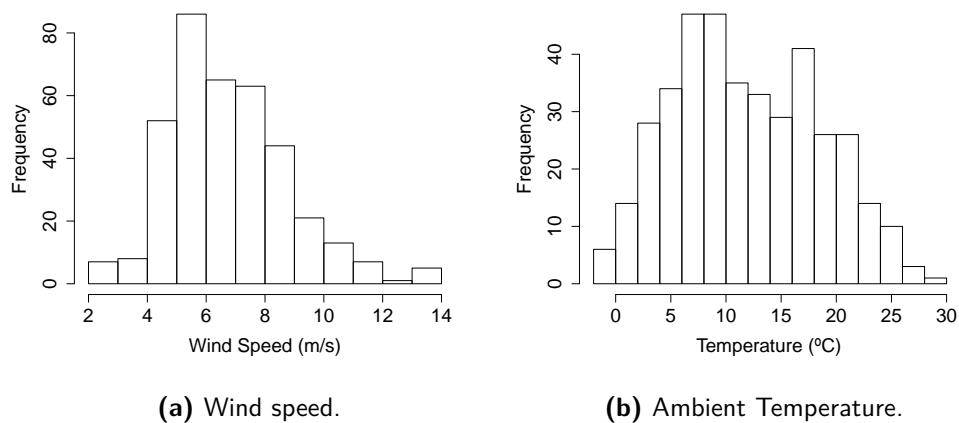
Figure 5.5. Histograms for the measured environmental data.

In Figure 5.8 the correlation between the different input covariates is shown. The TI was excluded, as it is only available for two WFs. Red squares indicate negative and blue squares positive correlation, and it can be seen that, although only pairwise correlation is displayed, many of the variables are highly correlated to each other. These correlations can result in serious problems during the modelling process. Hence, methods need to be used to eliminate the correlated input variables. This will be discussed in detail in the next sections.

88

(a) Relative humidity. **(b)** Precipitation.
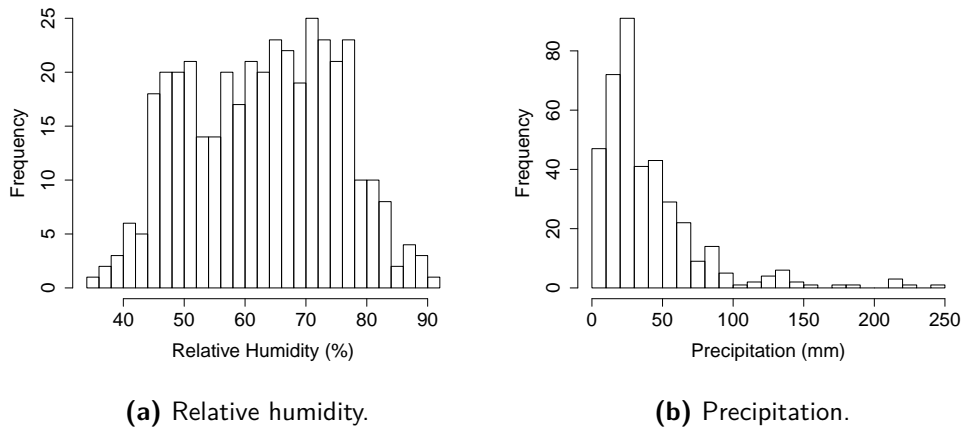
Figure 5.6. Histograms for the measured environmental data.

(a) Turbulence intensity. **(b)** Maximum wind speed.
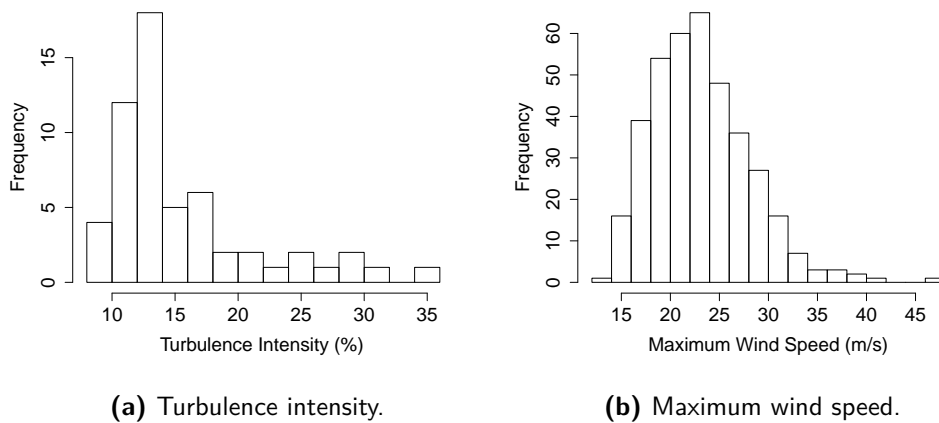
Figure 5.7. Histograms for the measured environmental data.

5.2.3 Model Memory

As the meteorological conditions can have a delayed and/or cumulative effect on the WT components, a so called model memory is introduced. This includes the environmental conditions of the month before the failure. In the following, the covariates representing the model memory will be denoted with the suffix *.mem*. With the model memory the total number of covariates rises from 11 to 17.
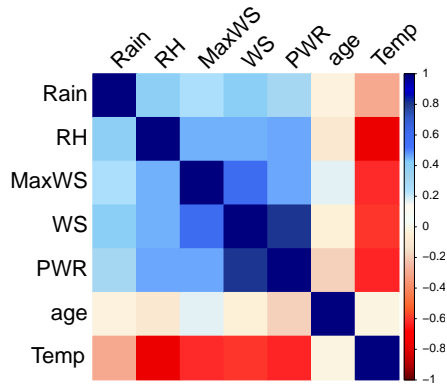
Figure 5.8. Pairwise correlation between the input variables.

5.3 Advanced Reliability Models and Problem Statement

In Section 5.1 reliability models based on the age of the system were discussed. These have found application in wind turbine failure modelling. For the above stated reasons, however, novel WT reliability models should incorporate the meteorological phenomena at the wind farm site.

Modelling the failure behaviour based on external variables could be carried out in several ways. Discrete responses, such as the number or the rate of occurrence of failures (ROCOF) over a given time interval are often modelled using Poisson regression. Nonetheless, these models have shown to be problematic in certain situations and are not able to represent all forms of data properly.

In this section the Poisson regression model as well as three advanced alternatives will be discussed. These include an extension of the Poisson distribution as well as two zero-inflated models. As these alternatives usually result in increased model complexity, more sophisticated parameter estimation and variable selection (PEVS) methods must be considered. The presented models along with the estimation techniques will be applied to the data in order to determine the most appropriate combination.

5.3.1 Failure Models

The Poisson distribution dictates that its mean is equal to its variance. This property can lead to serious problems when dealing with over-dispersed count data having

a larger variance than the Poisson distribution allows. Especially, the presence of unobserved effects (heterogeneity) and/or an abundance of zero counts in the data can lead to over-dispersion. When modelling the monthly failure events, for example, their distribution is usually highly right skewed, as the number of zeros is significantly larger than the non-zero counts - see Figures 5.4a and 5.4b. In these cases the Poisson models are likely to over-estimate the response variable.

In order to account for unobserved effects, the Poisson model is often extended with a Gamma distribution resulting in the negative-binomial (NegBin) model. This is frequently referred to as a Poisson-Gamma model.

Not only heterogeneity, but also a high number of zero-counts in the data (highly right skewed failure distributions) can cause over-dispersion. This could be avoided if two separate processes are considered: one generating the structural zeros using e.g. a binary distribution; the other one generating the counts including occasional zeros by a regular count process, such as the Poisson or the negative binomial distribution. These types of models are sometimes referred to as the zero-inflated Poisson (ZIP) and zero-inflated negative-binomial (ZINB) regression models and have been applied to a variety of different research areas, [146]. Being significantly more complex due to the two separate processes and the higher number of regressors, these models are often avoided. Nonetheless, if combined with a suitable parameter estimation technique other than the standard maximum likelihood estimation (MLE), they can lead to better results than conventional modelling techniques.

5.3.2 Parameter Estimation and Variable Selection Techniques

High dimensional regression problems often struggle with over-fitting (e.g. a model with too many covariates) and multicollinearity (strongly correlated covariates). By selecting suitable regularisation and variable selection techniques during parameter estimation, these issues could be avoided.

A standard technique for parameter estimation is the maximum likelihood estimation (MLE), which finds the set of model parameters β that maximizes a known log-likelihood function. As MLE does not provide any criteria for variable selection, it can lead to high coefficient variance and over-fitting when including many model covariates. Furthermore, MLE can result in highly biased outcomes if multicollinearity is present.

In order to select the best sub-set of model covariates and to avoid over-fitting,

various techniques have been developed. These include manual, forward, backward and stepwise model selection approaches. Yet, all of them are said to be very unstable and multicollinearity is still a remaining issue, [147, 148].

A more stable and computationally efficient form to avoid problems with over-fitting and multicollinearity can be achieved with penalised likelihood estimation. Frequently used penalisation techniques are the least absolute shrinkage and selection operator (LASSO) ℓ_1 [147] and ridge ℓ_2 (Tikhonov) [149] regularisation.

While ridge regularisation (on the ℓ_2-norm) deals with these issues by shrinking the coefficients to values close to zero, the LASSO using the ℓ_1-norm is actually able to set unimportant covariates to zero, with the aim of reducing the number of model covariates to the most relevant ones. Hence, it provides a form of variable selection for the regression problem. Yet, LASSO has shown to be still a highly biased estimate, as discussed in [150].

In order to overcome these bias issues, the minimax concave penalty (MCP) [151] and the smoothly clipped absolute deviation penalty (SCAD) [150] have been developed. Both, at first, apply the same penalisation rate as LASSO, but relax it towards zero with higher coefficient values. Hence, they reduce the bias by applying less shrinkage to the non-zero coefficients. Nevertheless, LASSO, MCP and SCAD have certain limitations in the presence of collinearity, as they assume the independence between penalty and correlation among predictors.

In the presence of highly correlated input data, the ridge regularisation usually performs better, [152]. However, as it does not perform variable selection, Zou *et al.* [152] proposed to combine the ℓ_1 and ℓ_2 penalties within the elastic net (Enet) regularisation. For the same reasons, Huang *et al.* [153] introduced the Mnet penalty, which is a combination of the the MCP and the ℓ_2 penalty. Both techniques, the Enet and Mnet, are able to delete groups of correlated covariates and, thus, reduce multicollinearity and over-fitting, carry out variable selection and help to avoid bias in the regression coefficients.

5.4 Methodology and Objectives

The objective of this chapter is to extend existing research by developing models that directly consider the environmental conditions and are able to capture their combined effect on the failure behaviour of WTs and their components. Previously

used regression techniques and parameter estimation methods, have not been found to entirely accomplish the objectives. Advanced alternatives need to be applied. Due to the very high numbers of (correlated) input covariates the latter, however, might require PEVS techniques other than MLE.

In order to find the most suitable modelling technique, four regression models: Poisson, NegBin, ZIP and ZINB; in combination with the three presented PEVS methods MLE, Enet and Mnet, are compared. The failure events are modelled on a monthly basis, as the environmental conditions significantly change throughout the year. As mentioned previously, the models will be applied to the failure data of the whole wind farm system, not distinguishing between the failed components, as well as on data containing only gearbox failures.

To cover the most relevant aspects the following steps are carried out and will be presented in this order in the results section:

1. **Evaluating the model performance with MLE:** At first the failures of the whole WT system and the gearbox will be modelled using Poisson, NegBin, ZIP and ZINB models in combination with MLE. As the MLE is not able to carry out any variable selection, it is used to understand how the model performs when all input covariates are included.

2. **Effect of the model memory:** Including the meteorological conditions of the month prior to failure can significantly enhance the model performance and will be analysed in detail.

3. **Performance of the PEVS methods:** By combining each regression model with either MLE, Enet or Mnet regularisation, the performance of the PEVS methods is evaluated. Figure 5.9 displays this modelling and evaluation process.

4. **Model Performance on data of a single wind farm:** To plan their O&M actions properly, operators often prefer separate reliability models for each WF. This is due to the fact that, on the one hand, the different WT technologies have a distinct failure behaviour, and on the other hand, the weather conditions vary strongly at each wind farm site. Hence, in a final step, the proposed models will be applied to the data of a single wind farm. Thus, the capability of the proposed models for establishing separate WF models will be assessed in a final step.
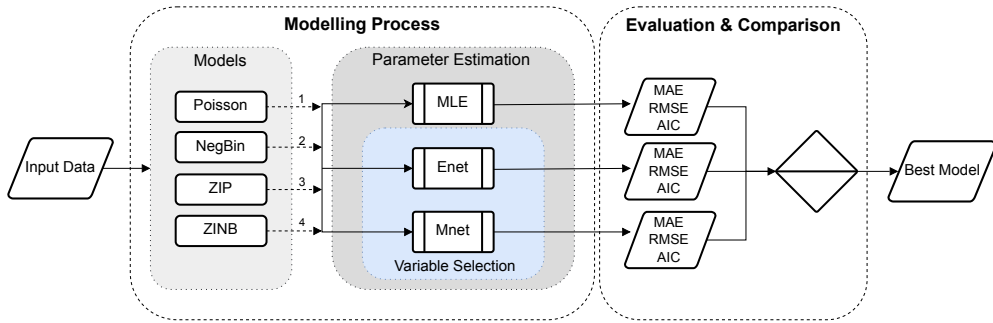
Figure 5.9. Modelling and evaluation process. AIC Akaike information criterion; Enet, elastic net; MAE, mean absolute error; MLE, maximum likelihood estimation; NegBin, negative binomial; RMSE, root mean square error; ZINB, zero-inflated negative binomial; ZIP, zero-inflated Poisson.

5.4.1 Model Evaluation

The evaluation metrics mean absolute error (MAE), root mean square error (RMSE) and the Akaike information criterion (AIC) will be used to compare the performance of the different modelling approaches.

The MAE and RMSE are given by:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\widehat{y}_i - y_i)^2} \quad , \tag{5.21}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(\widehat{y}_i - y_i)^2} \quad , \tag{5.22}$$

where $\widehat{y}$ are the modelled and y the measured data values. The AIC is a measure of the relative quality of statistical models that have been applied to the exact same data, [154, 155], and is defined as:

$$AIC = 2d - 2ln(\widehat{\mathcal{L}}) \quad , \tag{5.23}$$

where $\widehat{\mathcal{L}}$ is the maximum value of the likelihood function and d is the number of estimated parameters. If very small data samples are used, the AIC can be modified with a correction factor, [156], resulting in:

$$AICc = AIC + \frac{2d(d+1)}{n-d-1} \quad , \tag{5.24}$$

where n is the sample size.

Models with lower AIC and AICc values shall be preferred. By virtue of their definition, the AIC and AICc account for the model complexity and increase with an rising number of model parameters. Chi-square tests are often used for similar purposes, however, as this test is highly sensitive to low sample sizes (which is the case in this study), [157], most sources recommend the use of Fisher's *F-Test* instead. Nonetheless, as stated in [158, 159], the AIC and AICc have several advantages over the *F-Test*, and are able to find the best fitting model more reliably.

5.5 Mathematical Formulation of the Models

In Section 5.3 different possible alternatives to models have been explained. Problems of current WT failure models including over-fitting, over-dispersion, excess zeros in the response variable, variable selection, heterogeneity and multicollinearity have been discussed. Possible solutions to these problems were presented. Their mathematical formulations are introduced in this section. To account for the different numbers of turbines per WF, the failure counts are modelled with an offset of the number of turbines. This is essentially the same as modelling the failures per turbine over a certain time interval (ROCOF).

5.5.1 Regression Models

In the following, the four regression models: Poisson, NegBin, ZIP and ZINB will be introduced.

The probability distribution for the Poisson model is given by:[1]

$$Pr(y_i|x_i) = \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} \quad , \tag{5.25}$$

where y is the response variable of non-negative integer values, which in the present case is the monthly number of observed failures. The mean and variance are denoted

---

[1] For better readability and to distinguish the concepts, there is a slight notation difference between Eq. 5.12 and Eq. 5.25, however, in both cases the Poisson distribution is shown.

as $E(y_i|x_i) = Var(y|x_i) = \mu_i = \exp(x_i\beta_i)$, with the estimation coefficients β_i. The covariates x_i for each observation i include the variables described in Section 5.2:

$$
x_i = \begin{bmatrix} age \\ Max.WS \\ \vdots \\ Rain.mem \\ RH.mem \end{bmatrix}_i . \tag{5.26}
$$

The Poisson model can be extended to the NegBin model with the probability distribution:

$$
Pr(y_i|x_i, \vartheta) = \frac{\Gamma(y_i + \vartheta)}{y_i!\Gamma(\vartheta)} \left(\frac{\vartheta}{\vartheta + \mu_i}\right)^{\vartheta} \left(\frac{\mu_i}{\vartheta + \mu_i}\right)^{y_i} , \tag{5.27}
$$

with $\mu_i = \exp(x_i\beta_i)$. Here, the dispersion parameter ϑ is used to adjust the model regarding the degree of over-dispersion. The mean is given by $E(y_i|x_i) = \mu_i\tau_i = \exp(x_i\beta_i + \varepsilon_i)$, which accounts for the unobserved effects $\tau_i = \exp(\varepsilon_i)$ following a Gamma distribution. The variance of the NegBin model is denoted as $Var(y|x_i) = \mu_i + \mu_i^2/\vartheta$,

As stated in Section 5.3, zero-inflated models use one process governed by a binomial distribution and a second one governed by a count distribution such as Poisson or negative binomial.

The probability distribution for the ZIP model is denoted by:

$$
Pr(y_i|x_i) = \begin{cases} \sigma_i + (1 - \sigma_i)e^{-\mu_i} & \text{for } y_i = 0 \\ (1 - \sigma_i)\frac{\mu_i^{y_i}e^{-\mu_i}}{y!} & \text{for } y_i \geq 1 \end{cases}, \tag{5.28}
$$

with the zero-inflated regressors z_i, mean $E(y_i|x_i, z_i) = \mu_i(1 - \sigma_i)$, variance $Var(y_i|x_i, z_i) = E(y_i|x_i, z_i)[1 + \mu_i\sigma_i]$ and $\mu_i = exp(x_i\beta)$. The zero-inflation probability σ_i is given by the logistic link function $\sigma_i = \exp(\gamma_i z_i)/(1 + \exp(\gamma_i z_i))$

with parameters γ_i.

The probability distribution for the ZINB model is defined as:

$$
Pr(y_i|x_i, \vartheta) = \begin{cases} \sigma_i + (1 - \sigma_i)\left(1 + \frac{\sigma_i}{\vartheta}\right)^{-\vartheta} & \text{for } y_i = 0 \\ (1 - \sigma_i)\frac{\Gamma(y_i + \vartheta)}{\Gamma(y_i + 1)\Gamma(\vartheta)}\left(\frac{\vartheta}{\vartheta + \mu_i}\right)^{\vartheta}\left(\frac{\mu_i}{\vartheta + \mu_i}\right)^{y_i} & \text{for } y_i \geq 1, \end{cases} \tag{5.29}
$$

with the logistic link function given above. The mean is $E(y_i|x_i, z_i) = \mu_i(1 - \sigma_i)$ and the variance $Var(y_i|x_i, z_i) = E(y_i|x_i, z_i)[1 + \mu_i(\sigma_i + \vartheta^{-1})]$, with the model covariates x_i.

5.5.2 Parameter Estimation and Variable Selection Techniques

The PEVS techniques MLE, Enet and Mnet are used in this chapter and will be explained in the following.

Maximum likelihood estimation aims at finding the values for β that maximize the log-likelihood function $L(\beta|x_i) = \ln\left(\mathcal{L}(\beta|x_i)\right)$. Given the fact that maximizing the log-likelihood is essentially the same as minimizing the negative log-likelihood (loss-function), the MLE estimators are given by:

$$
\widehat{\beta}_{MLE} = \underset{\beta}{\mathrm{argmax}}\left\{L(\beta|x_i)\right\} = \underset{\beta}{\mathrm{argmin}}\left\{-L(\beta|x_i)\right\} \quad . \tag{5.30}
$$

Instead of minimizing the negative log-likelihood function, penalised regression aims at minimising the objective function $M(\beta)$:

$$
\widehat{\beta}_{pen} = \underset{\beta}{\mathrm{argmin}}\left\{M(\beta)\right\} = \underset{\beta}{\mathrm{argmin}}\left\{-L(\beta|X) + \lambda\Phi(\beta)\right\} \quad , \tag{5.31}
$$

with the penalty function $\Phi(\beta)$ and parameter λ that controls the trade-off between penalty and fit. By subtracting the penalty, sparsity is introduced and the magnitudes of the coefficients are lowered. The penalty function for the regularisation depends on the used technique.

The Enet regularization penalty is stated in [152] as:

$$
\lambda\Phi(\beta, \varphi) = \lambda\left(\varphi\|\beta\|_1 + \frac{1}{2}(1 - \varphi)\|\beta\|_2^2\right) \quad \text{with} \quad \lambda \geq 0 \quad . \tag{5.32}
$$

The parameter $\varphi \in [0, 1]$ controls the share of the ℓ_1 and ℓ_2 penalties. The Enet

estimates are defined as:

$$\widehat{\beta}_{Enet} = \underset{\beta}{\mathrm{argmin}}\Big\{ M(\beta; \lambda, \varphi)\Big\} \quad . \tag{5.33}$$

For $\varphi = 1$ a pure LASSO regression is obtained, which penalises the sum of the absolute values of the model coefficients $\|\beta\|_1 = \sum_{k=1}^{p} |\beta_k|$, where p is the number of model covariates. For $\varphi = 0$ a ridge regularisation with the model coefficients $\|\beta\|_2^2 = \sum_{k=1}^{p} \beta_k^2$ is obtained. In this chapter, by setting $\varphi = 0.5$, an Enet with equal share of LASSO and ridge penalty is used.

Contrarily to the Enet, which is using the ℓ_1 penalty in the first term, Mnet uses the MCP, which is defined in [151] as:

$$\lambda\Phi(\beta, \gamma) = \sum_{j=1}^{p} \rho(|\beta_k|; \lambda_1, \gamma) + \frac{1}{2}\lambda_2 \|\beta\|_2^2 \quad , \tag{5.34}$$

with $\lambda = (\lambda_1 \geq 0, \lambda_2 \geq 0)$ and regularisation parameter γ, as well as:

$$\rho(\beta; \lambda_1, \gamma) = \begin{cases} \lambda_1\beta - \frac{\beta^2}{2\gamma} & \text{, for } \beta \leq \gamma\lambda_1 \\ \frac{1}{2}\gamma\lambda_1^2 & \text{, for } \beta > \gamma\lambda_1 \end{cases} \quad . \tag{5.35}$$

The Mnet estimator is given in [153] as:

$$\widehat{\beta}_{Mnet}(\lambda, \gamma) = \underset{\beta}{\mathrm{argmin}}\Big\{ M(\beta; \lambda, \gamma)\Big\} \tag{5.36}$$

In order to find the appropriate shrinkage parameter λ that minimises the mean squared error, in all cases a K-fold cross validation (with $K = 10$) is carried out before fitting the models. More information on the presented regularisation techniques can be found in [151–153].

5.6 Results and Discussions - Failure Models

In this section the results of the modelling process shown in Figure 5.9 for general WT system and gearbox failures are presented. However, as the interpretation of the modelling results might not be straight forward, at first some comments will be made to aid the model interpretation.

5.6.1 Comments on the Model Interpretation

When analysing the influence of each regressor on the model outcome, several things need to be considered. This can only be analysed if at the same time all other regressors are assumed to be constant. Nevertheless, due to the naturally high correlation between many meteorological parameters, this is a very unlikely behaviour. The Enet and Mnet regularisations tend to eliminate these correlated regressors and the reader might get the impression that the latter are not important for the model interpretation.

Notwithstanding, as they are correlated to regressors that are included in the model, the deleted regressors still influence the model outcome and need to be considered in the interpretation. By taking a look at the correlation between all of the different input variables, this information can be combined.

In Figure 5.10 the pairwise Pearson correlation values are shown on the upper right triangle, the histograms and density functions for the measurements of each variable in the diagonal, and their scatter-plots with smoothed lines (LOESS) on the lower left triangle.
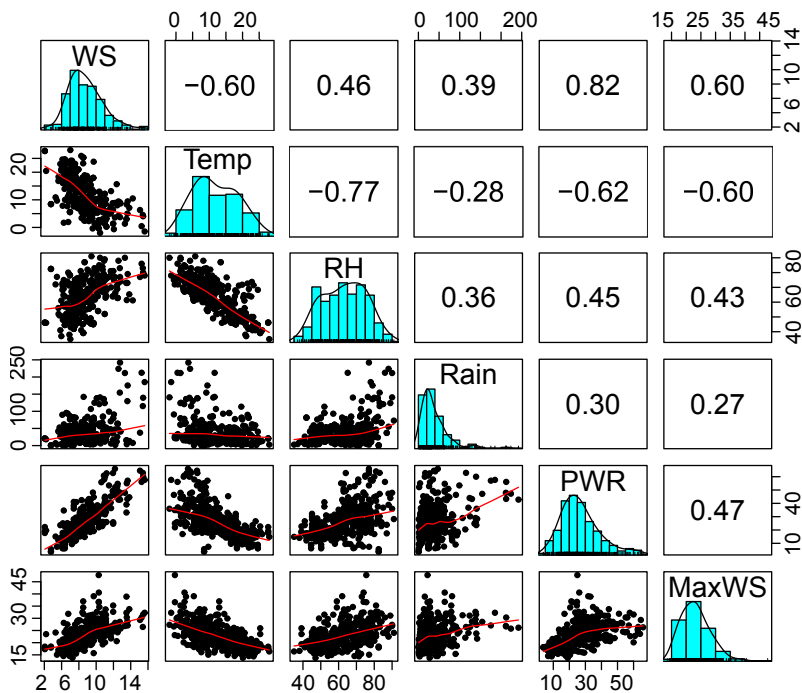
Figure 5.10. Correlation plots of the meteorological inputs.

The following relations can be obtained from the graph:

- The covariates WS, MaxWS and Temp are strongly negatively correlated, as with rising temperatures the WS and MaxWS decrease.
- The temperature and RH are negatively correlated.
- Precipitation is only weakly correlated with the other input parameters.
- WS and RH are positively correlated.
- The covariate average monthly active power in percent of the rated capacity (PWR) is highly positively correlated with WS and MaxWS; and negatively correlated with Temp.
- The RH is defined as the percentage of moisture held in the air in relation to the possible maximum moisture content at a given temperature. Hence, both, higher temperatures and/or less precipitation can cause lower RH. However, as shown in Figure 5.10 the temperature seems to be more important for the RH changes and there is almost no correlation between RH and precipitation.

These relations are of high interest for the interpretation of the model covariates later in this chapter, especially in Section 5.6.5. Firstly, however, the model performance using different PEVS techniques shall be evaluated.

5.6.2 Performance of the Regression Models using MLE

In Table 5.4 the model evaluation metrics for the four modelling techniques with MLE applied to data for general WT system failures are shown. Table 5.5 presents the results for the gearbox failures.

Table 5.4. Evaluation metrics for the wind turbine failure models with MLE.

Measure	Poisson	NegBin	ZIP	ZINB
AIC	1144.73	1004.02	1056.88	991.45
AICc	1146.92	1006.21	1065.94	1000.51
MAE	1.226	1.206	1.242	1.137
RMSE	1.903	1.848	1.883	1.812

It can be seen that in both cases, the the lowest values for all evaluation metrics were obtained for the ZINB-models. Hence, these models showed a substantially better fit to the data than the other ones. Furthermore, when comparing these

Table 5.5. Evaluation metrics for the gearbox failure models with MLE.

Measure	Poisson	NegBin	ZIP	ZINB
AIC	557.18	532.34	542.74	523.52
AICc	559.37	534.53	551.80	532.58
MAE	0.6005	0.6126	0.6430	0.5733
RMSE	0.7749	0.7827	0.7830	0.7572

results to those of other studies, e.g. [124], the error metrics for both models are significantly lower.

Figures 5.11 and 5.12 show hanging rootograms of the models (with MLE) for WT and gearbox failures, respectively. Rootograms were introduced in [160, 161] and are useful methods to display the model fit to the data. Furthermore, they can highlight issues such as over-dispersion and problems with excess zeros. In the figures, the red lines represent the expected counts and the gray bars are the observed counts, which are 'hanging' from the red lines. The distance between the bars and the reference line indicates the dissimilarity between expected and observed frequencies, [161]. The square root of the frequency is used on the y-axis, in order to display even small dissimilarities.

As presented in Figures 5.11a and 5.12a, the Poisson model has problems with high over-dispersion and the zero counts. The ZIP models shown in Figures 5.12c and 5.11c also present some degree of over-dispersion, but a significantly better fit for the zeros. Figure 5.11b and 5.11d as well as 5.12b and 5.12d present almost identical results for the NegBin and ZINB models, which handle over-dispersion and excess zeros significantly better than Poisson and ZIP.

Hence, having the lowest errors and dealing well with over-dispersion and excess zeros, the ZINB models show the best overall performance for both, the whole WT system and the gearbox failures. The error metrics for the model of the whole WT system, shown in Table 5.4, are fairly low, hence the model shows a very good fit. Nonetheless, as presented in Table 5.5, one can observe that the gearbox models have lower errors, and thus, the proposed ZINB model performs even better for modelling a single component. This is due to the fact that there is no distinction between the various components in the failure data of the whole WT system. They contain information on any failed component during the observation period,

(a) Poisson.

(b) NegBin.

(c) ZIP.

(d) ZINB.
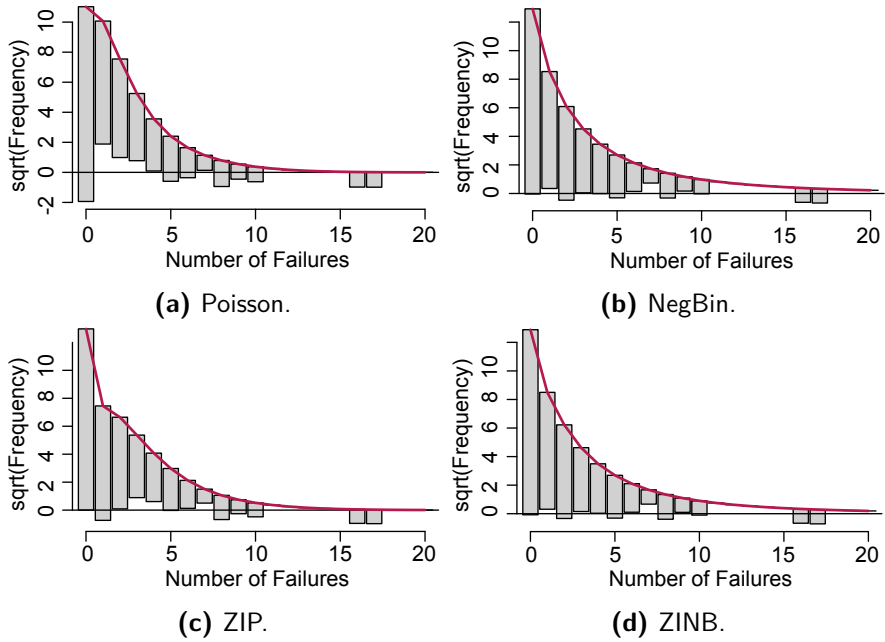
Figure 5.11. Rootograms for WT system failure models.

(a) Poisson.

(b) NegBin.

(c) ZIP.

(d) ZINB.
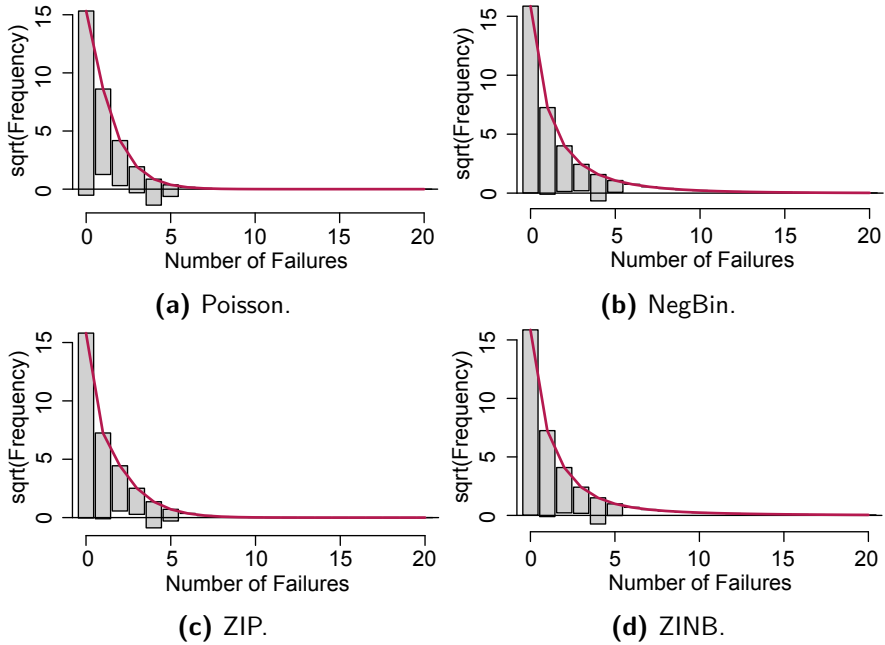
Figure 5.12. Rootograms for gearbox failure models.

and their failure behaviour is substantially harder to capture by models. As each component is affected differently by the meteorological conditions, the components should be modelled separately. Therefore in the following sections, only the results for the gearbox failure models will be displayed.

5.6.3 Results for the Model Memory

As the ZINB-model has shown the best results in the previous section, it will be used to display the effects of including a model memory. For this, the model with MLE will be used in order to emphasize the effects of the model memory without eliminating covariates, as Enet and Mnet do.

Table 5.6. Comparing ZINB-models with and without memory.

Measure	MLE	
	ZINB without memory	ZINB with memory
AIC	542.63	523.52
AICc	544.82	532.58
MAE	0.6702	0.5733
RMSE	0.7936	0.7572

In Table 5.6 it can be seen that including a model memory considerably decreases the MAE and RMSE about approximately 9%.

Event though the model with memory has more covariates, lower values for the AIC and AICc were obtained compared to those calculated for the model without memory. This leads to the assumption that certain input covariates of the model memory are important and need to be included in the failure model.

However, as introducing a memory results in a higher number of model covariates, the variable selection techniques discussed earlier in this chapter should be considered, as they can reduce model complexity by selecting the most relevant covariates. This will be discussed in the following.

5.6.4 Results of the Variable Selection Methods

To assess which of the variable selection methods works best, in the following the four considered models in combination with the Enet and Mnet penalty will be

applied to the gearbox failure data. In Table 5.7 the resulting evaluation metrics are presented. It can be seen, that the zero-inflated models outperform the Poisson and NegBin models in both cases, whereas ZINB again showed the best results. Only an approximately 1% difference in MAE and RMSE was found for ZINB-Enet and ZINB-Mnet. However, the values for the AIC and AICc are in favour of the ZINB-Mnet model. Comparing this to the evaluation metrics for the models with MLE, presented in Table 5.5, Mnet also outperforms MLE.

Table 5.7. Evaluation metrics for the models with Enet and Mnet .

	Enet			
Measure	Poisson	NegBin	ZIP	ZINB
AIC	554.19	530.85	512.36	508.32
AICc	555.34	532.18	513.88	510.80
MAE	0.6047	0.6103	0.6035	0.5468
RMSE	0.7776	0.7812	0.7768	0.7394
	Mnet			
Measure	Poisson	NegBin	ZIP	ZINB
AIC	552.43	527.39	503.27	494.45
AICc	553.76	528.72	504.42	495.43
MAE	0.6012	0.6090	0.5936	0.5503
RMSE	0.7754	0.7804	0.7704	0.7419

In Tables 5.8 and 5.9 the standardised model coefficients, standard errors and 95% confidence intervals for the ZINB-Mnet and ZINB-Enet models are shown respectively. The coefficients that were eliminated by the penalised regressions are not displayed in the tables, as they are not included in the resulting models. The variable ϑ is the model specific dispersion parameter, discussed in Section 5.5.

It is shown that the Enet model has 13 coefficients, while the Mnet model only has six. Both show a similar fit regarding the RMSE and MAE. However, the Mnet model shows lower AIC and AICc values.

As shown earlier in Figure 5.10, most input variables are highly correlated and Mnet successfully eliminated as many of them as possible in order to prevent collinearity, whereas the Enet model tends to include most of them. Thus, the Mnet model is able to describe the data very well by only using very few of the

input variables. As the model with less complexity shall be preferred, the Mnet model results will be discussed further.

Table 5.8. Results of the estimation of the ZINB models with Mnet .

Variable	Coefficient (β)	Standard Error	95% Confidence Interval
Intercept	-0.321	0.159	(-0.582, -0.060)
Rated Capacity	-0.327	0.160	(-0.591, -0.063)
Age	-0.430	0.137	(-0.655, -0.205)
Temp	-0.288	0.204	(-0.622, 0.047)
Rain	0.209	0.101	(0.043, 0.376)
RH.mem	-0.929	0.225	(-1.299, -0.559)
Log(ϑ)	1.139	0.777	–

Table 5.9. Results of the estimation of the ZINB models with Enet .

Variable	Coefficient (β)	Standard Error	95% Confidence Interval
Intercept	-0.395	0.189	(-0.706, -0.085)
Rated Capacity	-0.336	0.158	(-0.596, -0.077)
Age	-0.314	0.173	(-0.599, -0.029)
WS	0.176	0.234	(-0.208, 0.560)
Temp	-0.512	0.255	(-0.932, -0.092)
PWR	-0.112	0.250	(-0.523, 0.300)
RH	-0.372	0.263	(-0.805, 0.060)
Rain	0.286	0.120	(0.089, 0.484)
WS.mem	-0.187	0.259	(-0.614, 0.239)
RH.mem	-0.704	0.288	(-1.177, -0.230)
Rain.mem	-0.097	0.138	(-0.323, 0.130)
PWR.mem	0.214	0.233	(-0.170, 0.598)
MaxWS.mem	-0.240	0.197	(-0.564, 0.084)
Log(ϑ)	1.294	0.895	–

5.6.5 Interpreting the Model Covariates

The effect of each covariate on the model response can be interpreted by analysing the estimated coefficients of the ZINB-Mnet model given in Table 5.8

For the non-meteorological conditions the following observations were made:

- *Age:* The negative model coefficient for the covariate *age* suggests that the gearbox failures occur mainly in younger wind turbines. This result is consistent with literature, where premature gearbox failures are frequently considered one of the main problems of WTs. Although, being designed for a life-time of 20 years, literature states that gearboxes frequently suffer from serious damages within the first 2-11 years of their life-time, [162], or commonly fail at least once within the first 5 years, [163].

- *Rated Capacity:* A negative coefficient was obtained for the *Rated Capacity*, which means that turbines with lower rated capacity suffer from more gearbox failures. This, however, could be also because the data base itself includes more WTs with lower rated capacity.

The model coefficients for the meteorological covariates, showed the following behaviour:

- *Rain:* The covariate related to precipitation has a positive coefficient. This indicates that with more precipitation the gearboxes fail more frequently. Sudden and heavy rain can facilitate the air exchange between the gearbox and the surroundings, causing oil contamination and oxidation, which can accelerate the degradation of gears and gear bearings. This is explained in more detail later in this section.

- *Temperature (Temp):* The model further states that more gearbox failures occur for lower monthly mean temperatures. This is consistent with earlier findings, see Chapter 3. Furthermore, during months with colder mean temperatures (winter), usually the mean wind speeds are higher, causing increased wear on the gearbox. Additionally, lower temperatures affect the gearbox oil viscosity resulting in a less effective lubrication, as discussed below.

- *Relative Humidity (RH.mem):* The negative coefficient for the model memory covariate *RH.mem* shows that for lower mean relative humidities during the month prior to the gearbox failures, these occur more often. The higher *RH.mem*

values are most likely due to higher temperatures in the previous month (model memory). How the conditions in the month before failure affect the gearbox, is explained below.

Some of the model covariates are highly correlated, as discussed in Section 5.6.1. Hence, combinations of the latter need to be interpreted while taking into consideration the failure modes of the component.

Failure Modes. Wind turbine gearboxes can fail in several different ways. However, as shown in Chapter 2 and according to literature [54, 164, 165], a very high percentage of all gearbox failures are directly related to the gear bearings. The main reasons for WT gearbox bearing failures are oil degradation and contamination. Furthermore, temperature related changes in oil viscosity can affect the gearbox. Oil contamination is usually caused by moisture, particles and entrained air (foam), which can result in high vibrations and wear. These intrusions can enter the gearbox in a variety of ways, such as e.g. during manufacturing or maintenance . They could further be generated internally or be ingested through air exchange with the ambient air. The latter occurs frequently due to diurnal temperature variations that cause air to be sucked into the gearbox through the seals and 'breathers', as for example discussed in [166]. The gearboxes are not sealed completely and the typical breather systems in WT gearbox housings are usually not sufficiently preventing the contaminants from entering the system. The herein presented model suggests that the gearbox failures occur in the presence of these temperature variations, as discussed in the next two paragraphs.

The Month before Failure. The model indicates that the months before the failure events occur are characterised by lower relative humidity and thus higher temperatures, lower mean wind speeds and less PWR. Although, the relative humidity might be lower at higher temperatures, the increased air exchange contributes to an elevated risk of contamination inside the gearbox. As the effect of these contaminations usually occurs time-delayed, the component might only fail after a certain period of time or when the operational conditions change due to higher wind speeds and/or increased operational time.

The Month of Failure. The model states that the month, during which the failure events were registered, was defined by lower temperatures. This has also been shown by the data-driven learning framework presented in Chapter 3. As the previous month was rather characterised by higher temperatures, this is likely to indicate a transition month from warmer to colder seasons (e.g. during autumn). As stated in [167], especially, the daily temperature swings during these months can cause wear due to oil viscosity changes, which result in less oil flow. Further air exchange between the ambient air and the interior of the gearboxes through the breathers is facilitated by temperature variations due to periods of heavy rain. The correlations between temperature and rain are discussed e.g. in [168]. Additionally, along with lower temperatures usually higher mean wind speeds are measured, which cause a larger number of WT shut-down and start-up events and a longer time in operation. Under these conditions the gearboxes are mechanically challenged and possible damages due to previously entered oil contaminations can lead to a component breakdown. Hence, a combination of degraded and contaminated lubricants due to previous air exchange with the surroundings and problems with oil viscosity and higher loads during the failure month are affecting the gearbox life-time negatively. This has been successfully captured by the model.

5.6.6 Application of the Model to a single Wind Farm

In Sections 5.6.2 to 5.6.4 the ZINB-models with Mnet penalty have shown to perform significantly better than other combinations of modelling techniques. The ZINB-Mnet model has lead to very satisfying results when applied to a large data set containing several different WFs and turbine technologies.

However, the several WT technologies can react differently to combinations of environmental conditions, and operators often prefer modelling each technology separately. Thus, in order to test the model performance on a data set containing only one technology, a single WF taken from the data set presented in Section 5.2 is tested. *WF-C* was chosen for this analysis as its data contained information for the turbulence intensity, and with an age of 8 years the WF is well into its useful life without being very old. The results were expected to show slightly different model coefficients than for the whole data set.

In Figures 5.13 and 5.14 the modelled versus the original monthly failures for ZINB-Mnet with and without the covariate TI. Additionally, the constant failure

rate that is usually assumed by conventional reliability models during the useful life-time, as discussed in Section 5.1, is displayed. Figure 5.15 compares the kernel density plots for the two set-ups.

It can be obtained from the three graphs that the model including TI performs best. Furthermore, it is made clear that assuming a constant failure rate throughout the useful life-time leads to quite wrong results, which from a practical point of view can delay the repair and maintenance processes for months.
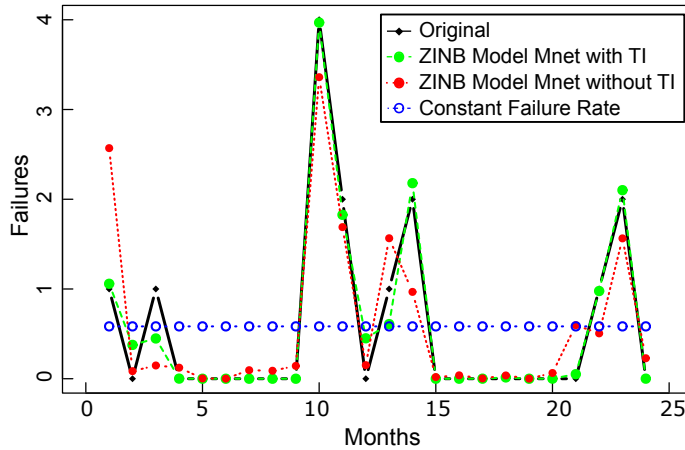
Figure 5.13. *WF-C*: gearbox failure model with and without TI – monthly failures .
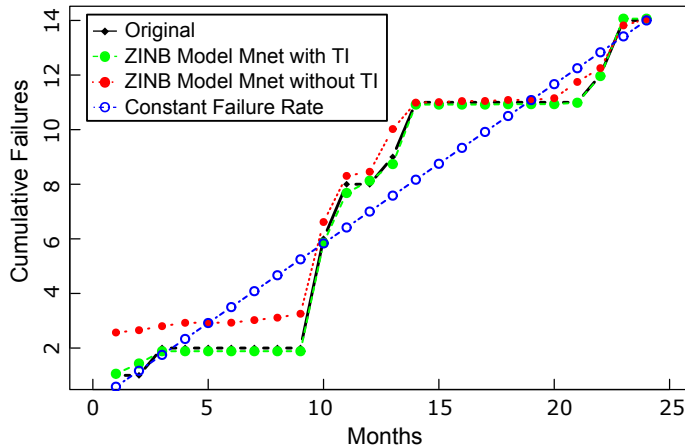
Figure 5.14. *WF-C*: gearbox failure model with and without TI – cumulative failures .
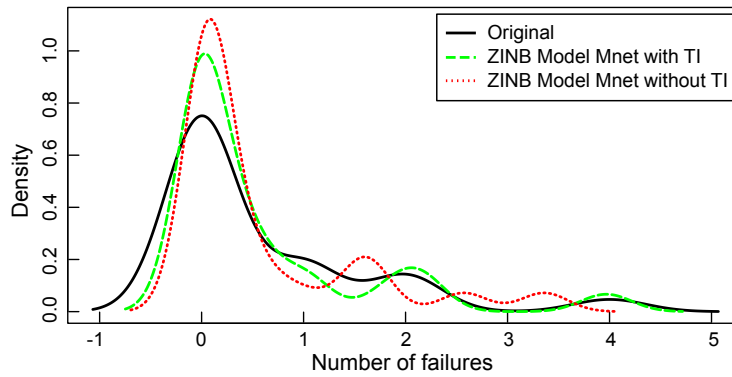
Figure 5.15. *WF-C*: kernel density plots of the original and modelled gearbox failures with and without TI.

Table 5.10 summarises the model evaluation metrics for the model with and without TI. It can be seen that including TI significantly increases the model performance, as lower values for AIC, MAE and RMSE were obtained. Having one additional covariate the difference between AIC and AICc for the model with TI is slightly higher. Nonetheless, the values are both lower than for the model that does not include TI.

When comparing these evaluation metrics to the results presented in Table 5.7 of Section 5.6.4, one can see that much lower errors were recorded when one WT technology is modelled separately. This was an expected result, as the failure behaviour of distinct technologies and different sites varies.

Table 5.10. Comparing the models with and without TI.

| | ZINB-Mnet | |
Measure	without TI	with TI
AIC	43.43	39.83
AICc	47.85	44.25
MAE	0.331	0.100
RMSE	0.485	0.192

As shown in Section 5.6.5 the terrain class was eliminated by both variable selection algorithms for modelling the gearbox failure data of all WT technologies. Nevertheless, as the terrain complexity affects the wind conditions, e.g. turbulence

intensity, which has lastly shown to play quite an important role for modelling gearbox failures, the terrain complexity enters indirectly into this model.

The coefficients obtained during the estimation process of the ZINB-Mnet model with TI are displayed in Table 5.11. The Mnet penalisation algorithm selected six environmental parameters. Here again, the model shows that the combination of low temperatures and high precipitation increases the number of gearbox failures. Furthermore, higher TI and MaxWS during the month of failure lead to more failures. Mnet did not eliminate the TI covariate, which shows that it is an important parameter for modelling the gearbox failures in this WF.

Table 5.11. *WF-C*: Results of the estimation of the ZINB model with Mnet including TI.

Variable	Coefficient	Standard Error	95% Confidence Interval
Intercept	-2.528	1.672	(-5.806 , 0.750)
Temp	-2.920	1.769	(-6.387 , 0.548)
Rain	2.154	0.845	(0.498 , 3.811)
TI	1.375	0.765	(-0.123 , 2.874)
MaxWS	4.286	2.021	(0.324 , 8.248)
WS.mem	0.746	0.542	(0.316 , 1.808)
TI.mem	0.518	0.312	(-0.095 ,1.130)

In this section the ZINB-Mnet model has shown to also perform very well for modelling one WT technology separately. The results contained even lower errors than when modelling several turbine technologies at the same time. As the data set for *WF-C* was fairly small, the coefficients show wider confidence intervals than those presented in Section 5.6.4. Bigger data bases (longer observation periods) would be needed to obtain lower confidence intervals. This is in turn likely to slightly affect the selection of the model coefficients.

5.7 Conclusion for the Reliability Modelling

In this chapter a novel approach for modelling WT failures, including the meteorological conditions the turbines are exposed to, was presented. Several regression models as well as suitable parameter estimation and variable selection techniques to reduce issues with high dimensional regression problems have been tested and

evaluated. To the author's knowledge this is the first work of this type in the context of modelling WT failures based on environmental conditions.

A model based on a zero-inflated negative binomial distribution in combination with an Mnet penalty is proposed for further use in the field. The model relies on two separate processes: one is generating the failure events as well as occasional zeros based on a negative binomial distribution; the other process generates the structural zeros and is governed by a binomial distribution. Possible unobserved effects (heterogeneity) in the model covariates are taken into account for using a Gamma distribution.

Furthermore, the Mnet penalty has shown to be able to select the most important input covariates very efficiently. Using the proposed model in combination with this penalisation, helps to prevent several problems such as over-dispersion, over-fitting, multicollinearity, etc. Additionally, including a model memory with the meteorological covariates of the previous month enhances the model performance significantly.

The proposed model was tested on a large data set and three specific test cases. At fist it was used for modelling WT system failures of different turbine technologies without further specifying the failed component. It showed a very good fit to the failure data. Nonetheless, when modelling the different WT components separately – instead of modelling the whole system – the models performed even better. This is related to the fact that the environmental conditions affect the reliability of the distinct components differently. This has then been demonstrated using a sub-set of the data containing only gearbox failures. This component is one of the most critical WT components and the model reveals that low temperatures, high maximum wind speeds and precipitation affect the gearbox failure behaviour negatively.

Finally, the model performance was tested on the failure data of a single wind farm. Different WT technologies behave differently when exposed to specific environmental conditions. Therefore, operators often use separate models for each technology or wind farm. Modelling the different technologies and their components separately has shown to lead to the best results and is recommended for future studies in this context. However, in order to analyse WFs separately, further work shall consider bigger data bases. Model improvements could also take into account noise in the input variables by using e.g. error-in-variable models. Correlated model

covariates were eliminated by the proposed model selection techniques. In order to test if the model accuracy can be further enhanced, further work, however, could consider introducing interaction terms between these covariates and subsequently reducing the number of total inputs using penalised regression.

In all three test cases the model performed very well. With this, it contributes to research in WT reliability modelling, by providing a robust technique for modelling WT failures including the effects of weather conditions.

6

A Bayesian Approach for WT Failure Detection

The previous chapter was concerned with modelling the failure behaviour of wind turbines and their components using probabilistic techniques. This was carried out based on the meteorological conditions on site, as well as some operational and turbine specific variables. The monthly number of failures were modelled.

In this chapter a method is introduced for failure detection on wind farm level based on these meteorological conditions. Instead of the number of failures per turbine, this model predicts the event of having one or more failures during a month within the whole wind farm given the surrounding weather conditions. Naive Bayesian classifiers are used to estimate the conditional probability of having one or more wind turbine failures in a wind farm. This is carried out using information on monthly wind turbine failures, turbine technology specific attributes, and the complex combinations of the weather variables introduced in the previous chapter. These include wind speed, wind gusts, precipitation (Rain), relative humidity and ambient temperature (Temp).

The trained models are then employed to predict the failure events of specific components in a wind farm during a prediction period of 36 months. Additionally to the failure events related to the whole turbine system, also component failures of the blades, gearbox, generator, main bearing, pitch and yaw system are predicted. Furthermore, an extensive sensitivity analysis is carried out to determine the most

suitable input variables for each model of six main WT components. With this, it can be determined which environmental variables affect the respective components the most. The results of this chapter have been published in Paper (VI) and are expected to significantly contribute to research in O&M and enable operators to predict and understand the conditional probabilities of having a failure under given meteorological conditions. Additionally to the results presented in Paper (VI), the performance of the naive Bayes classifier is compared to the performance of a logistic regression model, which is frequently used for similar purposes.

6.1 Data for Failure Prediction

In this chapter again a sub-set of the failure data presented in Chapter 2 is used. This comprises of historical failures obtained for 29 different wind farms during 1045 operational months. The wind farms operate in total 984 geared wind turbines, of which 638 are stall and 349 are pitch regulated. The machines are aged between 1 and 16 years with rated capacities of 300 kW to 2000 kW. In Figure 6.1 the meteorological and turbine specific model covariates are summarised.
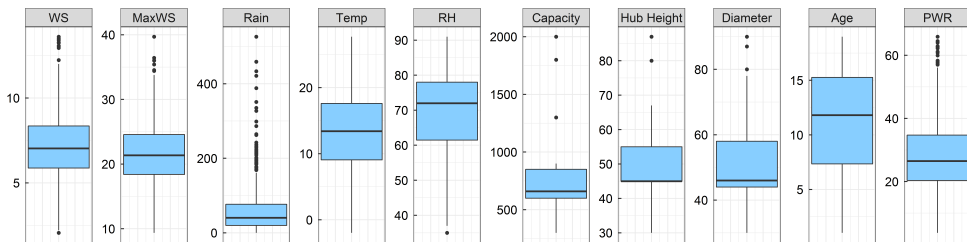
Figure 6.1. The environmental and turbine specific input data: WS (m/s), MaxWS (m/s), Rain (mm); Temp (°C), RH (%), Rated Capacity (kW), Hub Height (m), Diameter (m), Age (years), PWR (%).

The monthly mean measurements for the meteorological variables relative humidity (RH), ambient temperature (Temp), and the total monthly precipitation (Rain) were taken from close-by located weather stations or if available from the wind farms met mast. The monthly mean wind speed (WS) and maximum wind gust (MaxWS) measurements were taken directly from the wind turbines' SCADA systems. Furthermore, the measured monthly mean active power production in percent of the rated capacity (PWR), is included into the model in order to account

for the time in operation per month. This variable was taken from each turbine's SCADA system.

The failures of six main WT components, which have earlier been identified as the most critical components of the WT systems, are analysed. Additionally, the failures of the whole wind turbine system without distinguishing between the failed components will also be analysed. This will be referred to as 'all Failures' in the further.

In Table 6.1 the number of failure events per component over the whole observation period are displayed.

Table 6.1. Number of failure events per component in the data base used for this study.

Component	Blades	Gearbox	Generator	Main Bearing	Pitch	Yaw	**Total**
Failure events	232	230	104	14	6	18	**609**

6.2 Naive Bayes Classifier

In this part of the study a naive Bayesian classifier, [169], will be used, which is a special form of a Bayesian Belief Network (BBN). It is particularly useful for high dimensional input data and applies Bayes' theorem with a strong independence assumption among every pair of input features. Despite the fact, that this assumption does rarely hold true in reality, it has shown to perform surprisingly well in real world classification problems and reduces the complexity of the classification task significantly, [170]. This is partially due to the fact that this assumption does not seem to affect the posterior probabilities, especially when getting close to decision boundaries, [171]. Hence, in most cases, it does not impair the classification task at all. Furthermore, it allows to break down the multidimensional classification task into several one-dimensional ones, by being able to calculate the class conditional densities separately for each input variable, [171].

The conditional probability using Bayes' theorem is defined as:

$$Pr(y|\mathbf{x}) = \frac{Pr(y)Pr(\mathbf{x}|y)}{Pr(\mathbf{x})} = \frac{Pr(y)Pr(x_1, x_2, ...x_n|y)}{Pr(x_1, x_2, ..., x_n)} \quad , \tag{6.1}$$

with a vector of features $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, for which the posteriori probability $Pr(y|\mathbf{x})$ that $\mathbf{x}$ belongs to class y is determined. The class prior probability is given

by $Pr(y)$, the predictor prior probability by $Pr(\mathbf{x})$ and the the likelihood $Pr(\mathbf{x}|y)$. The 'naive' conditional independence assumption is defined as:

$$Pr(y|\mathbf{x}) = \frac{Pr(y) \prod_{i=1}^{n} Pr(x_i|y)}{Pr(\mathbf{x})} \quad . \tag{6.2}$$

The fact that $Pr(\mathbf{x})$ is constant given the input states that:

$$Pr(y|\mathbf{x}) \propto Pr(y) \prod_{i=1}^{n} Pr(x_i|y) \quad , \tag{6.3}$$

which results in the naive Bayes classifier model, given by:

$$\widehat{y} = \underset{j}{\mathrm{argmax}} \left\{ Pr(y) \prod_{i=1}^{n} Pr(x_i|y) \right\} \quad , \tag{6.4}$$

with the relative frequency $Pr(y)$ of class y in the training set. For classification problems, the likelihood $Pr(x_i|y)$ of the features is assumed to follow a Gaussian distribution:

$$Pr(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot e^{\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2} \right)} \quad , \tag{6.5}$$

where the parameters σ_y and μ_y can be estimated with maximum likelihood estimation (MLE), discussed in Chapter 5.

The response for the predictions is a boolean class variable indicating the event of having one or more failures in a wind farm during the respective month. In total 17 input covariates are considered, which include the environmental parameters presented in Section 6.1. All input features are discretised, as discussed in the following section.

As shown in the previous chapters, the meteorological conditions often have a delayed or even accumulative effect on the failure behaviour. Hence, in a similar manner to Chapter 5, the meteorological measurements taken throughout the previous month are also taken into consideration and are indicated with the suffix '.b', which stands for '.before'. In order to distinguish between the WT technologies, additionally, the turbine age, hub height, (pitch and stall) regulation, rated capacity and rotor diameter are added as model covariates. In Figure 6.2 the Bayesian Belief

Network (BBN) is exemplified with all input variables, to show how the covariates could be connected, however, the interconnections may vary for each component model.

The values for each environmental input variable are grouped into four equally-populated categories using quantile discretisation. Given the data of 29 wind farms and after eliminating observations with missing values, in total 716 monthly observations are available. The data are split randomly into training and testing sets, so that a training period of 680 months and a testing (prediction) period of 36 months are obtained. During the training phase, the conditional probabilities of having a failure event in the presence of the respective categories of each covariate are derived. Then, the trained naive Bayes classifier is used to predict failure occurrences in a wind farm during 36 months of operation.
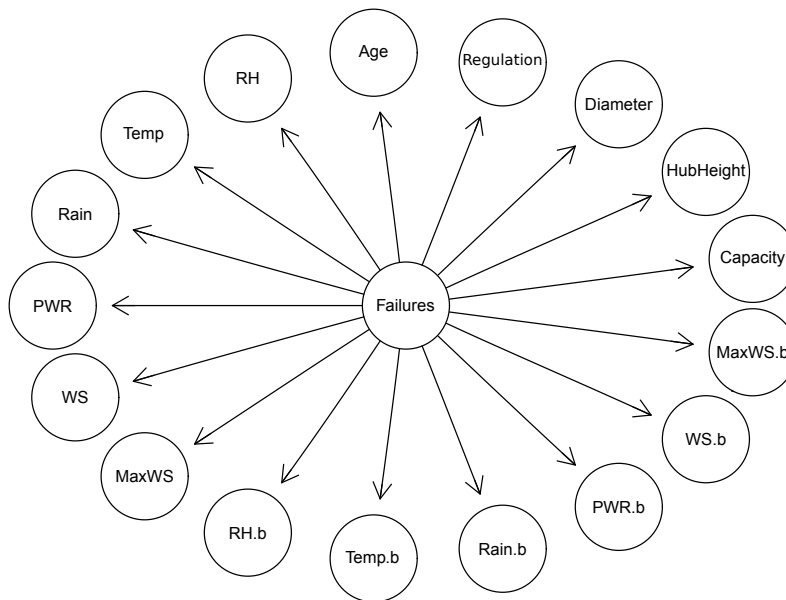

Figure 6.2. Example of a Bayesian Belief Network including all model covariates of this study.

6.2.1 Sensitivity Analysis

In the previous chapters it was shown that the distinct combinations of environmental conditions are not influencing the failure behaviour of all WT components

equally. Thus, for each component model a sensitivity study was carried our to find the most suitable model inputs, compromising between prediction accuracy and model complexity.

For the sensitivity study, separate models using all possible combinations of input covariates are trained and predictions are carried out on the test data set. Each combination can involve between 1 and 17 covariates (members). Hence, with $b = 17$ types of different variables in the data set and $m = \{1...17\}$ members possibly involved in each combination, the total number of models $_bC_m$ are trained for each component can be calculated by:

$$_bC_m(b, m) = \sum_{m=1}^{b} \binom{b}{m} = \sum_{m=1}^{b} \frac{b!}{(b-m)! \cdot m!} = \sum_{m=1}^{b} 2^{m-1} = 131071. \quad (6.6)$$

For evaluating the performance of the different classification models, each one of them is used to predict the classes of the response variable in the test data set. Subsequently, the confusion matrices of all predictions are compared. From these the information on the true positive (TP), true negative (TN), as well as false positive (FP) and false negative (FN) predictions can be obtained.

Finally, the best model will be selected based on four quality measures of the binary prediction. The sensitivity (true positive rate (TPR)) and specificity (true negative rate (TNR)) and the prediction accuracy (ACC) are defined as:

$$TPR = \frac{TP}{TP + FN} \quad ; \quad (6.7)$$

$$TNR = \frac{TN}{TN + FP} \quad ; \quad (6.8)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad . \quad (6.9)$$

Further, the Mathews correlation coefficient (MCC) is given by:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP(TN + FN))}}. \quad (6.10)$$

The TPR, TNR and ACC, can take any value on the interval $\{0, 1\}$, where 1 is the best case. The MCC can take any value between -1 to 1, where -1 represents a complete disagreement with the observed data, 0 is a random and 1 a perfect prediction. The MCC is often considered a more robust metric compared to TPR,

TNR and ACC, which are often found to introduce a certain bias to the model evaluation process. Additionally, the MCC works well for imbalanced data sets, as it avoids that the outcome is affected by the majority class, [172]. It can be seen in Section 6.1 that depending on the component, the input data are to some extend imbalanced (the number of failures recorded for each component is in some cases much lower than 50% of the overall number of observations). Hence, the MCC shall be used as primary evaluation metric in the results section.

6.2.2 Alternative Approaches

There are several alternatives to naive Bayes classifiers. Two of the most popular ones shall be compared to the herein proposed component models. These techniques will not be described fully at this point, as they are only used for comparison. Nonetheless, they will be explained in detail in Chapter 7, where they will be used for more sophisticated purposes.

Logistic regression models are often referred to as alternatives to naive Bayes classifiers. However, the fact that naive Bayes classifiers perform better when the size of the input data is limited, makes this technique very attractive for achieving the given objectives, [173]. In order to demonstrate this, a logistic regression model will be trained using a LASSO regularisation. As discussed in Chapter 5, the LASSO eliminates unimportant input variables and with this it achieves a similar goal as the sensitivity study carried out for the naive Bayes classifier. Another very popular method for classification problems are random forest classifiers. A random forest is composed of several weak decision trees, which as ensemble can perform better than a single tree. In this study a random forest with 80 trees is used.

6.3 Results and Discussion - Naive Bayes Classifiers

In this section, in order to find the best model for each component, at first, the results of the sensitivity study of the trained naive Bayes classifier models are presented. Subsequently, these results are compared to those obtained for the logistic regression models. Finally, the component models based on the naive Bayes classifier will be explained in detail including the conditional probabilities of their covariates.

6.3.1 Results Sensitivity Analysis

In Figure 6.3 the sensitivity versus specificity of the predictions obtained for each possible combination of input covariates with the failure model of the whole WT system ('all failures') are shown. The graphs are sorted by the number of possible members in each combination m. It is shown that the best compromise between sensitivity and specificity is obtained for mid-range m. Notably, the model using all input variable did not show the best performance out of all models. Nevertheless, with these kinds of graphs it remains a rather difficult task to determine the best performing model.
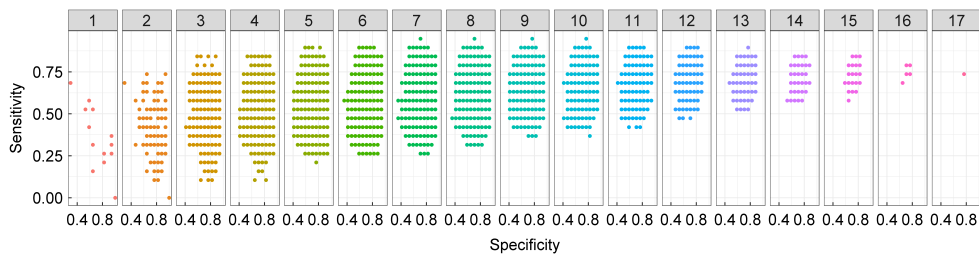
Figure 6.3. Sensitivity versus specificity for failures of the whole wind turbine system.

When looking at Figure 6.4, which shows the ACC and MCC for each m, this becomes clearer, however. It can be seen that the maximum values for both, ACC and MCC, were obtained for m between 6 and 11 members.

Figure 6.5 displays the MCC of the predictions for all components and all numbers of members per combination. One can see that the model for all WT failures obtains its maximum MCC value with six model covariates.

The model for the blades needs a minimum of $m = 9$ members in a combination
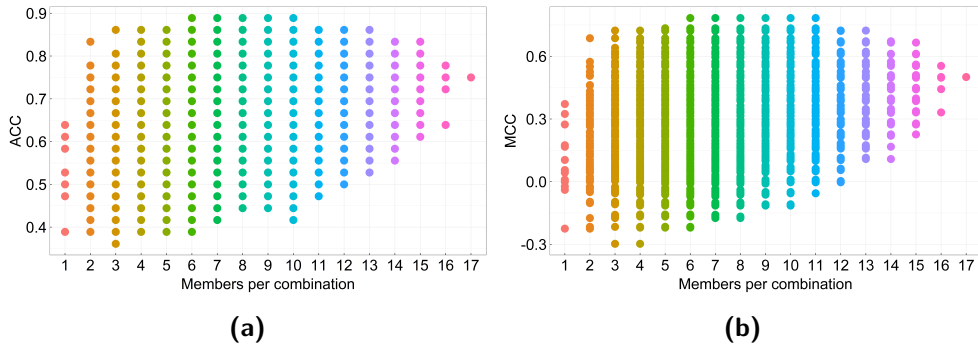
Figure 6.4. Metrics: (a) ACC and (b) MCC for failures of the whole wind turbine system.
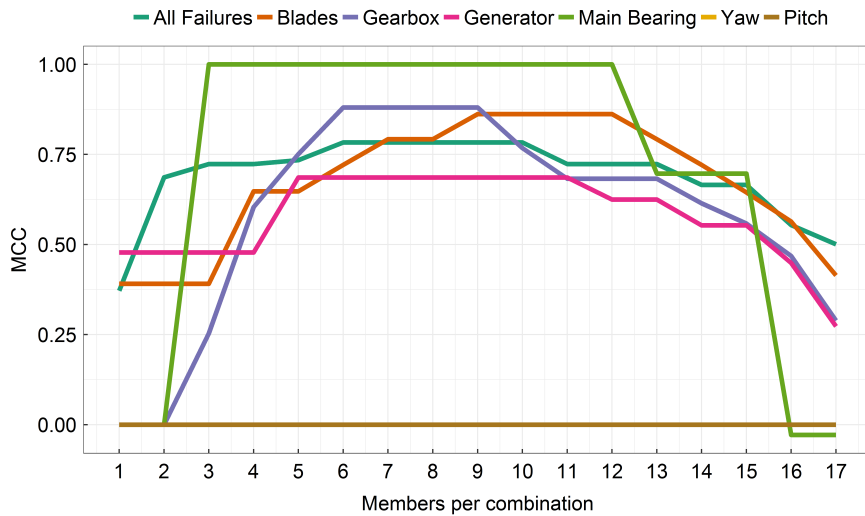
Figure 6.5. MMC of the predictions made for all components and all possible numbers of members per combination.

of covariates, while the gearbox needs $m = 6$, the generator $m = 5$ and the main bearing, at least, $m = 3$ members. The models for yaw and pitch system were no better than a random model, and thus, did not perform well in the predictions. This is likely to be due to the limited amount of failure events in the used data set for these components. Additionally, it could imply that the failure behaviour of these components is influenced more by other variables, which were not included into the models. The remaining component models, however, showed very high MCC values, confirming that these models serve as very good predictors.

As further displayed in Figure 6.5, many components showed a wide span of

possible m, for which the highest MCC values were achieved. It is worth pointing out that, as the objective of this chapter is to find the best compromise between model complexity and prediction accuracy, only the combinations of covariates with the minimum number of members for which the highest MCC value was obtained, will be considered in the further. Nonetheless, more covariates could possibly be added to the respective models, while having the same prediction performance.

Table 6.2. Results of the predictions for all naive Bayes component models with m_{best}.

Component	m_{best}	MCC	ACC	Sensitivity	Specificity	TP	FP	TN	FN
All Failures	6	0.783	0.889	0.842	0.941	16	1	16	3
Blades	9	0.862	0.944	0.800	1.000	8	0	26	2
Gearbox	6	0.880	0.972	0.800	1.000	4	0	31	1
Generator	5	0.686	0.944	0.500	1.000	2	0	32	2
Main Bearing	3	1.000	1.000	1.000	1.000	1	0	35	0
Pitch System	-	0	0.973	0	1.000	0	0	36	1
Yaw System	-	0	0.973	0	1.000	0	0	36	1

In Table 6.2 the evaluation metrics of the predictions obtained with all component models are summarised. The number of members per combination, which results in the best model, is denoted as m_{best}. It can be seen that only one false positive (FP) was obtained for all predictions. Generally, the models for the blades, gearbox, generator, main bearing and the whole WT system performed very well, having high sensitivities and specificities, as well as good detection rates.

As stated before, the number of failure events for each component in the available data set influences the accuracy of the predictions, as the models need a certain number of positive events during the training in order to learn the structure of the data. Hence, the models for the pitch and yaw system did not perform well and will be excluded from further discussion. Furthermore, the main bearing failure records only contained one event during the testing period. This was successfully predicted, however, the model covariates discussed in the next section might be biased with this restriction.

As shown in Table 6.3, in some cases there was little to no difference between the best and the second-best performing component model obtained for m_{best}. For the main bearing and gearbox models, the same performance was obtained for several combinations of input covariates. In any case, only one combination for each component will be analysed in detail.

Table 6.3. Comparing the best and the second-best performing naive Bayes component models.

Component	m_{best}	MCC - best model	MCC - 2nd best model	Comment
All Failures	6	0.783	0.734	–
Blades	9	0.862	0.792	–
Gearbox	6	0.880	0.880	7 models showed the same MCC
Generator	5	0.686	0.533	–
Main Bearing	3	1.000	1.000	11 models showed the same MCC

6.3.2 Comparing the Results obtained for the Naive Bayes and Alternative Models

In Tables 6.4 and 6.5 the results for the logistic regression model with LASSO penalisation and for the classification with random forests are shown, respectively. It can be seen that the naive Bayes classifier clearly outperformed both alternative approaches for most components. Only the logistic regression model for the main bearing shows the same values for all evaluation metrics. The discretisation of the input data had a positive effect on the performance of these two algorithms, which showed less accuracy when using the 'normal' input data set. This shall not be displayed here, but can be found in Appendix D. Being the best performing approach for the herein presented problem, in the further, only the models based on the naive Bayes classifier will be discussed in more detail.

Table 6.4. Results of the predictions using logistic regression.

Component	MCC	ACC	Sensitivity	Specificity	TP	FP	TN	FN
All Failures	0.460	0.722	0.895	0.529	17	8	9	2
Blades	0.322	0.750	0.400	0.885	4	3	23	6
Gearbox	0.468	0.806	0.800	0.806	4	6	25	1
Generator	0.156	0.833	0.250	0.906	1	3	29	3
Main Bearing	1.000	1.000	1.000	0.906	1	0	35	0
Pitch System	0	0	NA	0	0	36	0	0
Yaw System	0	0.972	0	1.000	0	0	35	1

Table 6.5. Results of the predictions using random forests.

Component	MCC	ACC	Sensitivity	Specificity	TP	FP	TN	FN
All Failures	0.353	0.667	0.895	0.412	17	10	7	2
Blades	-0.105	0.694	0	0.962	0	1	25	10
Gearbox	0	0.861	0	1.000	0	0	31	5
Generator	0.213	0.861	0.250	0.938	1	2	30	3
Main Bearing	-0.029	0.944	0	0.971	0	1	34	1
Pitch System	0	0	NA	0	0	36	0	0
Yaw System	0	0.973	0	1.000	0	0	36	1

6.3.3 Interpretation of the Resulting Models

This section discusses the predictions made with the naive Bayes classifier model that was determined as the best performing in the sensitivity study for the respective component, being displayed in Table 6.2.

In Figure 6.6 the observed and predicted failure events for each component during the prediction period of 36 months are shown. In general, a very good accuracy is obtained as most component failures were found.
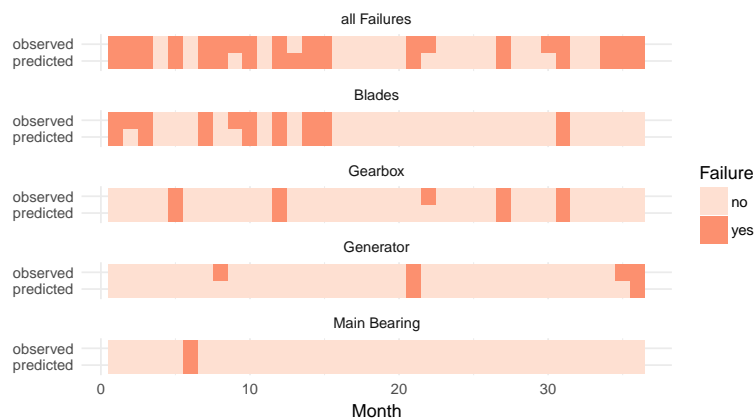

Figure 6.6. Monthly observed ans predicted failures (boolean).

Table 6.6 shows the levels to which the covariates were assigned to during the quantile discretisation. The conditional probabilities for having one or more failures of a certain component within the whole wind farm under the given values of the meteorological variable are shown in Figures 6.7 to 6.9 as well as Table 6.7.

Table 6.6. Assigned levels for the model covariates by quantile discretisation.

Level	WS	MaxWS	Rain	Temp	RH
SI-unit	m/s	m/s	mm	°C	%
low	2.1-5.7	9.4-18.4	0-20.7	(-2) - 8.6	35-59.9
mid low	5.8-6.8	18.5-21.6	20.8-35.5	8.7-13.6	60-69.9
mid high	6.9-8.1	21.6-25.3	35.6-70.3	13.7-17.9	70-76.9
high	8.1-13.6	25.3-40.5	70.4-526	18.0-27.7	77-91

Level	Capacity	Hub Height	Diameter	Age	PWR
SI-unit	kW	m	m	years	%
low	300-600	30-35	30-42	1-6.9	3.8-19.2
mid low	660 -850	43-55	44-58	7-11.6	19.3-25.7
mid high	900-1300	57-67	59-78	11.7-15.1	25.8-33.6
high	1800-2000	80-87	80-90	15.2-19	33.7-64.5

The figures display to what extend each covariate level contributed to the overall conditional probability of the respective covariate. As the minimum m for which the highest MCC value was obtained is used, these figures do not include all the model covariates that could possibly affect the failure behaviour.

In the following each covariate will be analysed separately for each component model. These findings are consistent with and extend those of Chapter 3, which showed a similar behaviour of the environmental conditions before the respective component failures. However, the reader shall keep in mind that the failure events occur under the condition of a combination of all of these covariates at specific levels.

- **All failures:** Figure 6.7a shows that the conditional probabilities for having a WT failure (without distinguishing between the failed components), are much higher for stall regulated turbines than they are for pitch regulated ones. Additionally, slightly elevated power production and relative humidity also contributed to higher probabilities of failure. Furthermore, it was found that WTs with lower rated capacities and smaller diameters are affected more often. This is not necessarily consistent with Chapter 2, where it was shown that pitch regulated turbines have higher failure rates. However, this could be due to the

fact that only a sub-set of the data introduced in Chapter 2 is used in this chapter.

- **Gearbox:** In Figure 6.7b it is shown that colder temperatures during the month before the failure as well as throughout the failure month, higher power production and higher mean wind speeds increase the probability of having a failure of this component. Furthermore, lower hub-heights showed to be affected more often.

- **Generator:** The conditional probabilities of the generator model indicated that the failures are likely to occur under higher relative humidity and mean wind speeds. Turbines of lower rated capacity were affected more often. During the month before failure slightly higher temperatures and lower maximum wind speeds resulted in a higher probability of failure. In literature, e.g. [174], it is shown that generator failures often occur under highly variable weather conditions, as for instance during transition periods from summer to winter. This might explain the difference of temperature and wind speed indicated by the model.

- **Main Bearing:** The compromise of best performing model and prediction accuracy for the main bearing failures only considered 3 covariates. According to these, younger turbines with higher rated capacity operating under wind conditions with low gust speeds, showed higher failure probabilities.

- **Blades:** High precipitation, high relative humidity during the month before failure, slightly higher wind speeds throughout the failure month as well as the month prior to the failure, resulted in higher probabilities of blade failures. Furthermore, marginally older turbines were affected.
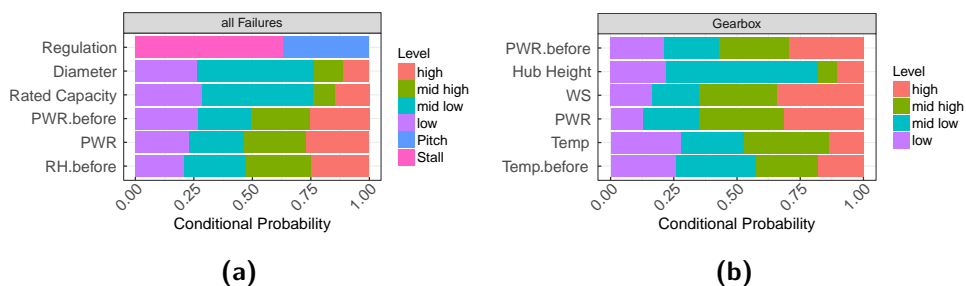

Figure 6.7. Model covariates and conditional probabilities for (a) all failures, (b) gearbox.
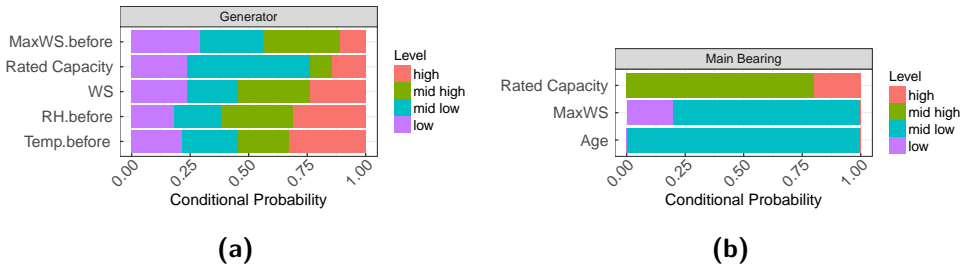
(a)   **(b)**

Figure 6.8. Model covariates and conditional probabilities for (a) generator, (b) main bearing failures.
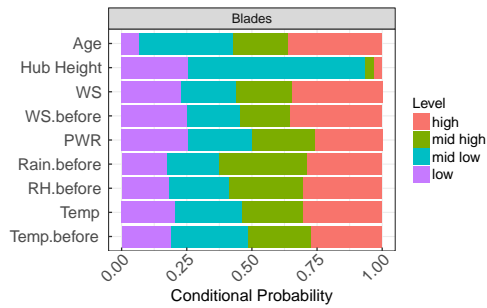

Figure 6.9. Model covariates and conditional probabilities for blade failures.

Table 6.7. Conditional Probabilities for all models and input variables.

		low	mid low	mid high	high
all Failures	Diameter	0.263	0.500	0.127	0.111
	Rated Capacity	0.285	0.475	0.095	0.146
	PWR	0.231	0.234	0.266	0.269
	PWR.before	0.269	0.228	0.250	0.253
	RH.before	0.209	0.266	0.278	0.247
	Regulation	Stall = 0.633		Pitch = 0.367	

		low	mid low	mid high	high
Gearbox	Hub Height	0.219	0.603	0.075	0.103
	WS	0.164	0.185	0.308	0.342
	PWR	0.130	0.219	0.336	0.315
	PWR.before	0.212	0.219	0.274	0.295
	Temp	0.281	0.247	0.336	0.137
	Temp.before	0.260	0.315	0.247	0.178

		low	mid low	mid high	high
Generator	WS	0.236	0.218	0.309	0.236
	MaxWS.before	0.291	0.273	0.327	0.109
	RH.before	0.182	0.200	0.309	0.309
	Temp.before	0.218	0.236	0.218	0.327
	Rated Capacity	0.236	0.527	0.091	0.145

		low	mid low	mid high	high
MB	Age	0.000	1.000	0.000	0.000
	MaxWS	0.000	1.000	0.000	0.000
	Rated Capacity	0.236	0.527	0.091	0.145

		low	mid low	mid high	high
Blades	Age	0.066	0.360	0.213	0.360
	WS	0.228	0.213	0.213	0.346
	WS.before	0.250	0.26	0.191	0.353
	PWR	0.257	0.243	0.243	0.257
	Rain.before	0.176	0.199	0.338	0.287
	RH.before	0.184	0.228	0.287	0.301
	Temp	0.206	0.257	0.235	0.301
	Temp.before	0.191	0.294	0.243	0.272
	Hub Height	0.257	0.676	0.037	0.029

6.4 Conclusions for the Failure Detection with Naive Bayes

In this chapter the failure events of six main WT components were predicted with a naive Bayes classifier based on meteorological and operational conditions as well as WT technology specific covariates. In order to find the best performing model for each of the six main components, an extensive sensitivity analysis was carried out, compromising between model complexity and prediction accuracy. It was shown that each component model is driven by different input covariates. The presented techniques are able to detect failure events in wind farms during 36 operational months reliably. Even with a certain degree of multicollinearity among the environmental input variables, as well as the independence assumption of the naive Bayes classifier, it performed very well. Most failures were detected by the models. Additionally, the effects of the environmental variables on the failure behaviour of the different WT components were discussed. These mainly match with the findings of Chapters 3 to 5. Moreover, the performance of the naive Bayes classifiers and the logistic regression models were compared, and it was demonstrated that the naive Bayes classifiers outperformed the logistic regression in all cases except for the main bearing, where both showed the same results. However, as for this component only very limited failure data was available, this might not be representative.

Hence, the herein proposed model based on the naive Bayes classifier can aid wind farm operators to detect and predict WT component failure events in their wind farms based on the environmental conditions at the wind farm site, and with this, lower the cost related to O&M actions.

Future studies should investigate the influence of different data pre-processing for the naive Bayes classifiers, as the form of input discretisation can affect the model outcome. Also, a kernel density estimation (kernel based naive Bayes) could be used to estimate the probability density functions of the input variables to the classifiers, which might enhance the predictions. However, this requires extensive input data, which was not available for the present study. Additionally, other meteorological variables should be included. Especially, the turbulence intensity is expected to have a significant impact on the failure behaviour of certain components. Also, data should be used that contain more failure information for the main bearing, pitch and yaw system. These were represented with a low number of failures in

the data set of this study. This results in highly imbalanced classes ('failure', 'no failure') and it was not possible to generate reliable prediction models for these components. Hence, in further work, the underlying class distribution should either be adjusted, or more failure events should be considered. Finally, future research could also focus on different failure modes of each component, as it is supposed that each failure mode is provoked by different combinations of meteorological and operational conditions.

After having proposed a well functioning technique for failure detection on wind farm level based on the site specific environmental conditions, in the subsequent chapter, approaches for failure prediction using operational and component specific condition monitoring data will be introduced.

7

Data-Driven Fault Prediction based on SCADA and CMS Data

In the previous Chapter 6, failure detection on wind farm level was carried out using naive Bayes classifiers. The presented method has shown to perform well when taking into consideration the weather conditions in the wind farm. This chapter will focus on data-driven fault detection in wind turbine components using SCADA and CMS data. Both systems collect a remarkable quantity of data, which contain important information on the health status of the WT components as well as on operational and external conditions. Analysing these vast amounts of data and deriving useful information is a very costly and difficult task. In order to use these data for fault detection purposes, appropriate data mining and machine learning techniques have to be identified. Furthermore, the relationships between the data generated from both systems need to be understood to eliminate unnecessary information and to select appropriate inputs for fault detection algorithms.

Some studies have compared the use of SCADA and CMS data for failure detection, [175, 176]. Furthermore, the synergies between different SCADA variables have been analysed in [177]. Few efforts have been made to join SCADA and vibration data, [178, 179]. These are, however, limited to time series measurements of vibrations obtained from the SCADA system. Hence, to date there has not been any approach combining SCADA and CMS data for enhanced WT condition monitoring, failure detection and remaining useful lifetime estimation, see also [7].

In this chapter SCADA data and CMS vibration measurements are merged and the relationships between both sources are analysed. Then, several applications suitable for condition monitoring based on both data sources are presented. Furthermore, a tool for failure detection in wind turbine components is presented, which compares the vibration data on wind farm level and triggers alarms when a component is at risk. The results presented in the following have been published in Paper (VII).

7.1 Background: Condition Monitoring of Wind Turbines

The objective of condition monitoring is to detect faulty or degraded components as early as possible in order to avoid unplanned downtimes of the wind turbines.

Condition monitoring systems have been used in several industries over many years, however, only recently WF operators have started to install these systems in their wind turbines. This might be due to the fact that installing CMS is quite expensive and it might not be financially feasible for all wind farms, [176, 180]. Nonetheless, in many cases it could be an attractive investment for wind farms, where the benefits of early failure detection can outweigh the initial cost of installation. Contrarily, SCADA systems are nowadays installed in nearly all operating wind turbines. Their data are often directly available to the WF operators, making them an attractive source for condition monitoring.

7.1.1 Condition Monitoring based on SCADA Data

Recently, research has increasingly been focusing on finding solutions for condition an performance monitoring using exclusively SCADA data. In that manner, additional cost for installing and operating CMS, as well as analysing the obtained data are avoided. Common variables monitored by SCADA systems are the wind conditions (e.g. wind speed, direction, turbulence intensity), operational parameters (for example power output, pitch angles, rotational speed, etc.), as well as operating temperatures and in some cases vibrations of certain components. These measurements generally have a sampling frequency of roughly 1 Hz, however, in most cases ten-minute average statistics are saved. Additionally, the SCADA systems trigger alarms if certain pre-defined thresholds for the monitored variables are exceeded.

Literature in condition monitoring with SCADA data has focused on the use of alarm logs, [181–184], as well as the development of damage models derived from SCADA measurements, [185, 186]. Most available literature, however, is concerned with the use of temperature measurements for fault detection in drive-train components. For this, clustering approaches, [187], and trend analyses, [175, 180], to discover similarities in the signals have been applied. Moreover, normal behaviour modelling has found several applications, which have lead to very promising results. Various normal behaviour models have been developed using adaptive neuro-fuzzy inference systems, [188], artificial neural networks, [189, 190], or other regression techniques, [191].

7.1.2 Condition Monitoring based on CMS Data

Condition monitoring based on CMS data has a long history frequently used in machinery throughout several different industries. In wind turbines, the installed systems usually consist of acoustic, oil level, strain or vibration sensors, which are spread over different WT components, [192]. In this chapter, the term CMS data will exclusively refer to vibration data. The latter is the most frequently obtained form of CMS measurements, due to the comparably low installation cost of the required equipment, [176].

The vibration measurements need to be pre-processed before being analysed further, as the raw time series signals are not very useful for failure detection. Frequently used processing techniques can be classified as time-domain analysis (e.g. Hilbert transform, statistical analysis, envelope analysis), frequency domain techniques (e.g. Fast-Fourier-Transform (FFT), Cepstrum analysis), as well as time-frequency techniques (e.g. wavelet-transform), [193].

The most frequently applied pre-processing technique is the FFT, which transforms the time domain signals into the frequency domain, and represents the vibration measurements in a single frequency spectrum. This spectrum is then analysed by experts, whereas certain spikes and harmonics can directly be related to component degradation or faulty conditions, [194, 195]. One limitation of FFT, however, is that it can only be used for stationary signals. Non-stationary signals might result in indistinct FFT results. For machinery operating in highly non-stationary conditions (e.g. wind turbines), several techniques have been developed to ensure stationarity before carrying out the FFT, [196, 197].

Envelope analysis is another often applied signal processing technique and is able to bring forward fault frequencies that are not necessarily shown in frequency spectra generated by FFT, such as for example shock impulse repetition and their harmonics, [198]. The enveloping process involves the application of a band pass filter to the time domain signal, which centres on the desired frequency energy region. The repetition rate is extracted from the filtered time signal using amplitude demodulation. When applying a FFT to the enveloped signal one can derive the fault specific characteristic 'impact frequencies' and their modulations, e.g. sidebands.

Another commonly used frequency-domain technique for vibration data pre-processing is the Cepstrum analysis, [199, 200]. It performs an inverse Fourier transform of the logarithmic power spectrum and can be used to analyse the lower level harmonics of the logarithmic power spectrum, [53, 194, 201]. Thus, it is very similar to auto-correlation analysis, which in turn carries out the inverse Fourier transform on the regular power spectrum.

By identifying harmonics and sidebands in the frequency spectra of the vibration data, specific fault frequencies of rotating machinery can be determined. While different harmonic families can be found with Cepstrum analysis, sidebands can be obtained by performing Envelope analysis, [202]. Thus, by combining FFT, Envelope and Cepstrum analysis as well as statistical methods such as the root mean square amplitude (RMS), various forms of failures can be found. This is essential for failure detection and many commercially available solutions rely on combinations of these techniques for fault diagnosis, [203]. Nonetheless, the results obtained from these tools always need to be evaluated by an expert, deciding on whether the analysed spectra indicate a faulty condition or not. By deriving features like side band energy and kurtosis, [204], applying deep learning networks, [205], or classifying wind and rotor speeds for monitoring purposes, [206], some authors have tried to automate failure detection with CMS data.

Comprehensive literature reviews concerning condition monitoring techniques for wind turbines based on vibration analysis are presented in [53, 194, 207]. It has been found that there is still a serious lack of fully automated, generic approaches for fault detection with CMS data requiring little to no human interaction.

7.2 Methodology

In Figure 7.1 the objectives of this chapter are summarised. The global aim is to demonstrate how SCADA and CMS data can be coupled for condition monitoring and fault detection and to build a basis for further research in the field. This work is intended to explore different possible solutions that can be achieved with the merged data.
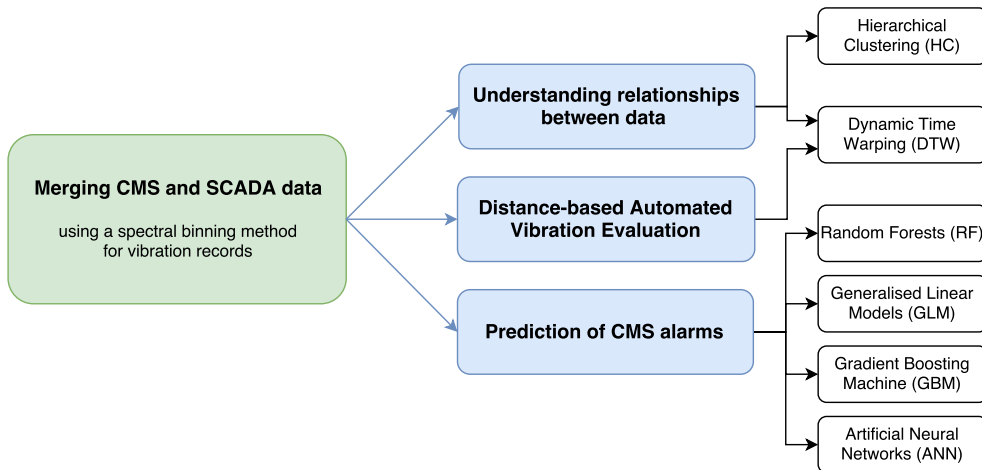

Figure 7.1. Application of the merged data for different purposes.

Firstly, a method is introduced to merge the 10-minute mean SCADA data with the CMS vibration data. These merged data serve as input to three different tasks, which are not necessarily building upon each other and should be treated as independent. These are:

A *Understanding relationships between data:* compare the different signals in the merged data set and find similarities and/or relationships.

B *Distance-based automated vibration evaluation:* an algorithm is developed for automated failure detection on wind farm level based on CMS data.

C *Prediction of CMS alarms:* with this the possibilities of triggering earlier warnings than the CMS system are evaluated. Different well known data-driven learning algorithms are used.

- *Count of alarms:* predict the number of CMS alarms only with SCADA data. This is done to examine the possibility of substituting condition monitoring systems.

- *Time dependent probability of alarm:* predict the probability of experiencing a failure event in a WT. This can be used as an early warning system based on only SCADA data.

- *Alarm time shifting:* this represents an alternative approach for early warnings. The aim is to predict the alarms in a classification set-up, investigating also on the benefit of adding CMS data to the prediction.

7.2.1 Data

The data used in this chapter were recorded over a period of four years (January 2013 to December 2016) at an onshore wind farm in Denmark operating 13 turbines (referred to as T01 to T13). The turbines have rated capacities of 2.3 MW and are three bladed and pitch regulated machines. Different data sources have been considered and consist of:

- *Component failures:* Both, the SCADA and CMS systems indicated severe main bearing failures in three turbines (T01, T03 and T08), which had to be replaced causing significant downtime. There were no further confirmed failures or major replacements.

- *SCADA data:* The SCADA data consist of 155 channels with 10-minute average values measured at various locations of the turbines.

- *SCADA alarm logs:* These contain high temperature alarms related to the main bearing generated automatically when a pre-defined temperature value was surpassed.

- *CMS vibration data:* Consisting of multiple Envelope (Env), FFT, Cepstrum and RMS records in irregular sampling intervals. For better readability in the further 'spectra' and 'frequency' will be used to describe CMS records, however, these also refer to the respective equivalents ('cepstra' and 'quefrency') in Cepstrum analysis. Each record is named after its characteristic sampling frequency and bandwidth, e.g. 'Env1000' or 'FFT1000', which indicate an Env and FFT with frequencies between 0 and 1000 Hz. The vibrations of seven WT components were measured:

  1. Generator Drive End (GDE),
  2. Generator Non-Drive End (GNDE),
  3. High Speed Shaft (HSS),

4. Intermediate Speed Shaft (IMS),

5. Main Bearing (MB),

6. Planetary Stage (PS),

7. Tower top acceleration, which only contained RMS.

In order to ensure quasi-stationary operating conditions, the vibration measurements were taken in seven different active power intervals. It will be shown later which of these active power intervals is the most suitable for the herein presented approaches.

- *CMS alarms per component*: Triggered by the commercial condition monitoring system, as soon as the vibration level of the component indicated by the Env, FFT, Cepstrum or RMS exceeded a certain threshold. In the case of main bearing failures this was mainly indicated by the RMS.

7.2.2 Merging SCADA and CMS Data

In order to set up a uniform data base for further application of the data driven prediction algorithms, the CMS and SCADA data need to be processed. Both are available in different temporal resolutions, and while SCADA data are frequently given as 10-minute average values, CMS measurements are obtained once per day, week or month.

Figure 7.2 shows an example for typical Envelope spectra (on log-log axes) obtained before and after a main bearing failure. For examining the health condition of a component one has to track down how the amplitudes of different spectral peaks, i.e. fault frequencies, side-bands and harmonics, change over time. This is normally done manually by experts and is a time consuming process that involves a great amount of experience. Component deterioration is usually indicated by a specific trend in the vibration behaviour, expressed by e.g. rising amplitudes of specific frequency ranges across subsequent measurements. This can be observed in Figure 7.2, where the amplitudes over certain frequencies increase steadily until the failure occurs in April 2016.

CMS spectra obtained from Envelope, FFT and Cepstrum analysis can contain a high number of data points. Hence, including them entirely in machine learning algorithms can lead to unnecessarily high computational effort and can affect the

accuracy of the results. Thus, reducing the complexity of these input data is necessary.
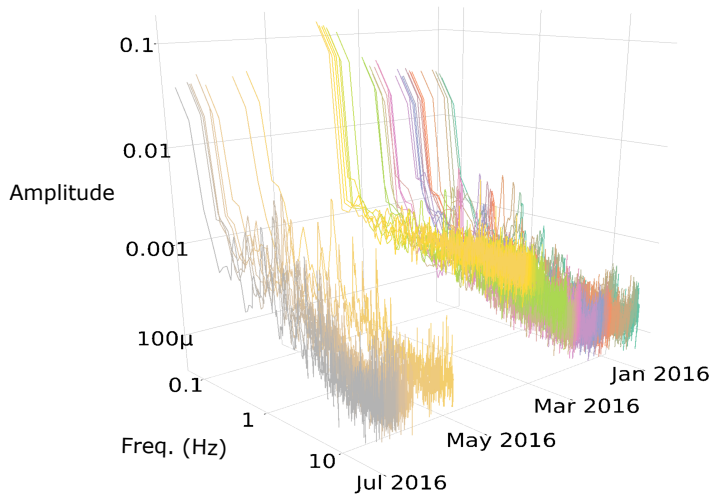

Figure 7.2. Example of Envelope records in time-frequency domain (log-log axes) for manual analysis by experts. The amplitudes of certain frequencies rise significantly prior to a failure (in April 2016) .

For this, a binning approach was developed, which splits each Envelope, FFT and Cepstrum record into bins of frequencies and takes the integral over the frequency and amplitude recorded in each bin. In Figure 7.3 this process is displayed.
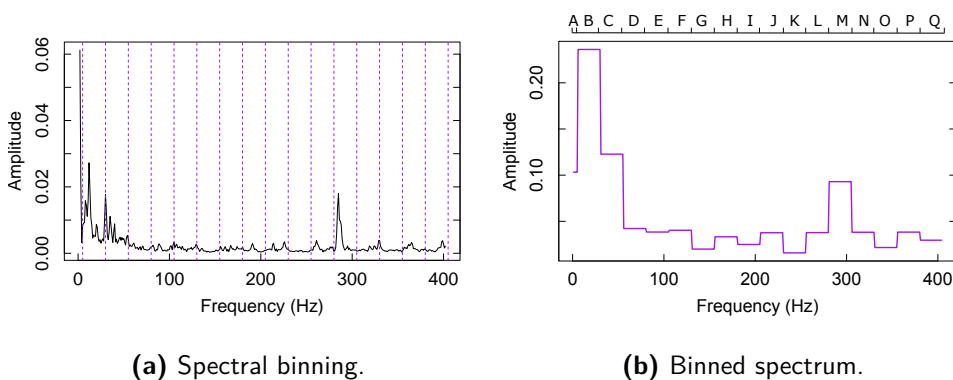

(a) Spectral binning.

(b) Binned spectrum.

Figure 7.3. The binning process developed for this chapter.

After having carefully evaluated the different CMS records (FFT, Cepstrum and

Envelope), the total number of bins was set to 17, which are labelled in alphabetical order *A,B,...,Q*. This number of bins enables to capture the different peaks in the spectrum that can indicate the fault frequencies, harmonics and side-bands, while reducing the dimensionality of the spectrum significantly.

This is illustrated in Figure 7.4, where the Box-Whisker plots of all binned Envelope and FFT spectra obtained during healthy and faulty operation of one WT main bearing are compared. It is shown that the variations and mean values obtained in the healthy and faulty states differ substantially from each other. Furthermore, it can be seen that the changes occur in different regions of the binned spectra. This is due to the fact that both techniques, FFT and Envelope, treat the raw signal differently. Hence, a significant dimensionality reduction is achieved by applying the proposed binning approach, whilst preserving the important properties of the spectra for distinguishing between healthy and faulty states.
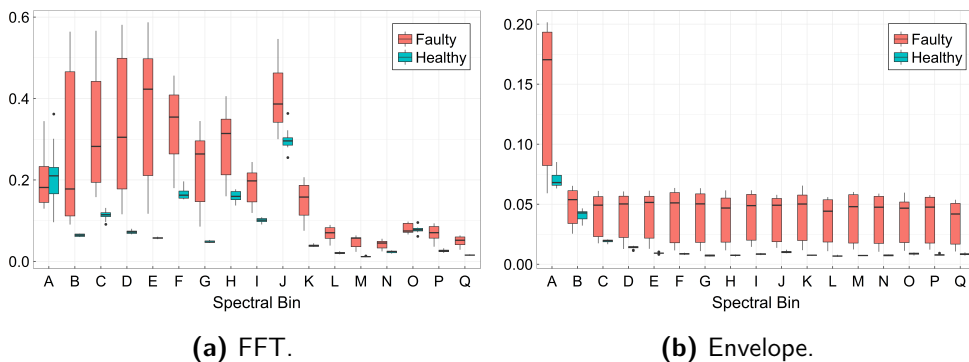

(a) FFT.   **(b)** Envelope.

Figure 7.4. Comparing binned FFT and Envelope records in a healthy and faulty state of a wind turbine main bearing.

In a final step, the temporal resolution of the CMS and SCADA records need to be matched. As fewer CMS measurements are available, they will be adjusted to match the resolution of the SCADA measurements, by keeping the CMS values constant (i.e. it is assumed that they do not change) if no measurements are available and are updated as soon as there are new measurements.

The merged CMS and SCADA data were then standardized: centred to a mean equal to 0 and divided by their standard deviation.

7.2.3 Understanding Relationships between Data

The relationships between SCADA and CMS data are analysed in order to understand how similarities change in the case of a component failure. With this it can be seen which signals are appropriate for data-driven failure detection and which ones can be omitted. Here, a simple correlation analysis would not be effective, due to the irregular temporal resolution of CMS data and the large number of signals. Instead, Hierarchical Clustering (HC) and Dynamic Time Warping (DTW) are applied.

HC is a frequently used unsupervised machine learning tool to group data bottom-up (agglomerative) or top-down (divisive), [208]. In this section an agglomerative HC is used with the average linkage method and Euclidean distances.

DTW was originally developed for speech recognition, [209] and is a method to measure similarities in two time-dependent signals. These signals may have characteristics that are out of phase, and DTW re-aligns them by finding the best 'warping' path that results in the lowest sum of pairwise distances. The warping path is required to be monotonic and boundaries might be set to limit the adjustment. For this task, DTW is applied with a Euclidean distance and a maximal adjustment window of two weeks. Groups of data are compared by a 2-dimensional DTW distance that stretches all signals of each group together.

7.2.4 Distance-based automated vibration evaluation

In this section, the methodology for the herein developed distance-based automated vibration evaluation (DAVE) framework is described. It uses CMS vibration data and is intended as a generic tool for WT health assessment based on the fact that it is unlikely that all turbines in a farm experience a failure of the same component at the exact same moment. This tool is especially useful for automated failure detection and as early warning system on wind farm level. By calculating the pairwise distances between the measured vibrations of the components of all WTs, it detects deviations of the measured vibrations of a 'faulty' WT component from 'healthy' operation.

At first, this tool needs to be set up for the wind farm during normal operation. This procedure involves analysing the CMS records for each component separately. The initial configuration is then used for the actual failure detection with online

measurements obtained for the same component. In Figure 7.5 the automated failure detection process is visualised.
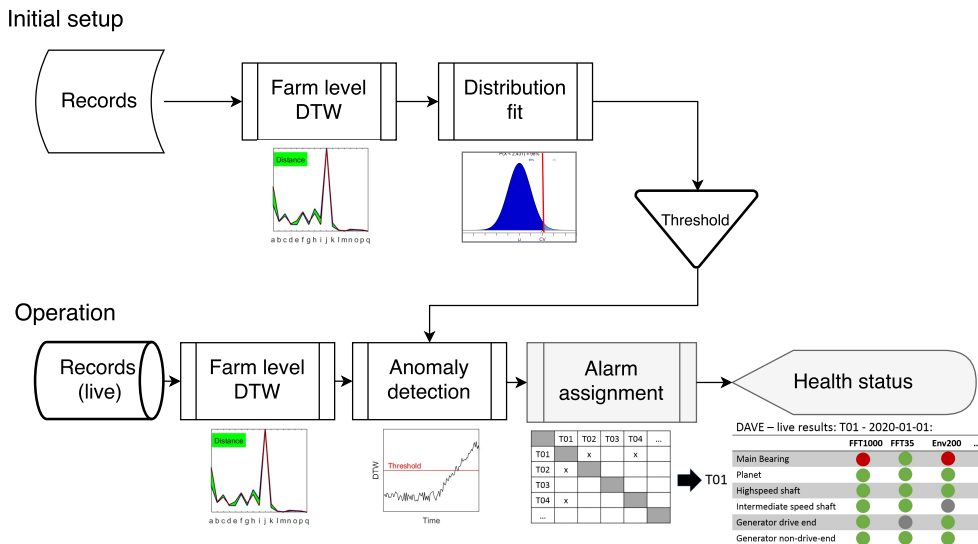

Figure 7.5. Distance-based Automated Vibration Evaluation (DAVE) workflow .

This procedure involves the following steps:

1. *Initial setup:* The thresholds to distinguish abnormal from normal behaviour/operation have to be defined. This could be accomplished either by consulting expert judgement or by applying data-driven methods. The latter will be used for setting up the tool. It uses the measured vibrations during 'healthy' operation and includes two sub-tasks:

    (a) *DTW calculations on wind farm level:* At each time step t, the spectra of the vibration measurements of two turbines are compared using DTW distances (e.g. the vibration records of turbine T01 are compared to those of T02, T03, etc.). Hence, for each point in time and each combination of WTs, the similarity of two binned vibration spectra is assessed. The DTW algorithm is used to identify similar trends in the different vibration spectra. By virtue of its flexibility, DTW can achieve this even if the spectral peaks are at slightly different frequencies. To ensure that similar fault frequencies are captured, the warping window is limited to a maximum of two neighbouring bins (one at each side of the actual bin). This allows a slight shift of peaks without considering peaks at opposite ends of the spectra as corresponding.

143

(b) *Distribution fit and threshold definition:* By fitting a distribution to the distances and setting a threshold for the healthy condition, the anomalies are defined for the subsequent on-line application. The threshold for this is a critical value, which is set to a certain percentile of the distribution. All values higher than this critical value are flagged as anomalous subsequently.

2. *Operation:* The on-line data recorded during operation are then used for the following:

(a) *DTW calculation on wind farm level:* Analogue to step 1 (a), for each point in time all pairwise distances between the vibration measurements of all turbines are calculated. With this a distance matrix containing all turbines is created.

(b) *Anomaly detection:* Based on the threshold defined during the initial setup, anomalies within the pairwise distances are identified.

(c) *Alarm assignment:* To determine whether an alarm has to be triggered and to which turbine it corresponds, the number of anomalies from pairwise comparisons is counted for each turbine (at each time step). If the count is larger than one, an alarm is issued. The alarm is assigned to the turbine with the highest count.

(d) *Health status:* With the above procedure, alarms are generated for each turbine, component and CMS record. The overall health status can be visualised with a dashboard summarising all alarms.

The DAVE framework is exemplified in this chapter for the main bearing, for which several failures were recorded in the data base.

7.2.5 Algorithms used for the Prediction of CMS alarms

Usually CMS alarms are triggered significant time before the component actually fails. Being able to predict these CMS alarms could lead to considerable benefits in terms of early failure detection. The third objective of this chapter is to predict the CMS alarms using classic machine learning algorithms. Here, three different approaches are investigated. (1) The count of CMS alarms is modelled. (2) The probabilities of having an alarm over time are obtained. (3) The alarms are predicted in a classification set-up with an alarm time shifting approach. As there have been

observed failures in the main bearings (MB) of three turbines, the focus here is only on MB alarms.

For this, four regression and classification techniques are used: Generalised Linear Model (GLM), Random Forests (RF), Gradient Boosting Machines (GBM), and Artificial Neural Network (ANN). These algorithms will be shortly explained in the following.

Generalised Linear Model (GLM) A GLM, [210], is a generalisation of the ordinary linear model and allows for having an error distribution other than the Gaussian distribution. Hence, depending on the modelled response variable, an appropriate distribution has to be chosen. Assuming a distribution of the exponential family, the probability density function can be written as:

$$f(y; \varphi, \vartheta) = e^{\frac{y\varphi - b(\varphi)}{\vartheta} + c(y,\vartheta)} \quad , \tag{7.1}$$

with the response variable y, parameters φ and ϑ, as well as the functions $b(\cdot)$ and $c(\cdot)$, which determine which distribution of the exponential family is used. The conditional mean and variance of the distribution are given by: $E[y_i|x_i] = \mu_i = b'(\varphi_i)$ and $Var[y_i|x_i] = \vartheta b''(\varphi_i)$. The link function, which denotes the dependence of the conditional mean on the regressors x_i, is given by:

$$g(\mu_i) = x_i^T \beta \quad , \tag{7.2}$$

where β is a vector of regression coefficients. For modelling the count of CMS alarms a Poisson distribution will be used. It was introduced in Eq. 5.25 in Chapter 5. For better readability, however, it will be given again at this point. The probability density function of the Poisson distribution is defined as:

$$f(y; \mu) = \frac{\mu_y}{y!} \cdot e^{-\mu} \quad . \tag{7.3}$$

The dispersion parameter is equal to one $\vartheta = 1$ hence, the mean and the variance of the distribution are identical. In a Poisson distributed GLM, a log-linear relationship between the linear predictor and the mean of the distribution is used, which results in the logarithmic canonical link function:

145

$$g(\mu_i) = log(\mu_i) \quad . \tag{7.4}$$

For modelling the time dependent probability of having a CMS alarm a logistic regression is applied. This is achieved by using a binomial distribution for the GLM in combination with a logit-link function. The probability density function of the binomial distribution is given by:

$$f(y|q) = \binom{q}{y} \pi^y (1-\pi)^{q-y} \quad , \tag{7.5}$$

with parameter π, and y is the number of successes when running q trials. The logit-link function can be defined as:

$$g(\pi) = logit(\pi_i) = \frac{\pi_i}{1-\pi_i} \quad . \tag{7.6}$$

In order to eliminate redundant covariates, the GLMs will be used with a LASSO penalised likelihood estimation, which was explained in Chapter 5. The estimated standardised coefficient magnitudes serve then as indicator for the importance of each input variable.

Random Forests (RF)   In RF, [211, 212], an ensemble of single decision tree predictors is trained using a so called 'bagging' method. Each decision tree predictor is considered a weak learner, however, a combination of these weak learners results in a strong learner, and thus, in a better predictor. RFs are very often used in machine learning, as they can handle both, classification and regression problems. One of the major advantages of RFs is that they do not have problems with over-fitting, by selecting a random set of predictor variables for each tree node, as explained below.

In Figure 7.6 the RF classifier is exemplified for three decision trees. At first, subsets of the entire data base are randomly generated using a Bootstrap Aggregation (bagging) algorithm, [213]. The number of sub-sets depends on the number of trees that are to be grown. Each sub-set should approximately include 2/3 of the whole data set, [212]. The remaining 1/3 are called out-of-bag samples. Then, full trees for each random sub-set are grown, while the out of bag samples serve to determine the unbiased estimate of the error. For growing each node of

each tree, a set of d predictor variables are randomly chosen out of all predictors, whereas d should be significantly lower than the total number of predictors D in the data ($d << D$). Usually, for regression problems $d = D/3$, and for classification problems $d = \sqrt{D}$ are chosen. At the next node of the tree, a new set of d predictor variables are chosen.

After having grown all decision trees, the random forest is used to predict the response variable. This is exemplified in Figure 7.6 for a classification problem. Here, each single decision tree makes a prediction on the resulting class variable. The final prediction is the one that was predicted by the majority of the decision trees.

For regression problems the procedure is similar. However, instead of a majority vote to determine the final prediction, usually an average or weighted average of all of the resulting nodes is taken.
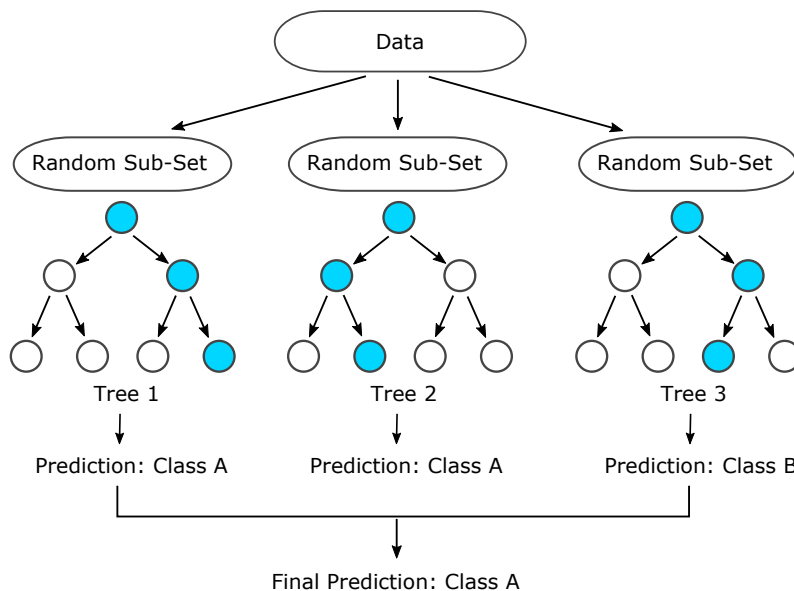

Figure 7.6. Example of a random forest classifier with three decision trees.

Given the high number of channels in the input data, in this study aa RF with 60 individual decision trees will be used.

The importance of the predictor variables (feature importance) is then obtained using a method called permutation, which is also known as Mean Decrease Accuracy (MDA). The idea behind this method is to investigate how the accuracy of the

prediction (indicated by e.g. the coefficient of determination R^2, or an error metric such as MAE, RMSE, etc.) decreases if the investigated feature is not present. This could be done, as shown in Chapter 6, by re-training the classifier leaving out one or more predictor variables. As the model in this chapter is significantly more complex, however, re-training would be extremely computationally expensive. Hence, in order to avoid re-training, one can remove the input variable only from the test data set, and calculate the score without this feature. As simply removing one input variable from the test data would not work, it is replaced by random noise that follows the same distribution as the original input variable. With this, the permutation importance for each input variable can be calculated and compared.

Gradient Boosting Machines (GBM) Similar to RF, the GBM, [214], produces a prediction model based on a combination of weak learners. Whereas, RF uses fully grown trees (high variance, low bias), and lowers the prediction error by reducing the variance, the GBM uses very shallow decision trees with low variance and high bias, and uses 'boosting' mainly to reduce bias (but also variance). The boosting is a sequential process (not parallel tree growing like in RF) and in each iteration a decision tree is added, enhancing the previous model. Hence, the boosting process is a way of increasing the model accuracy by learning from the mistakes of the previous model. This involves, firstly, defining a loss function, e.g. an error metric like the Mean Squared Error (MSE). The gradient boosting aims at minimising this loss function by continuously adding weak learners to the model, while leaving the existing ones unchanged. The predictions are updated until the residuals are close to the minimum and the predictions are similar to the observed values. Unlike RF, it is possible to over-fit a GBM, this risk can for example be lowered by using very shallow decision trees (reducing the nodes per tree), or by defining a (high enough) minimum number of observations that have to be included in a node for it to be considered for splitting. Furthermore, reducing the learning rate (shrinkage) or carrying out a cross validation can be used to avoid this problem.

In this study a GBM with 100 decision trees is applied. Each tree has a maximum depth of 10 levels and a learning rate of 0.1 is applied. Furthermore, a 10-fold cross validation is carried out. Again, permutation is used to obtain the significance of the different variables.

Artificial Neural Networks (ANN) ANNs are among the most famous machine learning techniques and have been used in a vast number of classification and regression problems. A network consist of various interconnected nodes, often called neurons, which are distributed over several layers. The neurons are connected via so called 'synapses', which are weighted paths from one neuron to another. More important connections have bigger weights than unimportant ones. In Figure 7.7 an example of an artificial neural network with one input, two hidden, and one output layer is shown. In this study, a feed-forward ANN set-up is used, which implies that the connections between the nodes only follow one direction: from the input layer towards the output layer. Three hidden layers of 50, 40 and 20 nodes respectively are trained. This set-up showed the best performance in an iterative testing of different configurations starting with one hidden layer and a small number of nodes. The ANN is trained using Levenberg-Marquardt backpropagation [215].
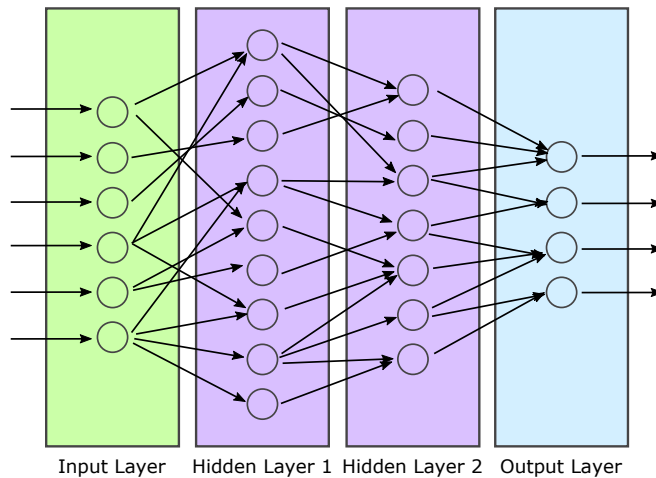

Input Layer   Hidden Layer 1 Hidden Layer 2   Output Layer

Figure 7.7. Example of an artificial neural network with two hidden layers.

This method aims at changing the internal structure of the ANN, i.e. the weights and biases at each node, in order to minimise a given cost function. A commonly used cost function for ANNs is the mean squared error (MSE), which is given by:

$$\Psi(w, B) = \frac{1}{2n} \sum_x \|y(x) - a\|^2 \quad ,$$

(7.7)

where x is the individual input from the training data set, $y(x)$ is the desired output, a are the actual outputs for x, n is the number of inputs in the training set, w

are the weights and B denotes the bias. The backpropagation algorithm assumes that the cost function can be written as an average of the cost functions for all x, which is applicable to the MSE. In a fist step, the weights of the neural network are selected randomly and the outputs are calculated using one of the inputs. With this, the cost function is determined, indicating how close the outputs are to the actual values. Subsequently, the output is propagated back through the network to calculate the error of all hidden and output neurons. Then, the weights and biases of the neurons are modified accordingly. In order to determine how changes in weight and bias of the j-th neuron on layer l affect the cost function, its partial derivatives $\delta\Psi/\delta w_{j,k}^l$ and $\delta\Psi/\delta B_j^l$ need to be computed.

This can be achieved by applying the chain rule starting from the output layer and propagating backwards in direction of the input layer. This shall be explained in the further for the adjustments of the weights of each nodes, however, it is carried out in a very similar manner for the bias. The chain rule for the changes is given by:

$$\frac{\delta\Psi}{\delta w_{j,k}^l} = \frac{\delta\Psi}{\delta o_k^l}\frac{\delta o_k^l}{\delta z_k^l}\frac{\delta z_k^l}{\delta w_{j,k}^l} \quad , \tag{7.8}$$

where o_k^l is the output of neuron k on layer l, $w_{j,k}^l$ is the weight of the connection between neuron j and k, z_k^l is the weighted input value of neuron k, and the error of the output neuron:

$$\delta = \frac{\delta\Psi}{\delta o_k^l} \cdot \frac{o_k^l}{z_k^l} \quad . \tag{7.9}$$

For example on the output layer, the final value of the derivative $\delta\Psi/\delta o_k^l$ is obtained by:

$$\frac{\delta\Psi}{\delta o_k^l} = \frac{\delta\Psi}{\delta a} = a - y(x) \quad . \tag{7.10}$$

As in this case the output of the neuron is equal to the output value of the whole ANN, the derivative is the expected output value minus the obtained one in this neuron. The derivative of the output of neuron k regarding its input can be obtained by:

$$\frac{\delta o_k^l}{\delta z_k^l} = \frac{\delta \gamma(z_k^l)}{\delta z_k^l} = \gamma(z_k^l)(1 - \gamma(z_k^l)) \quad , \tag{7.11}$$

where $\gamma(z_k^l)$ is the sigmoid activation function, denoted as:

$$\gamma(z_k^l) = \frac{1}{1 + e^{-z_k^l}} \quad . \tag{7.12}$$

Note, activation functions other than the sigmoid function can be used, as explained e.g. in [215]. The derivative of the weighted input with respect to the weight of the incoming connection is written as:

$$\frac{\delta z_k^l}{\delta w_{j,k}^l} = \frac{\delta w_{j,k}^l o_k}{\delta w_{j,k}^l} = o_k^l \quad . \tag{7.13}$$

With this, the error δ on the neuron of the output layer can be defined as:

$$\delta = (a - y(x)) \cdot a(1 - a) \quad . \tag{7.14}$$

In order to calculate the error on a neuron of the next hidden layer, it is assumed that the cost function depends on all inputs from the neurons that receive input from this neuron. Hence, the derivative is obtained by the sum of the derivatives of all neurons in the layer closer to the output layer that are connected to this neuron. This is given by:

$$\frac{\Psi}{\delta o_k^l} = \sum_{e \in E} \left(\frac{\delta \Psi}{z_e^l} \frac{\delta z_e^l}{\delta o_k^l} \right) = \sum_{e \in E} \left(\frac{\delta \Psi}{\delta o_e^l} \frac{\delta o_e^l}{\delta z_e^l} w_{k,e} \right) \quad . \tag{7.15}$$

Here, the neurons of the layer that is located closer to the output and that serve as input to the neuron of interest, are denoted with e. This results in the error for this neuron:

$$\delta = \sum_{e \in E} (\delta_e w_{k,e}) a(1 - a) \quad . \tag{7.16}$$

This, however, only works if the errors of all neurons of the layer closer to the output have been calculated beforehand. Finally, the weights of the connections have to be adjusted, using gradient descent optimisation, with the aim of minimising the cost function. The adjustment of the weights is given by:

$$\Delta w_{j,k}^l = -\kappa \frac{\delta \Psi}{\delta w_{j,k}^l} = -\kappa y^j \delta^k \quad , \tag{7.17}$$

where κ is the learning rate, that has to be pre-defined. With higher learning rates, the algorithm converges faster, is however, likely to be less accurate. This procedure is carried out for all neurons of all layers of the ANN starting from the output layer and progressing towards the input layer. For further information, a comprehensive explanation of backpropagation and gradient descent optimisation can be found for example in [216, 217].

The variable importance is determined using functional analysis of the final weight matrix, as introduced in [218]. This method derives measures based on the architecture of the trained ANN, not exclusively on the data, and has shown to outperform traditional magnitude based input significance for ANN.

7.2.6 Prediction of CMS alarms

Three distinct approaches for predicting the CMS alarms are used. These contain predicting the count of CMS alarms, the time dependent probability of having a CMS alarm, and predicting the events using an alarm shifting method. Each approach will be explained in detail in the following.

Count of alarms. For predicting the number of CMS alarms, only SCADA data are used. The SCADA channels that were containing constant or cumulative values were eliminated for this task, as they would contain unnecessary information.

For this task, the previously mentioned algorithms are used in a regression set-up. This means a discrete quantity output for the count of alarms is predicted. Thus, the GLM is set with a Poisson distribution and logarithmic link function. The other algorithms are used as explained above.

Two distinct concepts for training and testing are applied:

- Random sampling: The data from all turbines are split with a ratio of 80% training and 20% testing.
- Blind testing: The training is carried out with data from all turbines except one. Then blind testing is then done on the data of the remaining turbine. This leads to an approx. share of 92% of the data for training and 8% for testing.

As in the data set the number of time-steps containing entries for alarm events is much lower than the time-steps where no alarms were observed, the class distribution needed to be adjusted using a method called 'under-sampling'. This technique re-balances the representation of different targets, [219]. Hence, the resulting training dataset is under-sampled, which limits the number of time-steps without alarm to 80% of the data.

Time dependent probability of alarms. For this task, the aim is to calculate the time-dependent probability of getting a CMS alarm using a classification approach. Again, only SCADA data are used for the predictions. For the training, the data are under-sampled to a ratio of 80% and 20% for the two respective classes {0,1}, which stand for {'no alarm','alarm'}. For the prediction, the *blind testing* approach is applied, as any random sampling would hinder a time-dependent evaluation.

The algorithms introduced earlier are used in a set-up for classification problems. Thus, the GLM is applied with a binomial error distribution and a *logit* link-function. Also, when modelling binary response variables, the other learning algorithms can be used as probabilistic classifiers. Then, the output from the predictions is a posterior class probability at each point in time, [220]. To compare the performance of the techniques, both, turbines with and without failures are analysed.

Alarm time shifting. Here, again a binary classification approach is used. The predictions are carried out by shifting the dataset to an earlier time step and thus training the model with shifted data. A lead time equal to zero, corresponds to modelling the alarm, while higher lead times correspond to prediction of the alarm in advance. Although this is a static way to represent a dynamic process, it can provide useful information about the performance of the classifier before the actual occurrence of an event. Various combinations of the different input sources are investigated: 1) SCADA, 2) SCADA and RMS and 3) all data. Here 'all data' stands for SCADA, RMS and all MB spectra from one active power interval. With this the contribution of each data source to the prediction accuracy is evaluated. For this task, the classifiers RF and ANN are employed. Furthermore, data from all turbines are used with random sampling and are again under-sampled, as described before. A 10-fold cross validation is carried out on the data leading to 10 distinct predictions, which are then gathered to provide an average performance.

7.3 Results: Failure Prediction based on merged CMS and SCADA Data

This section presents the results for the application of the processed and merged CMS and SCADA data for the three defined objectives: (1) understanding relationships between the data, (2) automated failure detection (DAVE) and (3) predicting the CMS alarms. These three approaches do not interact and their results should be considered separately.

7.3.1 Understanding Relationships between Data

In Figures 7.8 - 7.9 the results of the HC analysis of all data per turbine are shown using simplified dendrograms and statistics of the most separated clusters. Generally speaking, a cluster analysis can show which signals are similar, i.e. joined in a cluster, and which are more different, i.e. in separated clusters. In agglomerative HC, the process of building the clusters is more important than the final result as all signals are eventually joined to one cluster. This process can be visualised using so called dendograms, which show how the clusters are formed by connecting two 'leaves', i.e. sub-clusters, to form the next cluster. The x-axis shows the sub-clusters or in the lowest level all signals. Due to the high number of signals used in this study (3375), the lowest level can not be shown, and the dendrograms were cut at 20 leaves. The height in the y-axis represents the difference of the two joined sub-clusters with respect to the Euclidean distance. Hence, a dendrogram with vertically close clusters indicates very similar signals.

It can be seen in Figure 7.8a that in fault-free operation, there was no clear cluster separation obtained. The two clusters that are joined at the top of the dendrogram consisted here of one big cluster and one smaller cluster with only a few signals (the most separated cluster). The second most separated cluster, which is not a sub-cluster of the most separated cluster, showed a similarly small number of signals. However, in the case of the MB failures clear separation of more signals is shown in Figures 7.8b and 7.9. The two most separated clusters were formed by many MB signals for T03 and T08. This proves that there are clearly deviating features in a number of signals before the failure event. Noticeably, there were mainly GNDE and not MB contributors in the case of the MB failure in T01.
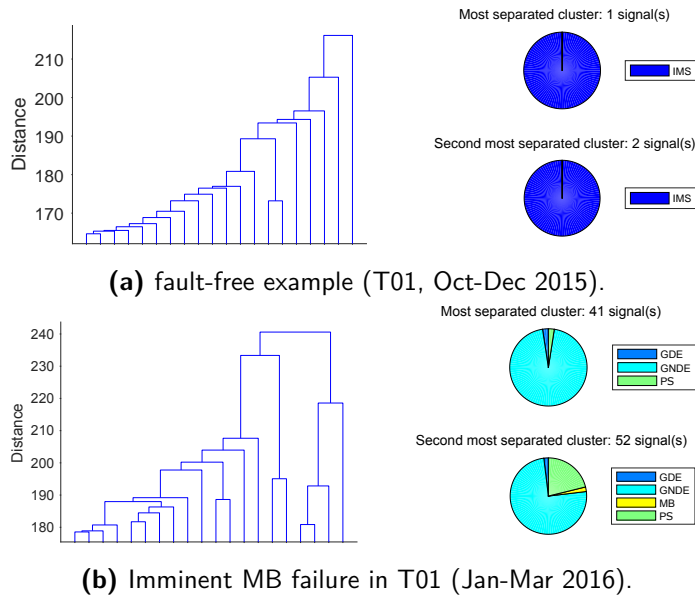
(a) fault-free example (T01, Oct-Dec 2015).

(b) Imminent MB failure in T01 (Jan-Mar 2016).

Figure 7.8. Simplified dendrogram and contribution to most separated clusters in T01.

(a) Imminent MB failure in T03 (Apr-Jun 2015).

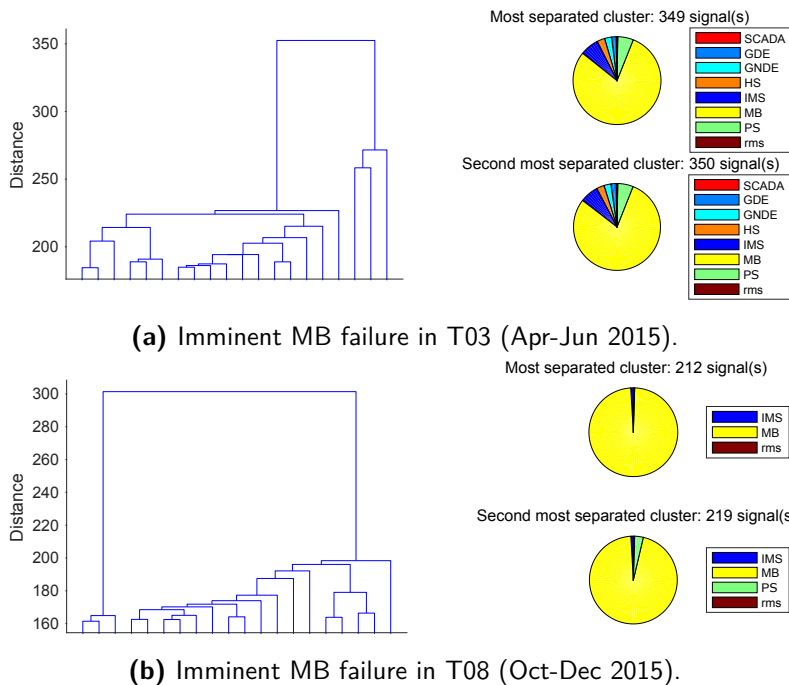**(b)** Imminent MB failure in T08 (Oct-Dec 2015).

Figure 7.9. Simplified dendrogram and contribution to most separated clusters in case of MB failure for turbines T03 and T08.

In a second step, the relationships between the CMS records for different active power intervals were evaluated using DTW distances. As stated before, in order to ensure nearly stationary operating conditions for further analysis, the condition monitoring systems separate the vibration records according to the respective active power interval, in which they were recorded. The aim of this comparison was to see, which of the seven active power intervals is the most relevant to indicate faulty conditions, such that this can be used in further application for failure prediction. The DTW distance was calculated by comparing two spectra, each measured in a different active power interval. Here, the calculated distance is normalised by the number of samples in the signal for better readability.

The pairwise comparison of the spectra obtained in different active power intervals showed that for fault-free operation, the normalised distances were usually greater than 0.2. However, in the case of MB failures, the spectra obtained for all active power intervals became more similar, showing values as low as 0.05. The normalised distances between the spectra recorded before the failure of turbine T01, T03 and T08 are displayed in Figures 7.10, 7.11a and 7.11b. In these graphs, the pairwise distances of all active power intervals are shown for various combinations of MB records. Each set of seven nodes represents the seven active power intervals. The heading above each set of nodes identifies the two compared record types according to the direction of the arrows connecting the nodes (always starting at the lower number). The colour of the arrows shows the similarity of the two connected spectra, as indicated by the DTW distance. With this, higher distances imply higher dissimilarities.

As shown in Figure 7.10 for T01, some spectra were only similar for certain power intervals. Very low distances and accordingly similar features were obtained for all intervals in the FFT1000 spectra. Contrarily, interval 1 appeared to have unique features for Env200 spectra, as they are not connected to any other node. If the different types of records are compared, it can be seen that the FFT1000 and Env200 are more similar, but the two FFT records FFT1000 and FFT35 are less similar.

The resulting DTW distances for the MB failures in T03 and T08 shown in Figures 7.11a and 7.11b are comparable to T01, except for interval 6 with unique features (instead of interval 1). However, all FFT1000 and Env200 distances were slightly larger in T08 – a trend that was amplified in T03 with even higher distances.

To sum it all up, the DTW analysis confirmed that the CMS records from the different active power intervals showed mostly similar features, i.e. using data from only one active power interval is a reasonable compromise to handle large number of records. Thus, in the further, the study is based on data from the fourth active power interval corresponding to 58-69% rated capacity.
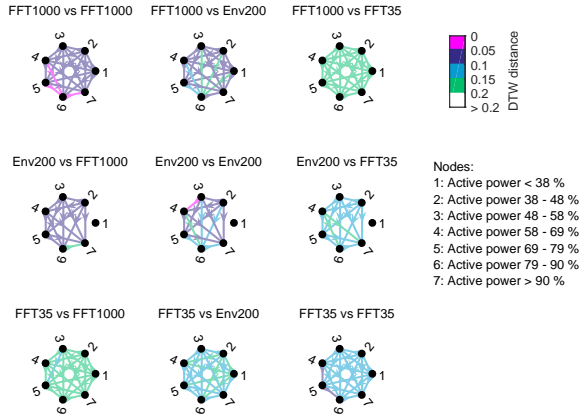

Figure 7.10. DTW distances for various MB records with different active power intervals in the case of a MB failure for T01 (Jan-Mar 2016).
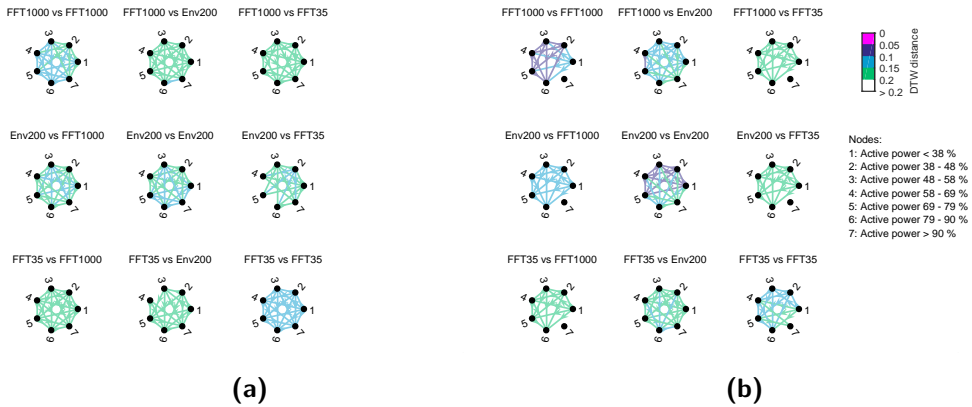

Figure 7.11. DTW distances for various MB records with different active power intervals in the case of a MB failure: (a) T03 (Apr-Jun 2015), (b) T08 (Oct-Dec 2015).

7.3.2 Distance-based automated vibration evaluation

In this section the performance of the automated failure detection tool DAVE is demonstrated for CMS records obtained for the main bearing.

The initial setup of DAVE was achieved by deriving the thresholds from the first two observed (healthy) years of the entire dataset. A Weibull distribution showed the best fit to the calculated distances and the threshold for an anomaly was set to the 99.9 percentile of the distribution. In Figure 7.12 the alarms, triggered by CMS, SCADA and DAVE (for FFT1000) as well as the downtime caused by the failure during an observation period of four years, are displayed. The turbines T09 to T13 are not shown in this graph as none of them had CMS nor SCADA alarms and DAVE did not indicate any alarms for these WTs.

DAVE showed very early alarms for the three MB failures in T01, T03 and T08. Furthermore, no false positives or false negatives were recorded during the four year period.

In Table 7.1 the number of days DAVE was able to anticipate the component problem in comparison to the CMS and SCADA alarms are displayed. It can be seen that using FFT1000, DAVE was able to detect the problems up to 72 days before the CMS did. The other MB records, Env200 and FFT35 gave less reliable warnings.
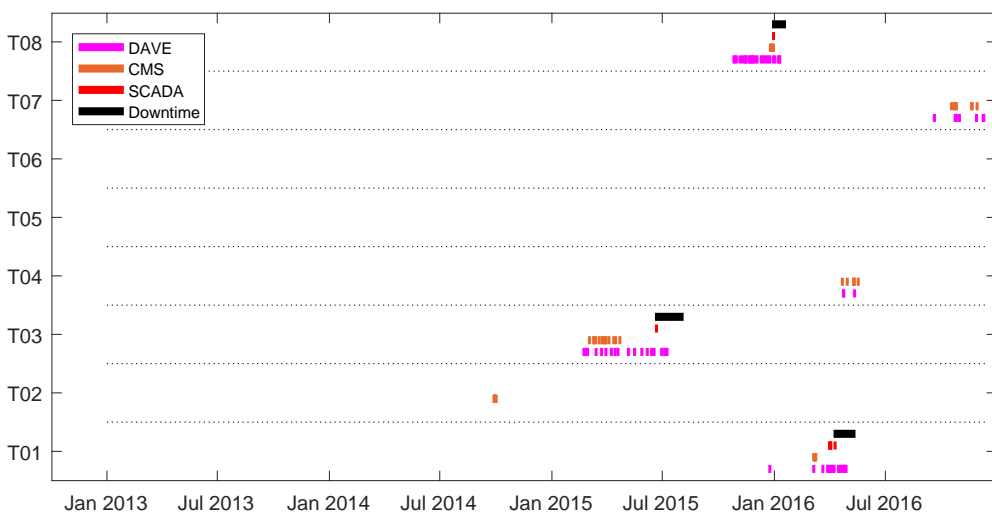

Figure 7.12. Results for the automated failure detection (DAVE). Comparing CMS, SCADA and DAVE (FFT1000) alarms.

Table 7.1. Number of days the different systems triggered alarms before failure.

WT	DAVE			CMS	SCADA
	FFT1000	FFT35	Env200		
T01	106.5	11.8	-	34.5	9.3
T03	119.9	-	105.1	110.1	0
T08	65.1	-	93.8	5.5	0

7.3.3 Predicting the Count of CMS alarms

Here, the results for modelling the number of CMS alarms using only SCADA data are presented. The performance of the four different algorithms is evaluated using the coefficient of determination R^2, the mean absolute error (MAE) and the root mean squared error (RMSE). Furthermore, the most important model covariates for each set-up are determined. In Table 7.2 the evaluation metrics for training and testing of each technique using *random sampling* are shown.

Table 7.2. Evaluation metrics for modelling CMS alarm counts using only SCADA data with random sampling.

	Metric	RF	GBM	GLM	ANN
Train	R^2	0.893	0.942	0.142	0.610
	MAE	0.151	0.192	0.982	0.388
	RMSE	1.261	0.929	3.573	2.438
	Metric	RF	GBM	GLM	ANN
Test	MAE	0.149	0.224	0.982	0.379
	RMSE	1.151	1.182	3.592	2.468

Figure 7.13 displays the recorded CMS alarms and the predictions obtained from each model. RF and GBM performed best, intermediate results were obtained for ANN, while GLM showed a rather poor performance.

In Figure 7.14 the importance of the different variables in each of the models
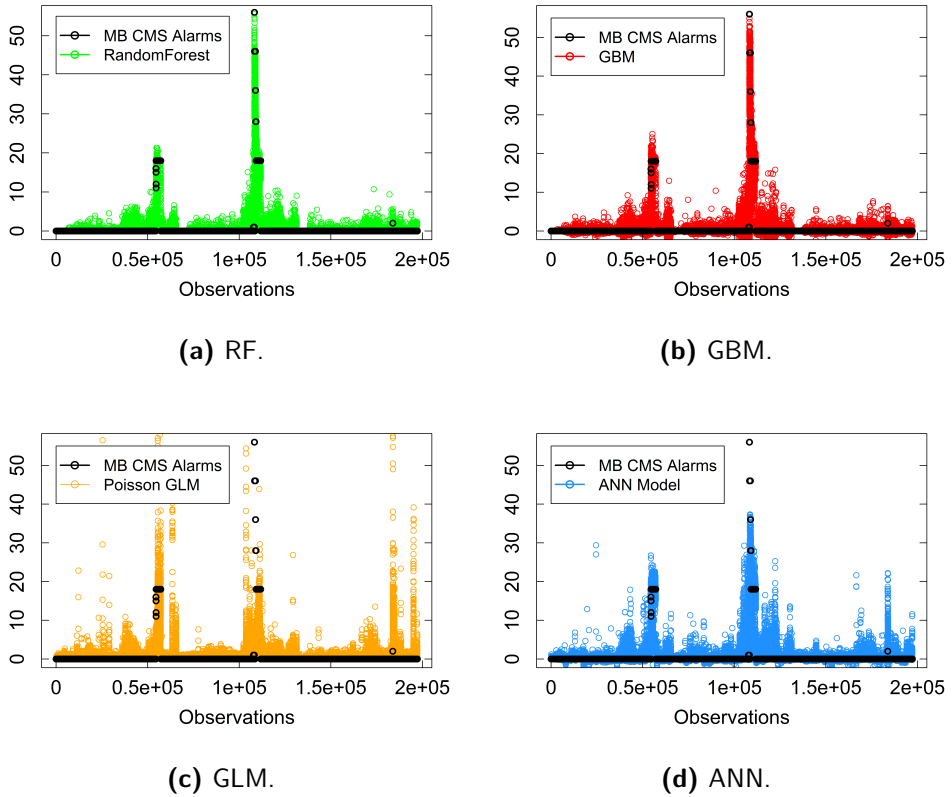
159

(a) RF.

(b) GBM.

(c) GLM.

(d) ANN.

Figure 7.13. Results for the predictions of CMS alarm counts with random sampling.

are shown. These indicate that in all cases the average main bearing temperature was the most important model covariate.

The results for *blind testing* are displayed in Figure 7.15 and Table 7.3. It can be seen that during training and testing RF and GBM performed better than ANN and GLM. Additionally, it is shown that the predictions using the testing dataset were characterised by substantially higher errors than for the predictions with random sampling. This is most likely due to the splitting process used in blind testing. As one turbine is left during in the training phase, the algorithm can not capture the variations in the turbine-dependent operational and environmental conditions for this turbine. Hence, the predictions are less accurate.
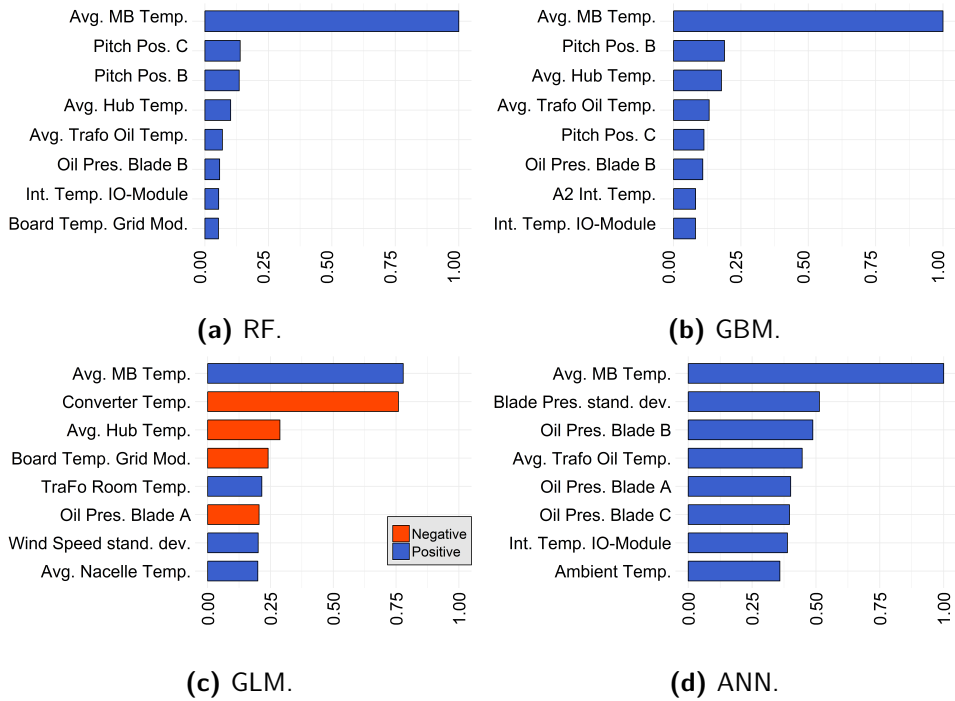
(a) RF.

(b) GBM.

(c) GLM.

(d) ANN.

Figure 7.14. Variable importance for the algorithms used for random sampling.

(a) RF.

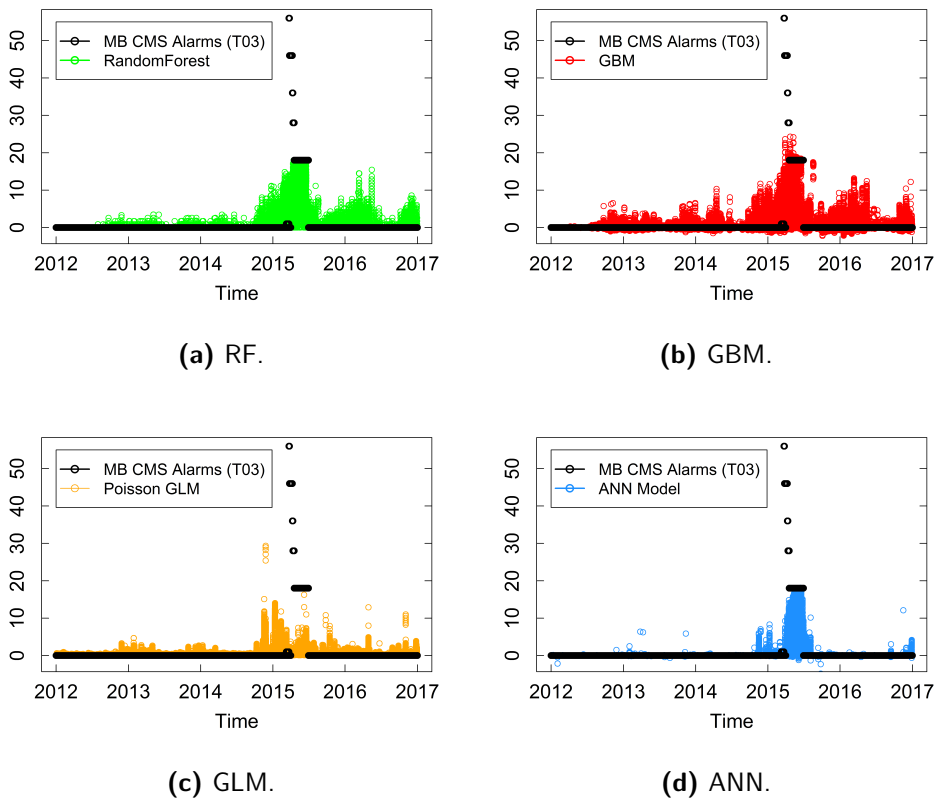**(b)** GBM.

(c) GLM.

(d) ANN.

Figure 7.15. Results for modelling of CMS alarm counts with blind testing.

Table 7.3. Evaluation metrics for modelling CMS alarm counts using only SCADA data with blind testing.

	Metric	RF	GBM	GLM	ANN
Train	R^2	0.979	0.989	0.218	0.549
	MAE	0.028	0.053	0.504	0.195
	RMSE	0.364	0.256	2.211	1.672
	Metric	RF	GBM	GLM	ANN
Test	MAE	0.974	0.950	1.350	1.079
	RMSE	4.698	4.689	5.640	5.498

7.3.4 Predicting the Time Dependent Probability of Alarms

This section shows the results of predicting how the probability of having a CMS alarm is evolving over time. The predictions are carried out using only SCADA data. This is evaluated graphically by plotting the alarm events as well as the predicted probabilities over time obtained with all four algorithms. The aim is to see how the probabilities behave before the failure event and which algorithm was able to indicate an upcoming alarm more reliably.

In Figures 7.16a to 7.16c the results for turbines T01, T03 and T08 are shown, for which CMS MB alarms were recorded during the observation period.

Furthermore, in Figures 7.16d to 7.16f turbines T06, T10 and T11, which did not experience any MB CMS alarms, are displayed. For an easier interpretation, the probabilities over time are smoothed using a moving average filter.

It can be seen that GBM performed best. It showed a rising probability towards the time of occurrence of the alarms for turbines T01, T03 and T08. Contrarily, for turbines T06, T10 and T11, which did not experience any CMS alarms, the model resulted in near-zero probability. The algorithms GLM and RF indicated a quite high probability of having an alarm for T06. Generally, GLM performed poorly for all cases. A certain seasonality was observed for the ANN predictions of T01 and T06 as well as the GLM predictions of T01, T10 and T11. Furthermore, in T08 GBM and RF showed a peak in the obtained probability approximately one year before the failure. This could be caused either by seasonality, or a separate problem with the MB or any other component in T08. Further investigation is required to fully understand the observed trends.
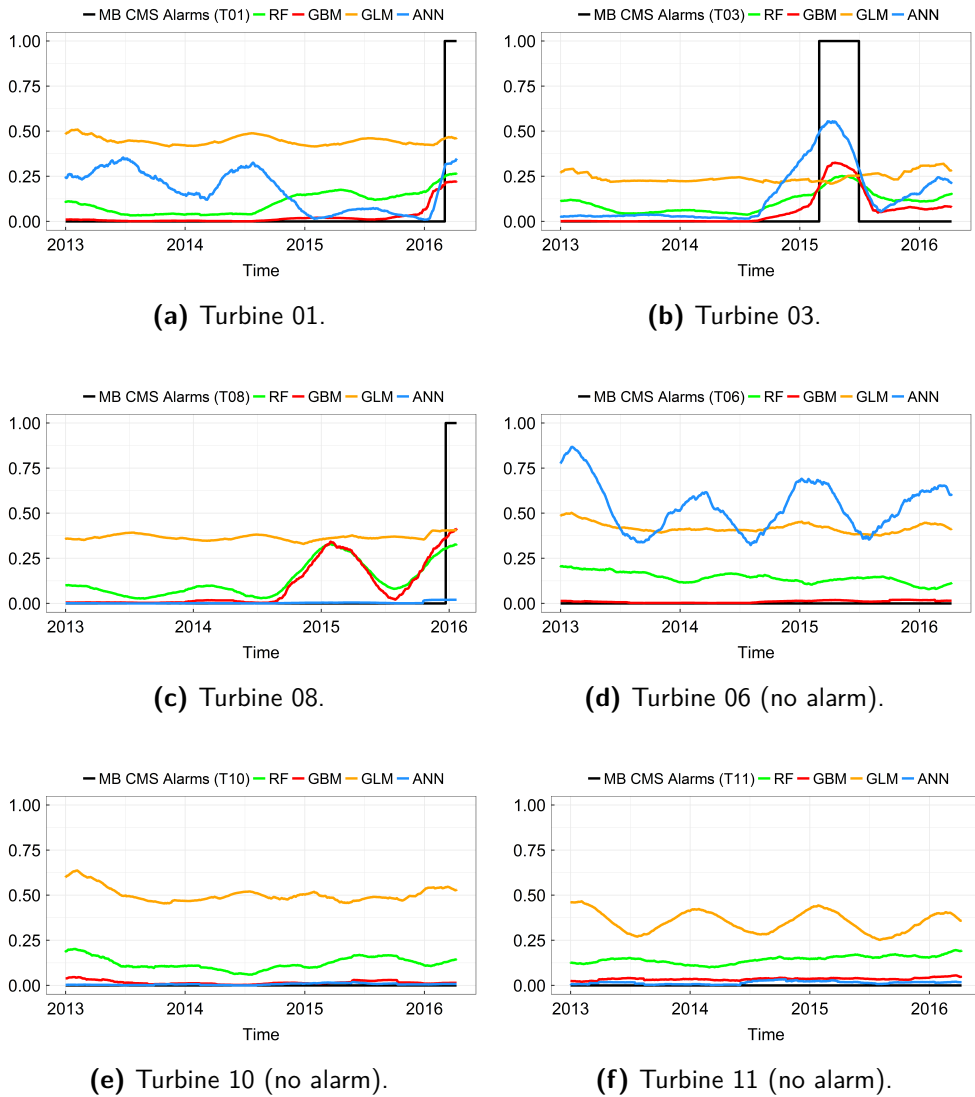
(a) Turbine 01.

(b) Turbine 03.

(c) Turbine 08.

(d) Turbine 06 (no alarm).

(e) Turbine 10 (no alarm).

(f) Turbine 11 (no alarm).

Figure 7.16. Probability of having a CMS alarm for blind testing.

7.3.5 Alarm Time Shifting

The performance of the predictions using alarm time shifting is evaluated by receiver operator characteristic (ROC) curves at different fixed lead times. These curves show the hit rate (probability of detection (POD)) versus the false alarm rate (FAR) as a function of the threshold for an alarm in a probabilistic setting. In Figures 7.17a and 7.17b the results for the predictions made with ANN and RF, with a lead

time of 0 hours, are shown. The figures consist of ten different ROC curves for each of the predictions made during cross validation. The red line is the average ROC curve.
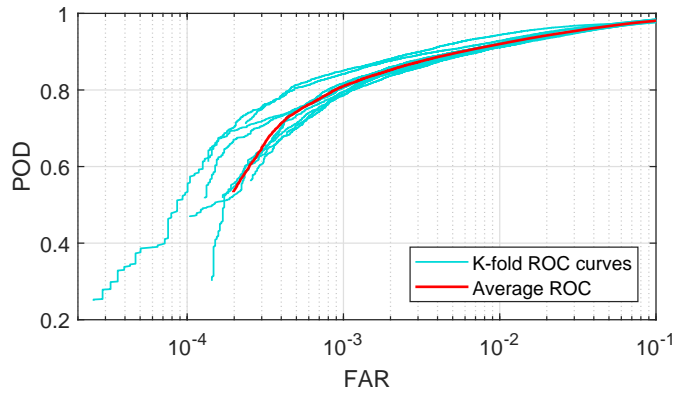
(a) ANN.
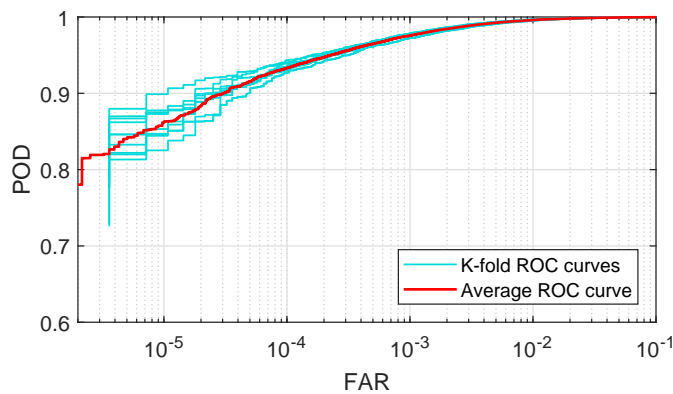
(b) RF.

Figure 7.17. Example of variation in the results of the 10-fold cross-validation.

Figures 7.18a and 7.18b display the average ROC curves for different combinations of input data and various lead times. It has to be noted that in the figures, different x-axis scales are used in order to emphasise the difference between different combinations of data, rather than comparing the performance of the two classifiers. Furthermore, in the legend of the figures, 'all data' stands for the input set combining SCADA, RMS, FFT and Envelop records.

The RF algorithm performed significantly better than the ANN in all cases. Nonetheless, the POD of both algorithms in predicting the CMS alarms using only
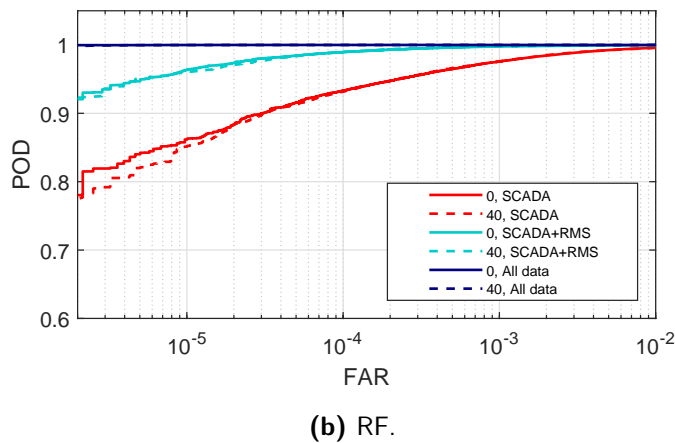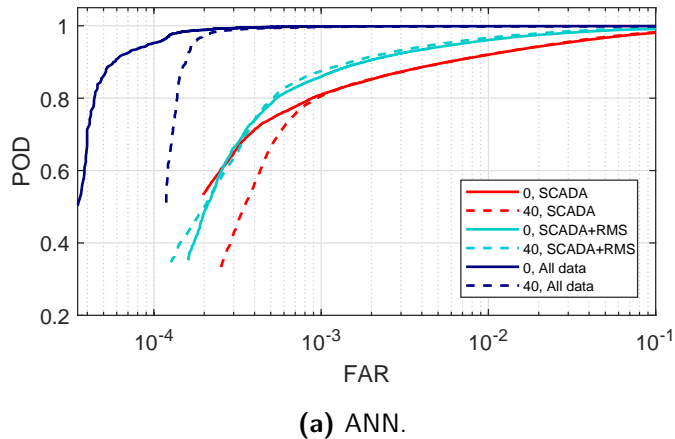
(a) ANN.

(b) RF.

Figure 7.18. Average ROC curves for ANN and RF at lead times 0 and 40 hours.

SCADA data is quite high. But, using more data improved the predictions. Hence, the combination of all CMS and SCADA data performed best, followed by the combined SCADA and RMS data. This shows that the added value of CMS data is large.

Additionally, it is shown that with a lead time of 0 hours the predictions for all combinations of input data are better. This was also an expected result, since closer to an actual event the separability between the classes increases. It is worth noting that the order of the POD and FAR displayed is affected by the used pre-processing approach for the data, in particular the fact that the CMS records are kept constant if there was no measurement.

166

7.4 Concluding Remarks for the Data-Driven Fault Prediction

In this chapter a method for merging CMS, SCADA and alarm data was proposed. Subsequently, a thorough analysis was carried out on the possible applications with respect to enhancing predictive O&M practice.

Analysing the relationship between the different data sources gave insight on how the signals obtained during faulty conditions differ from the ones measured during 'healthy' operation. The HC analysis showed that in the case of normal conditions, there is no significant cluster separation in the data obtained for the whole turbine. In the case of imminent failures, however, a strong separation of clusters was seen. Hence, failures are clearly reflected in a number of signals. Evaluating the DTW distances of different CMS records shows that the information contained in the MB records does not differ significantly for various active power intervals.

An automated failure detection framework (DAVE) was developed and tested on three registered MB failures within the analysed wind farm. It was demonstrated that this tool was able to issue alarms up to 120 days prior to the failure. Furthermore, the low number of false alarms proves that DAVE gives not only earlier alarms than the CMS, but is also very reliable.

In order to predict the CMS alarms, three different approaches were chosen. Each one of these approaches was used to accomplish distinct objectives.

Firstly, the number of CMS alarms were modelled with RF, GBM, ANN and GLM by using only SCADA data as model input. It was shown that it is possible to anticipate the CMS alarms by only using the data taken from SCADA systems. For this, the algorithms RF, GBM and ANN performed best. Additionally, by analysing the importance of the input covariates on the model outcome, it was found that the main bearing temperature was the most important model covariate.

Secondly, the probability of having a CMS alarm over time was calculated also only based on SCADA data. This could for example be used by operators as an early warning system. It was stated that the GBM algorithm reliably indicated an increase in the probability of having an alarm several months before the actual alarm. Furthermore, for turbines that did not experience any CMS alarm, the GBM indicated a probability close to zero over the whole observation period.

Finally, the CMS alarms were predicted using a method to shift the alarm time

to an earlier time-step. For this, the performances of RF and ANN with different combinations of input data were examined using ROC curves. The alarms were predicted with different lead times (0 and 40 hours) before the actual alarm event. As expected, the use of combined CMS and SCADA data showed the best results for predicting CMS alarms. However, only using SCADA data still led to a remarkably low number of false alarms.

This chapter intends to exemplify the use of merged CMS and SCADA data for several applications in the field of failure prediction and detection. Further studies, can build upon and extend this work by for example including:

- A more detailed analysis of the relationships between the different data sources by considering the detailed composition of clusters and DTW similarities inside these clusters.
- The DAVE tool should be evaluated for other failure modes and components. Additionally, it could be equipped with a user interface in order to obtain a tool framework for failure detection that can be used by operators.
- The methodology of predicting the probabilities of having an alarm could be integrated into an online monitoring framework.
- The prediction of CMS alarms could be extended using other learning algorithms, maximising the possible lead times for several components.
- The effectiveness of the classification based predictions should be evaluated from a risk perspective.
- The different approaches presented in this work could be merged, for example by predicting alarms generated by DAVE.

Generally speaking, the synergies between CMS and SCADA data should be explored more precisely. In order to avoid high frequency time series storage, a data processing directly in the turbine could be established.

It can be concluded that with the proposed methods for merging CMS and SCADA data, it is possible to develop several different applications for predictive O&M. The ones presented in this paper have shown to perform very well and could contribute to lowering the costs related to maintenance actions.

8

General Conclusions of the Thesis

Operation and maintenance is one of the main cost drivers of modern wind farms. Current O&M practice is still primarily based on corrective or preventive actions, however, applying predictive techniques could notably decrease the maintenance related costs and with this enhance the wind farms' overall revenue.

Novel approaches, such as advanced reliability models and failure prediction algorithms would enable operators to anticipate WT failures and to adapt their maintenance strategy accordingly. It has been found that, to date, there is still a significant lack of research regarding these predictive techniques. At present, WT reliability models are mainly based on the turbine age, similar to models that were originally established for machinery operating in fairly steady surroundings. This is not reflecting reality sufficiently well, as the turbines are exposed to a highly varying environment. Furthermore, failure prediction is often carried out using either SCADA or CMS data. However, in modern wind turbines both sources are available and could be used simultaneously. By merging both sources, it has to be discovered, how these data depend on each other and if it is possible to omit one or the other system.

The main objective of this thesis is to enhance current WT O&M practice by, firstly, developing advanced data-driven WT reliability and failure models including the environmental and operational conditions measured within the respective wind parks. And, secondly, establishing novel failure detection algorithms using environmental, operational and CMS data. These two objectives were achieved by splitting

them into four independent tasks, which included: (1) A wind turbine failure and downtime analysis; (2) understanding the meteorological conditions before failure events of main WT components; (3) developing a reliability model that incorporates these conditions; and (4) failure prediction based on environmental conditions or vibrational and operational data.

An extensive failure analysis was carried out taking into account historical failure data of around 4600 wind turbines of different turbine technologies including various drive train set-ups, ages, power regulation, etc. For this, firstly, a taxonomy was developed for classifying the WT system into 7 sub-systems, 45 assemblies and 199 sub-assemblies. Developing this taxonomy was necessary, as there was hardly any component break-down publicly available. Furthermore, it extends the existing ones, which are mainly based on outdated technologies, by incorporating information on modern WTs of different technologies. This taxonomy is applied to the data base and the yearly failure rates and downtime for each WT sub-system and assembly are determined. The results presented for this task enhance previous research on WT failures significantly, as they give a detailed insight on the failure behaviour of different components and WT technologies, including both, older and newer ones. Furthermore, a failure mode analysis is carried out for each assembly. The results were presented for geared and direct drive WTs with rated capacities below, over or equal to 1 MW, as well as pitch and stall regulated turbines. It was shown that geared turbines have higher failure rates and downtime than direct drive turbines. This is consistent with previous research. Furthermore, the sub-system and assembly related failure rates and downtimes obtained for each of the analysed technologies differed in some cases remarkably. The failure mode analysis showed that there are often prevalent failure modes for certain components. These include e.g. gear bearing problems in the gearbox, surface cracks in the blades; gear rim faults in yaw systems, and bearing issues leading to generator failures. All of the presented results are very important for predictive O&M of wind turbines, as they serve as input to reliability and maintenance models. Moreover, operators and manufacturers can obtain information on which components are the most affected ones in each of the WT technologies, as well as on how they fail.

Subsequently, the meteorological conditions before the failures of five main WT components were analysed. These main components were identified in the previous failure analysis as the most critical ones in terms of failure rates and

downtime. At first, a data-driven learning framework based on apriori rule mining was developed to analyse the conditions during a relatively short period of time before failure. These included the general conditions over the whole observation period, the combinations of environmental variables at the exact time of failure as well as variations in the weather parameters. The results show that the failures occurred mainly during winter time, with high relative humidities, low temperatures and fairly high wind speeds (especially for the pitch system). Furthermore, a loss in production efficiency was recorded before all component failures.

As previous research had shown that high wind speeds affect the components' failure behaviour, and it was supposed that variations and recurrent patterns in wind speed time series before failures could affect the components' life-time negatively, these were analysed in this thesis using anomaly detection and motif discovery techniques. To the author's knowledge these variations and patters have not been subject to previous studies. The findings show that during the month of failure, the average number of anomalies detected per time series before component failure, are much higher than during a 'regular' month. Hence, it can be concluded that the variations in wind speed (not only the mean wind speeds) highly affect the components, especially the blades, converter and pitch system. The motif discovery also was able to detect recurrent patterns before converter failure, however, this technique will have to be further enhanced. The findings regarding the meteorological conditions leading to WT component failures further contribute to the field and extend previous research by clearly indicating which factors influence the failure behaviour of the components and by providing a tool that enables to analyse big data bases in future studies.

After having shown that certain environmental conditions affect the components' failure behaviour, an advanced failure model was established taking into account these conditions. Here, weather data obtained during a longer period before failure (two months) are considered. The model is based on two separate processes, one is generating the failure events as well as occasional zeros with a negative binomial distribution, the other process generates the structural zeros and follows a binomial distribution. Possible unobserved effects (heterogeneity) are taken into account using a Gamma distribution. To the author's knowledge, no previous approaches have been made to model the failure behaviour of wind turbines taking into account the complex combinations of weather conditions. This model has sown to perform

very well for modelling failure events of the whole wind turbine system as well as a single component (i.e. the gearbox). Using sophisticated regularisation techniques, several common problems of statistical models are avoided. The herein developed model contributes to current research providing a novel and more realistic reliability model, which is recommended for further use in the field.

For the purpose of failure detection on wind farm level, a naive Bayesian classifier was trained also based on the site specific environmental conditions. Instead of predicting the number of failures as in the previous chapter, the model detects the probability of having one or more failures within a wind farm given the surrounding weather conditions. The approach reliably predicted most of the component failures given in a test data set. Furthermore, the conditional probabilities of each meteorological input variable were extracted and compared to the results of Chapters 3 to 5. This technique provides a valuable tool to wind farm operators, and enables them to detect problems in the wind farm even without having to analyse extensive CMS and SCADA data.

Nonetheless, fault detection based on CMS and SCADA data is becoming a well established area and failures can be predicted significant time ahead of the actual event with quite low uncertainties. However, current research focuses either on SCADA data or CMS data. Combining both sources with the aim of enhancing failure detection has not been subject to previous research. Hence, the last chapter of this thesis explored data-driven fault detection in wind turbine components using a combination of SCADA and CMS data. For this, the relationship between the different data sources were investigated. Then, a technique for merging CMS vibrations and SCADA data of different time resolutions was proposed. As CMS alarms usually occur way ahead of the actual failure event, predicting these alarms implies an early detection of the failure. The alarms were predicted with several machine learning algorithms. It was found that Random Forests and Gradient Boosting Machines provide very good results in predicting the count of CMS alarms as well as the changing probability of having a CMS alarm over time. It was further shown that it is possible to predict the CMS alarms only using SCADA data. This leads to the assumption that, under certain circumstances, there could not be a conspicuous reason for using time consuming and expensive CMS data analysis. Nevertheless, the prediction accuracy was significantly enhanced when including the CMS vibration data into the algorithms.

Finally, the DAVE framework was presented, which is an on-line wind farm level failure detection tool based only on CMS vibrations. It uses Dynamic Time Warping to calculate the distances between the vibrations obtained at the same component of different wind turbines. This tool has shown to be able to detect failures up to 119 days ahead of their actual occurrence and outperformed the alarm system embedded in the condition monitoring system.

8.1 Future Work

Future work could focus on including SCADA alarms into the herein presented models and tools. It is supposed that certain alarms, even if they are not triggered by the respective component, can indicate faulty components. For example, a combination of cooling system and high interior temperature alarms could indicate a gearbox problem even before a gearbox alarm is issued. This information, together with environmental, operational and vibration measurements, could be capable of enhancing the failure prediction for certain WT components.

The environmental conditions before failures could be analysed in more detail focusing on their cumulative and long term effects on the components. This could be carried out by adapting the herein presented data-driven learning framework respectively.

Moreover, commercial on-line monitoring tools could be developed based on the Bayesian classifier and the DAVE framework presented in this thesis. For this, it would be necessary to test these approaches on bigger scenarios in order to ensure their robustness.

The motif discovery has shown great potential to detect recurrent patterns in wind speed time series, this should be analysed further and appropriate techniques to evaluate the results of the algorithm must be taken into account.

As the approaches presented in this thesis all follow slightly different objectives, it would be interesting to combine the anomaly detection and motif discovery algorithms, the DAVE framework, the naive Bayesian classifier approach and the reliability model, in one predictive O&M tool, which combines the different objectives and generates several forms of early warnings.

9

Conclusiones Generales de la Tesis

Las tareas de Operación y Mantenimiento (O&M) representan una de las mayores partidas dentro de los costes de explotación de los parques eólicos modernos. La práctica actual de O&M está todavía basada en acciones correctivas o preventivas que se basan en reparaciones, generalmente caras, de componentes averiados o bien en acciones que se repiten de forma periodica sin tener en cuenta el estado real del activo. Sin embargo, la aplicación de técnicas predictivas podría disminuir notablemente los costes relacionados con el mantenimiento y con ello mejorar los beneficios globales de los parques eólicos.

Nuevos enfoques tales como modelos avanzados de fiabilidad y algoritmos de predicción de fallos, permitirían a los operadores anticipar los fallos de los aerogeneradores y así adaptar su estrategia de mantenimiento. Hasta la fecha, todavía existe una carencia significativa de investigación en estas técnicas predictivas. En la actualidad, los modelos de fiabilidad de turbinas eólicas se basan principalmente en la edad de la turbina de forma similar a los modelos que se establecieron originalmente para maquinaria que opera en entornos 'amigables'. Esto no refleja la realidad lo suficientemente bien, ya que los aerogeneradores están expuestos a un entorno altamente variable y agresivo. Además, la predicción de fallos se lleva a cabo habitualmente utilizando datos SCADA o CMS. Sin embargo, en las turbinas eólicas modernas ambas fuentes están disponibles y podrían usarse simultáneamente.

175

Al combinar ambas fuentes, debe que descubrirse, como dependen estos datos el uno del otro y si es posible omitir, en parte o completamente, alguno de los sistemas.

El objetivo principal de esta tesis es mejorar las actuales prácticas en Operación y Mantenimiento de turbinas eólicas desarrollando, en primer lugar, modelos avanzados de fiabilidad basados en las condiciones ambientales y operativas existentes realmente en el parque. Y, en segundo lugar, proponer nuevos algoritmos de detección de fallos utilizando datos ambientales, operativos y CMS. La consecución de estos objetivos se logró mediante su división en cuatro tareas independientes que incluyeron: (1) análisis de datos de fallos y paradas de turbina eólicas; (2) comprensión del comportamiento de las condiciones meteorológicas previas a los fallos de los componentes principales; (3) desarrollo de un modelo de fiabilidad que incorpora dichas condiciones; y (4) predicción de fallos basada en condiciones ambientales o datos de vibraciones (CMS) y operacionales.

Se ha realizado un extenso análisis de fallos teniendo en cuenta los datos históricos de alrededor de 4600 aerogeneradores de diferentes tecnologías, incluyendo varias configuraciones de transmisión, antigüedad, regulación de potencia, etc. Para esto, en primer lugar, se completó una taxonomía que permite clasificar los distintos elementos de la turbina eólica en 7 componentes principales, 45 sistemas y 199 subsistemas. El desarrollo de esta taxonomía era necesario, ya que apenas existía información, disponible públicamente, sobre este tipo de descomposición por componentes. Para ello, se trabajó partiendo de resultados existentes, que estaban basados principalmente en tecnologías anticuadas, con la incorporación de información de turbinas modernas de diferentes tecnologías. La aplicación de esta taxonomía a la base de datos permitió determinar las tasas anuales de fallos y el tiempo de parada para cada componente general y sistema.

Los resultados obtenidos y presentados en esta tarea mejoran significativamente los estudios previos sobre fallos de aerogeneradores, ya que proporcionan una visión detallada del comportamiento de fallo de diferentes componentes y tecnologías, incluyendo tanto las antiguas como las más modernas. Además, también se realizó un análisis del modo de fallo para cada tipo de turbina. Los resultados se presentaron para turbinas con multiplicadora y de eje directo con capacidades nominales inferiores, iguales o superior a 1 MW, así como turbinas de paso de pala ('pitch') variable y turbinas reguladas por pérdida aerodinámica. Se demostró que las turbinas con

multiplicadora tienen mayores tasas de fallo y tiempos de parada que las turbinas de eje directo, resultado que es consistente con investigaciones previas. Además, las tasas de fallo y los tiempos de parada relacionados con el componente principal y el sistema obtenidos para cada una de las tecnologías analizadas difirieron en algunos casos notablemente. El análisis del modo de fallo mostró que a menudo existen modos de fallo dominantes para ciertos componentes. Estos incluyen, por ejemplo problemas de rodamientos en la multiplicadora, grietas superficiales en las palas; fallos en el anillo de orientación y problemas de rodamientos que conducen a fallos del generador. Todos los resultados presentados pueden ser muy importantes para las actividades de O&M predictivo de turbinas eólicas, ya que sirven como datos de entrada para modelos de fiabilidad y mantenimiento. Además, los operadores y fabricantes pueden obtener información sobre qué componentes son los más afectados para cada una de las tecnologías estudiadas, así como también sus modos de fallo.

Posteriormente, se procedió al estudio del efecto de las condiciones meteorológicas existentes antes de los fallos de cinco de los componentes principales. Estos componentes principales se identificaron en el análisis de fallos anterior como los más críticos en términos de tasas de fallo y tiempo de parada. Inicialmente, se desarrolló un marco de aprendizaje basado en datos utilizando téccnias de búsqueda de reglas apriorísticas ('apriori rule mining') para analizar las condiciones durante un periodo de tiempo relativamente corto antes del fallo. Se incluyeron aquí las condiciones generales durante todo el periodo de observación, las combinaciones de variables ambientales en el momento exacto del fallo, así como las variaciones en los parámetros. Los resultados muestran que los fallos ocurrieron principalmente durante el invierno, con altas humedades relativas, bajas temperaturas y velocidades de viento bastante altas (especialmente para el sistema de control de paso del ángulo de pala, 'pitch'). Además, se registró una pérdida en la eficiencia de producción antes de todos los fallos de los componentes.

Como los resultados previos habían demostrado que velocidades altas de viento afectan al comportamiento de fallo de los componentes, y se suponía que las variaciones y patrones recurrentes en las series temporales de velocidad de viento podían afectar negativamente a la vida útil de dichos componentes, se procedió a analizar las series temporales de velocidad de viento utilizando técnicas de detección de anomalías y descubrimiento de patrones repetitivos. Los resultados muestran que

durante el mes del fallo, el número promedio de anomalías detectadas antes del fallo del componente para cada serie temporal es mucho más alto que durante un mes 'regular'. Por tanto, se puede concluir que las variaciones en la velocidad de viento (y no solo las velocidades medias) afectan mucho a los componentes, especialmente a las palas, el convertidor y el sistema de control de paso de pala, 'pitch'. La técnica para la deteción de patrones también fue capaz de detectar patrones recurrentes antes del fallo del convertidor. Sin embargo, hay que decir que esta técnica debería mejorarse para obtener resultados más fiables. Los hallazgos relativos a las condiciones meteorológicas que producen fallos de los componentes contribuyen aún más al campo y amplían la investigación previa al indicar claramente qué factores influyen en el comportamiento de fallo y al proporcionar una herramienta que permite analizar grandes bases de datos en estudios futuros.

Después de haber demostrado que ciertas condiciones ambientales afectan al comportamiento de fallo de los componentes, se desarrolló un modelo avanzado de fiabilidad teniendo en cuenta estas condiciones. Para el desarrollo de este modelo, se utilizaron los datos meteorológicos obtenidos durante un periodo más largo antes del fallo (dos meses). El modelo se basa en dos procesos separados, uno que genera los eventos de fallo así como ceros ocasionales con una distribución binomial negativa, y un segundo proceso que genera los ceros estructurales siguiendo una distribución binomial. Los posibles efectos no observados (heterogeneidad) se tienen en cuenta en el modelo utilizando una distribución Gamma. No se han encontrado, hasta la fecha, trabajos previos en los que se hayan considerado combinaciones complejas de las condiciones para modelar el comportamiento de fallo de turbinas eólicas. Este modelo ha demostrado comportarse muy bien tanto para predecir sucesos globales del aerogenerador como para detectar fallos de componentes específicos, por ejemplo la multiplicadora. Cabe destacar además que los problemas más comunes de los modelos estadísticos, como por ejemplo el 'over-fitting', sobredispersión y multicolinealidades, se han evitado utilizando técnicas de regularización avanzadas. El modelo aquí desarrollado contribuye a la investigación actual proporcionando un modelo de fiabilidad novedoso y más realista que los existentes hasta la fecha por lo que se recomienda para su aplicación posterior en campo.

Para la detección de fallos a nivel de parque, se planteó y entrenó un clasificador bayesiano basado también en las condiciones ambientales específicas del emplazamiento. En lugar de predecir el número de fallos como en los desarrollos previos,

el modelo detecta la probabilidad de tener uno o más fallos dentro de un parque eólico a partir de las condiciones climatológicas. El sistema predijo de manera fiable la mayoría de los fallos de componentes dados en un conjunto de datos de prueba. Además, se extrajeron las probabilidades condicionales de cada parámetro meteorológico de entrada y se compararon con los resultados de los capítulos 3 a 5. Esta técnica proporciona una herramienta valiosa para los operadores de parques eólicos, y les permite detectar problemas en el parque, incluso sin tener que analizar datos extensos de CMS y SCADA.

Además de los métodos previamente descritos, la detección de fallos basada en datos CMS y SCADA se está convirtiendo en un técnica bien establecida y los fallos pueden predecirse mucho antes del suceso real con incertidumbres bastante bajas. Sin embargo, la investigación actual se centra en datos SCADA o CMS. La combinación de ambas fuentes con el objetivo de mejorar la detección de fallos no ha estado sujeta a investigaciones previas. Por lo tanto, en el útimo capítulo de esta tesis se exploró la detección de fallos en componentes de aerogeneradores utilizando una combinación de datos SCADA y CMS. Para esto, se investigó la relación entre las diferentes fuentes de datos. Luego, se propuso una técnica para fusionar datos CMS de vibraciones y SCADA con diferentes resoluciones temporales. Como las alarmas de CMS generalmente ocurren mucho antes del suceso de fallo, la predicción de estas alarmas implica una detección temprana del fallo. Las alarmas se predijeron con varios algoritmos de inteligencia artificial. Se demostró que los 'Random Forests' y los 'Gradient Boosting Machines' generan muy buenos resultados para predecir las alarmas CMS y la probabilidad de tener una alarma CMS a lo largo del tiempo. Se demostró además que es posible predecir las alarmas del CMS solo con datos SCADA. Esto lleva a la suposición de que, en ciertas ocasiones, puede no existir una razón de peso para utilizar el análisis de datos de CMS. La precisión de predicción se mejoró cuando se incluyeron los datos de vibración CMS en los algoritmos. Finalmente, se presentó el entorno DAVE, que es una herramienta 'on-line' de detección de fallos a nivel de parques eólicos basada únicamente en vibraciones (CMS). Utiliza técnicas de 'Dynamic Time Warping' para calcular las distancias entre los espectros de vibraciones obtenidos para el mismo componente de diferentes aerogeneradores. Esta herramienta ha demostrado que puede detectar fallos hasta 119 días antes de su ocurrencia real y superando así al sistema de alarma integrado en el sistema de CMS.

9.1 Trabajo Futuro

El trabajo futuro podría centrarse en incluir alarmas SCADA en los modelos y herramientas presentados aquí. Se supone que ciertas alarmas, incluso si no son activadas por el componente respectivo, pueden indicar componentes defectuosos. Por ejemplo, una combinación de alarmas del sistema de refrigeración y de alta temperatura interior podría indicar un problema en la multiplicadora incluso antes de que se emita una alarma de dicho componente. Esta información, junto con las variables ambientales, operacionales y de vibración, podría ser capaz de mejorar la predicción de fallo para ciertos componentes de la turbina.

Las condiciones ambientales previas a los fallos podrían analizarse con más detalle, centrándose en sus efectos acumulativos y de largo plazo en los componentes. Esto podría llevarse a cabo adaptando el entorno de aprendizaje presentado en esta tesis.

Además, se podrían desarrollar herramientas comerciales de monitorización en línea basadas en el clasificador bayesiano y el entorno DAVE presentado en esta tesis. Para esto, sería necesario probar estos enfoques en escenarios más grandes a fin de garantizar su funcionamiento.

El descubrimiento de patrones ha mostrado un gran potencial para detectar patrones recurrentes en series temporales de velocidad del viento, esto debe analizarse más detalladamente teniendo en cuenta técnicas apropiadas para evaluar los resultados del algoritmo.

Como los enfoques presentados en esta tesis persiguen objetivos ligeramente diferentes, sería interesante combinar los algoritmos de descubrimiento de anomalías y descubrimiento de patrones, el entorno DAVE, el clasificador bayesiano y el modelo de fiabilidad, en una herramienta predictiva para O&M, que combinara los diferentes objetivos y generase varias formas de alertas tempranas.

Bibliography

[1] D. Fraile, A. Mbistrova, I. Pineda (editor), P. Tardieu (editor), Wind in power 2017: Annual combined onshore and offshore wind energy statistics, Tech. rep., Wind Europe (2018).

[2] A. Nghiem, I. Pineda, P. Tardieu (editor), Wind Energy in Europe: Scenarios for 2030, Tech. Rep. September, Wind Europe (2017).

[3] M. Tomescu, I. Moorkens, F. Meinke-Hubeny, L. Emele, E. Laes, EEA Report No 23/2017 - Renewable energy in Europe: Recent growth and knock-on effects, Tech. rep., European Environment Agency (2017).

[4] J. Carroll, A. McDonald, I. Dinwoodie, D. McMillan, M. Revie, I. Lazakis, Availability, operation and maintenance costs of offshore wind turbines with different drive train configurations, Wind Energy 20 (2) (2017) 361–378. `doi:10.1002/we.2011`.

[5] International Renewable Energy Agency, Renewable Power Generation Costs in 2017, Tech. rep., IRENA, Abu Dhabi (2018). `doi:10.1007/SpringerReference_7300`.

[6] R Core Team, R: A Language and Environment for Statistical Computing, 2015, R Foundation for Statistical Computing, Vienna, Austria, Available online, last accessed: 02/05/2018.
URL `https://www.R-project.org/`

[7] G. A. M. van Kuik, J. Peinke, R. Nijssen, D. Lekou, J. Mann, J. N. Sørensen, C. Ferreira, J. W. van Wingerden, D. Schlipf, P. Gebraad, H. Polinder, A. Abrahamsen, G. J. W. van Bussel, J. D. Sørensen, P. Tavner, C. L. Bottasso, M. Muskulus, D. Matha, H. J. Lindeboom, S. Degraer, O. Kramer,

S. Lehnhoff, M. Sonnenschein, P. E. Sørensen, R. W. Künneke, P. E. Morthorst, K. Skytte, Long-term research challenges in wind energy – a research agenda by the European Academy of Wind Energy, Wind Energy Science 1 (1) (2016) 1–39. `doi:10.5194/wes-1-1-2016`.

[8] AWESOME - Project Website, available online, last accessed: 06/04/2018. URL `http://awesome-h2020.eu`

[9] J. F. Manwell, J. G. McGowan, A. L. Rogers, Wind Energy Explained: Theory, Design and Application, 2nd Edition, John Wiley & Sons Inc., Massachusetts, USA, 2010.

[10] D. Rivkin, L. Silk, Wind Turbine Operations, Maintenance, Diagnosis, and Repair, Jones & Bartlett Learning, Burlington, MA, USA, 2012.

[11] R. R. Hill, V. Peters, J. Stinebaugh, P. S. Veers, Sandia Report: Wind Turbine Reliability Database Update, Tech. Rep. March, Sandia National Laboratories (2009).

[12] A. Stenberg, Wind energy statistics in Finland, in: International Statistical Analysis on Wind Turbine Failures, Kassel, Germany, 2011, pp. 117–122.

[13] VGB-PowerTech, VGB-Standard RDS-PP Application specification Part 32: Wind energy, Tech. rep., VGB-PowerTech, Essen, Germany (2014).

[14] M. Wilkinson, B. Hendriks, F. Spinato, T. Van Delft, Measuring wind turbine reliability, results of the reliawind project, in: European Wind Energy Association Conference, 2011, pp. 1–8.

[15] R. Wiser, M. Bolinger, 2014 Wind Technologies Market Report, Tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO (United States) (2015). `doi:10.2172/1155074`.

[16] A. Stenberg, H. Holttinen, Analysing failure statistics of wind turbines in Finland, in: European Wind Energy Conference and Exhibition, 2010, pp. 20–23.

[17] WindStats, Newsletter Volume 12 No 4 (Autumn 1999) to Volume 26 No 1 (Spring 2013), Tech. rep., Denmark.

[18] P. J. Tavner, J. Xiang, F. Spinato, Reliability analysis for wind turbines, Wind Energy 10 (1) (2007) 1–18. `doi:10.1002/we.204`.

[19] P. J. Tavner, F. Spinato, G. J. W. van Bussel, E. Koutoulakos, Reliability of Different Wind Turbine Concepts with Relevance to Offshore Application, in: Proceedings of the European Wind Energy Conference, 2008.

[20] B. Hahn, Zuverlässigkeit, Wartung und Betriebskosten von Windkraftanlagen, Tech. rep., Institut für Solare Energieversorgungstechnik Verein, Universität Kassel, Kassel (2003).

[21] B. Hahn, M. Durstewitz, K. Rohrig, Reliability of Wind Turbines, in: Wind Energy, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 329–332. `doi:10.1007/978-3-540-33866-6_62`.

[22] S. Faulstich, M. Durstewitz, B. Hahn, K. Knorr, K. Rohrig, Windenergy Report Germany 2008: Written within the Research Project Deutscher Windmonitor, Tech. rep., German Federal Ministry for the Environment Nature Conversation and Nuclear Safety, Bonn, Germany (2009).

[23] E. Echavarria, B. Hahn, G. J. W. van Bussel, T. Tomiyama, Reliability of Wind Turbine Technology Through Time, Journal of Solar Energy Engineering 130 (3) (2008) 031005. `doi:10.1115/1.2936235`.

[24] J. Gayo, Project final report: ReliaWind, Tech. Rep. November, Gamesa Innovation and Technology, Egues, Spain (2011).

[25] J. Carroll, A. McDonald, D. McMillan, Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines, Wind Energy 19 (6) (2016) 1107–1119. `doi:10.1002/we.1887`.

[26] J. Carroll, A. McDonald, D. McMillan, Reliability Comparison of Wind Turbines With DFIG and PMG Drive Trains, IEEE Transactions on Energy Conversion 30 (2) (2015) 663–670. `doi:10.1109/TEC.2014.2367243`.

[27] J. Ribrant, L. M. Bertling, Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005, IEEE Transactions on Energy Conversion 22 (1) (2007) 167–173. `doi:10.1109/TEC.2006.889614`.

[28] S. Pfaffel, S. Faulstich, K. Rohrig, Performance and Reliability of Wind Turbines: A Review, Energies 10 (11) (2017) 1904. `doi:10.3390/en10111904`.

[29] M. Reder, E. Gonzalez, J. J. Melero, Wind Turbine Failures - Tackling current Problems in Failure Data Analysis, Journal of Physics: Conference Series 753 (072027). `doi:10.1088/1742-6596/753/7/072027`.

[30] V. A. Peters, A. B. Ogilvie, C. R. Bond, Continuous Reliability Enhancement for Wind (CREW) Database : Wind Plant Reliability Benchmark, Tech. rep., Sandia National Laboratories, Albuquerque, NM, USA (2012).

[31] C. Carter, B. Karlson, S. Martin, C. Westergaard, Continuous Reliability Enhancement for Wind (CREW) Program Update, Tech. rep., Sandia National Laboratories, Albuquerque, NM, USA (2016).

[32] Y. Lin, L. Tu, H. Liu, W. Li, Fault analysis of wind turbines in China, Renewable and Sustainable Energy Reviews 55 (2016) 482–490. `doi:10.1016/j.rser.2015.10.149`.

[33] N. Carlstedt, Driftuppföljning av Vindkraftverk: Arsrapport 2012: > 50 kW, (2013), available online, last accessed: 06/04/2018.
URL `http://www.vindstat.nu/stat/Reports/arsrapp2012.pdf`

[34] G. J. W. Van Bussel, M. B. Zaaijer, Estimation of Turbine Reliability Figures within the DOWEC Project, (2003), Tech. Rep. 10048, available online, last accessed: 06/04/2018.
URL `https://www.ecn.nl/fileadmin/ecn/units/wind/docs/dowec/10048_004.pdf`

[35] J. Schmid, H. Klein, Performance of European Wind Turbines: A Statistical Evaluation from the European Wind Turbine Database EUROWIN, Elsevier Applied Science (2003).

[36] J. Schmid, H. Klein, EUROWIN. The European Windturbine Database. Annual Reports. A Statistical Summary of European WEC Performance Data for 1992 and 1993, Tech. rep., Fraunhofer Institute for Solar Energy Systems, Freiburg, Germany (1994).

186

[37] K. Harman, R. Walker, M. Wilkinson, Availability trends observed at operational wind farms, in: Proceedings of the European Wind Energy Conference, Brussels, Belgium, 2008.

[38] Jiangtao Chai, Guoyin An, Zhiyong Ma, Xiaowei Sun, A study of fault statistical analysis and maintenance policy of wind turbine system, in: International Conference on Renewable Power Generation (2015), Institution of Engineering and Technology, 2015, p. 4. `doi:10.1049/cp.2015.0443`.

[39] R. Lynette, Status of the U.S. wind power industry, Journal of Wind Engineering and Industrial Aerodynamics 27 (1-3) (1988) 327–336. `doi:10.1016/0167-6105(88)90047-5`.

[40] G. J. Herbert, S. Iniyan, R. Goic, Performance, reliability and failure analysis of wind farm in a developing Country, Renewable Energy 35 (12) (2010) 2739–2751. `doi:10.1016/j.renene.2010.04.023`.

[41] Commitee for Increase in Availability/Capacity Factor of Wind Turbine Generator System and Failure/Breakdown Investigation of Wind Turbine Generator Systems Subcommittee: Summary Report, Tech. rep., New Energy Industrial Technology Development Organization, Kanagawa, Japan (2004).

[42] J. A. Andrawus, J. Watson, M. Kishk, Wind Turbine Maintenance Optimisation: Principles of Quantitative Maintenance Optimisation, Wind Engineering 31 (2) (2007) 101–110. `doi:10.1260/030952407781494467`.

[43] Y. Feng, P. J. Tavner, H. Long, Early experiences with UK Round 1 offshore wind farms, Proceedings of the Institution of Civil Engineers : Energy 163 (4) (2010) 167–181. `doi:10.1680/ener.2010.163.0.1`.

[44] C. Su, Y. Yang, X. Wang, Z. Hu, Failures analysis of wind turbines: Case study of a Chinese wind farm, Proceedings of 2016 Prognostics and System Health Management Conference, PHM-Chengdu 2016 (2017) 1–6`doi:10.1109/PHM.2016.7819826`.

[45] Portfolio Review 2016; System Performance, Availability and Reliability Trend Analysis (SPARTA); Northumberland, UK (2017).

[46] V. Turkia, H. Holttinen, Tuulivoiman tuotantotilastot: Vuosiraportti 2011, Tech. rep., VTT Technical Research Centre of Finland, Espoo, Finland (2013).

[47] S. Sheng, Report on Wind Turbine Subsystem Reliability - A Survey of Various Databases, Tech. rep., National Renewable Energy Laboratory, Golden, CO, USA (2013).

[48] P. J. Tavner, J. Xiang, F. Spinato, Reliability analysis for wind turbines, Wind Energy 10 (1) (2007) 1–18. `doi:10.1002/we.204`.

[49] Fraunhofer IWES, Complex Systems Require New Strategies and Methods, Tech. rep., Fraunhofer IWES, Munich, Germany (2017).

[50] S. Pfaffel, S. Faulstich, B. Hahn, Performance and reliability benchmarking using the cross-company initiative WInD-Pool, in: Proceedings of the RAVE Offshore Wind R&D Conference, Bremerhaven, Germany, 2015.

[51] S. Pfaffel, S. Faulstich, B. Hahn, J. Hirsch, V. Berkhout, H. Jung, Monitoring and Evaluation Program for Offshore Wind Energy Use. Implementation Phase, Tech. rep., Fraunhofer-Institut für Windenergie und Energiesystemtechnik, Kassel, Germany (2016).

[52] H. Jung, S. Pfaffel, S. Faulstich, F. Bübl, J. Jensen, R. Jugelt, Abschlussbericht: Erhöhung der Verfügbarkeit von Windenergieanlagen EVW-Phase 2, Tech. rep., FGW e.V. Wind Energy and Other Decentralized Energy Organizations, Berlin, Germany (2015).

[53] S. Sheng, Wind Turbine Gearbox Condition Monitoring Round Robin Study - Vibration Analysis, Tech. rep., National Renewable Energy Lab. (NREL), Golden, CO (United States) (2012). `doi:10.2172/1048981`.

[54] S. Sheng, Gearbox Typical Failure Modes, Detection and Mitigation Methods, in: AWEA Operations & Maintenance and Safety Seminar, National Renewable Energy Laboratory, 2014, pp. 1–24.

[55] B. Hahn, Zeitlicher Zusammenhang von Schadenshäufigkeit und Windgeschwindigkeit, FGW-Workshop Einfluss der Witterung auf Windenergieanlagen,.

[56] P. Tavner, C. Edwards, A. Brinkman, F. Spinato, Influence of Wind Speed on Wind Turbine Reliability, Wind Engineering 30 (1) (2006) 55–72. `doi: 10.1260/030952406777641441`.

[57] S. Faulstich, P. Lyding, P. Tavner, Effects of wind speed on wind turbine availability, European Wind Energy Conference.

[58] P. Tavner, D. M. Greenwood, M. W. G. Whittle, R. Gindele, S. Faulstich, B. Hahn, Study of weather and location effects on wind turbine failure rates, Wind Energy 16 (2) (2013) 175–187. `doi:10.1002/we.538`.

[59] M. Wilkinson, T. Van Delft, K. Harman, The Effect of Environmental Parameters on Wind Turbine Reliability, European Wind Energy Conference, Copenhagen, Denmark (2012).

[60] G. Sheikholeslami, C. Science, B. Buffalo, S. Chatterjee, A. Zhang, Wave-Cluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases, in: VLDB '98 Proceedings of the 24rd International Conference on Very Large Data Bases, New York, USA, 1998, pp. 428–439.

[61] J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, in: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Vol. 1, Berkeley, California, 1967, pp. 281–297.

[62] E. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications, Biometrics 21 (3) (1965) 768–769.

[63] S. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory 28 (2) (1982) 129–137. `doi:10.1109/TIT.1982.1056489`.

[64] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set, Journal of Statistical Software 61 (6) (2014) 1–36. `doi:10.18637/jss.v061.i06`.

[65] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, R-Package: Determining the Best Number of Clusters in a Data Set (2015).

[66] T. A. Sørensen, A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species and its Application to Analyses of the Vegetation on Danish Commons, Biologiske Skrifter 5 (1948) 1–34.

[67] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, ACM SIGMOD Record 22 (2) (1993) 207–216. `doi:10.1145/170036.170072`.

[68] M. Zaki, Scalable algorithms for association mining, IEEE Transactions on Knowledge and Data Engineering 12 (3) (2000) 372–390. `doi:10.1109/69.846291`.

[69] J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, in: Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00, ACM Press, New York, New York, USA, 2000, pp. 1–12. `doi:10.1145/342009.335372`.

[70] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules, in: VLDB '94 Proceedings of the 20th International Conference on Very Large Databases, 1994, pp. 487 – 499.

[71] M. Hahsler, A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, (2015), available online, last accessed: 02/05/2018. URL `http://michael.hahsler.net/research/association_rules/measures.html`

[72] P.-N. Tan, M. Steinbach, V. Kumar, Association Analysis: Basic Concepts and Algorithms, in: Introduction to Data mining, 2005, Ch. 6, pp. 327–414. `doi:10.1111/j.1600-0765.2011.01426.x`.

[73] G. Wilson, D. McMillan, Assessing Wind Farm Reliability Using Weather Dependent Failure Rates, Journal of Physics: Conference Series 524 (2014) 1–10. `doi:10.1088/1742-6596/524/1/012181`.

[74] M. Kanamitsu, W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, G. L. Potter, NCEP–DOE AMIP-II Reanalysis (R-2), Bulletin of the American Meteorological Society 83 (11) (2002) 1631–1644. `doi:10.1175/BAMS-83-11-1631`.

[75] M. U. Kemp, E. Emiel van Loon, J. Shamoun-Baranes, W. Bouten, RNCEP: global weather and climate data at your fingertips, Methods in Ecology and Evolution 3 (1) (2012) 65–70. `doi:10.1111/j.2041-210X.2011.00138.x`.

[76] G. Jiang, J. Keller, P. L. Bond, Determining the long-term effects of H2S concentration, relative humidity and air temperature on concrete sewer corrosion, Water Research 65 (2014) 157–169. `doi:10.1016/j.watres.2014.07.026`.

[77] G. O. Lloyd, Atmospheric corrosion, in: ACIDITY - Sources, Consequences and Abatement, Elsevier Science Publishers Ltd, 2000, pp. 1–8.

[78] Y. Xiang, Z. Wang, Z. Li, W. Ni, Long term corrosion of X70 steel and iron in humid supercritical CO 2 with SO 2 and O 2 impurities, Corrosion Engineering, Science and Technology 48 (5) (2013) 395–398. `doi:10.1179/1743278213Y.0000000099`.

[79] M. Reder, J. J. Melero, Assessing Wind Speed Effects on Wind Turbine Reliability, in: Wind Europe Conference, Hamburg, Germany, 2016. `doi:10.13140/RG.2.2.11134.59200`.

[80] M. Reder, J. Melero, Time series data mining for analysing the effects of wind speed on wind turbine reliability, in: M. Cepin, R. Bris (Eds.), Safety and Reliability - Theory and Applications, Proceedings of the European Safety and Reliability Conference, Portoroz, Slovenia, 2017. `doi:10.1201/9781315210469-93`.

[81] T. C. Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (1) (2011) 164–181. `doi:10.1016/j.engappai.2010.09.007`.

[82] S. S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium 2010, IEEE, 2010, pp. 1–8. `doi:10.1109/NAPS.2010.5619586`.

[83] J. Tautz-Weinert, S. J. Watson, Using SCADA data for wind turbine condition monitoring – a review, IET Renewable Power Generation 11 (4) (2017) 382–394. `doi:10.1049/iet-rpg.2016.0248`.

[84] V.
Surveys 41 (3) (2009) 1–58. `doi:10.1145/1541880.1541882`.

[85] D. Cheboli, Anomaly Detection of Time Series, *Masters Thesis*. University of Minnesota (2010).

[86] C. Kleist, Time Series Data Mining Methods: A Review, *Masters Thesis*. Humboldt-Universität zu Berlin School (2015).

[87] O. Vallis, J. Hochenbaum, A. Kejariwal, A Novel Technique for Long-Term Anomaly Detection in the Cloud, in: 6th USENIX Workshop on Hot Topics in Cloud Computing, Philadelphia, PA, 2014.

[88] B. Rosner, On the Detection of Many Outliers, Technometrics 17 (2) (1975) 221–227. `doi:10.1080/00401706.1975.10489305`.

[89] B. Rosner, Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics 25 (2) (1983) 165. `doi:10.2307/1268549`.

[90] R. B. Cleveland, W. S. Cleveland, J. E. McRae, I. Terpenning, STL: A Seasonal-Trend Decomposition Procedure Based on Loess, Journal of Official Statistics 6 (1) (1990) 3–73.

[91] F. E. Grubbs, Sample Criteria for Testing Outlying Observations, The Annals of Mathematical Statistics 21 (1) (1950) 27–58. `doi:10.1214/aoms/1177729885`.

[92] P. J. Huber, E. M. Ronchetti, Robust Statistics, 2nd Edition, John Wiley & Sons, Inc., Cambridge, Massachusetts, USA, 2009.

[93] Information Technology Laboratory - US Department of Commerce, Extreme Studentized Deviate Test, (2015), available online, last accessed: 06/04/2018.
URL `http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/esd.htm`

[94] J. Lin, E. Keogh, S. Lonardi, P. Patel, Finding motifs in time series, in: Proceedings of the 2nd Workshop on Temporal Data Mining, 2002, pp. 53–68. `doi:10.1.1.19.6629`.

[95] P. Patel, E. Keogh, J. Lin, S. Lonardi, Mining motifs in massive time series databases, in: 2002 IEEE International Conference on Data Mining, 2002. Proceedings., 2002, pp. 370–377. `doi:10.1109/ICDM.2002.1183925`.

[96] D. J. Berndt, J. Clifford, Using Dynamic Time Warping to FindPatterns in Time Series, in: 1994 Workshop on Knowledge Discovery in Databases, volume 10, Seattle, WA, 1994, pp. 359–370.

[97] B. Chiu, E. Keogh, S. Lonardi, Probabilistic discovery of time series motifs, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining KDD 03 304 (2003) 493. `doi:10.1145/956750.956808`.

[98] D. Yankov, E. Keogh, J. Medina, B. Chiu, V. Zordan, Detecting time series motifs under uniform scaling, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining KDD 07, 2007, p. 844. `doi:10.1145/1281192.1281282`.

[99] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: A novel symbolic representation of time series, Data Mining and Knowledge Discovery 15 (2) (2007) 107–144. `doi:10.1007/s10618-007-0064-z`.

[100] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A Symbolic Representation of Time Series, with Implications for Streaming Algorithms, in: SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003, pp. 2–11. `doi:10.1145/882082.882086`.

[101] P. A. Pevzner, S. H. Sze, Combinatorial approaches to finding subtle signals in DNA sequences, Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology 8 (2000) 269–278.

[102] J. Buhler, M. Tompa, Finding Motifs Using Random Projections, Journal of Computational Biology 9 (2) (2002) 225–242. `doi:10.1089/10665270252935430`.

[103] C. E. Lawrence, A. A. Reilly, An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences, Proteins: Structure, Function, and Genetics 7 (1) (1990) 41–51. `doi:10.1002/prot.340070105`.

[104] T. L. Bailey, C. Elkan, The value of prior knowledge in discovering motifs with MEME, in: Proceedings of the International Conference on Intelligent Systems in Molecular Biology, 1995, p. 21ff.

[105] E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, A. J. S. López, Handbook of Research on Machine Learning Applications and Trends, IGI Global, 2010. `doi:10.4018/978-1-60566-766-9`.

[106] S. Moritz, , T. Bartz-Beielstein, Comparison of different methods for univariate time series imputation in R, The R Journal Vol. 9/1 (2017) 207–218.

[107] R. Hyndman, R Package 'forecast': Forecasting functions for time series and linear models, (2016), available online, last accessed: 26/04/2018.
URL   `https://cran.r-project.org/web/packages/forecast/forecast.pdf`

[108] E. Artigao, A. Honrubia-Escribano, E. Gomez-Lazaro, Current signature analysis to monitor DFIG wind turbine generators: A case study, Renewable Energy 116 (2018) 5–14. `doi:10.1016/j.renene.2017.06.016`.

[109] E. Artigao, S. Koukoura, A. Honrubia-Escribano, J. Carroll, A. McDonald, E. Gómez-Lázaro, Current Signature and Vibration Analyses to Diagnose an In-Service Wind Turbine Drive Train, Energies 11 (4) (2018) 960. `doi:10.3390/en11040960`.

[110] N. Castro, P. Azevedo, Time Series Motifs Statistical Significance, Proceedings of the SIAM International Conference on Data Mining (SDM) (2011) 687–698.

[111] M. Rausand, A. Hoyland, System Reliability Theory: Models and Statistical Methods, 2nd Edition, John Wiley & Sons, 2004.

[112] J. Faulin, A. A. Juan, S. Martorell, J. E. Ramirez-Marquez, Simulation Methods for Reliability and Availability of Complex Systems, Springer Series

in Reliability Engineering, Springer London, London, 2010. `doi:10.1007/978-1-84882-213-9`.

[113] P. D. T. O'Connor, A. Kleyner, Practical Reliability Engineering, 5th Edition, John Wiley & Sons, Ltd., West Sussex, United Kingdom, 2012.

[114] J. A. Nachlas, Reliability Engineering: Probabilistic Models and Maintenance Methods, 1st Edition, CRC Press Taylor & Francis Group, 2005.

[115] P. Hokstad, The failure intensity process and the formulation of reliability and maintenance models, Reliability Engineering & System Safety 58 (1) (1997) 69–82. `doi:10.1016/S0951-8320(97)00053-7`.

[116] J. J. Higgins, S. Keller-McNulty, Concepts in Probability and Stochastic Modeling, 1st Edition, Duxbury Press, 1995.

[117] J. F. C. Kingman, Poisson Processes, Oxford University Press, Oxford, UK, 1992.

[118] Information Technology Laboratory - US Department of Commerce, Engineering Statistics Handbook, (2015), available online, last accessed: 06/04/2018. URL `https://www.itl.nist.gov/div898/handbook/index.htm`

[119] S. M. Ross, Stochastic Processes, 2nd Edition, Wiley, 1996.

[120] L. A. Escobar, W. Q. Meeker, A Review of Accelerated Test Models, Statistical Science 21 (4) (2006) 552–577. `doi:10.1214/088342306000000321`.

[121] D. Cox, Models and Life-Tables Regression, Journal of the Royal Statistical Society. Series B (Methodological) 34 (2) (1972) 187–220.

[122] H. Guo, S. Watson, P. Tavner, J. Xiang, Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation, Reliability Engineering and System Safety 94 (6) (2009) 1057–1063. `doi:10.1016/j.ress.2008.12.004`.

[123] J. a. Andrawus, J. Watson, M. Kishk, Wind Turbine Maintenance Optimisation: principles of quantitative maintenance optimisation, Wind Engineering 31 (2) (2009) 101–110. `doi:10.1260/030952407781494467`.

[124] S. Faulstich, V. Berkhout, J. Mayer, D. Siebenlist, Modelling the failure behaviour of wind turbines, Journal of Physics: Conference Series 749 (2016) 1–11. `doi:10.1088/1742-6596/749/1/012019`.

[125] P. J. Tavner, J. Xiang, F. Spinato, Improving the Reliability of Wind Turbine Generation and its Impact on Overall Distribution Network Reliability, 18th International Conference on Electricity Distribution (4) (2005) 1–4. `doi:10.1049/cp:20051199`.

[126] F. Spinato, P. Tavner, Reliability-Growth Analysis of Wind Turbines from Fleet Field Data Fabio, in: ARTS Conference, Loughborough, 2007.

[127] F. Spinato, P. Tavner, G. Van Bussel, E. Koutoulakos, Reliability of wind turbine subassemblies., IET Renewable Power Generation 3 (4) (2009) 387–401. `doi:10.1049/iet-rpg.2008.0060`.

[128] J. Carroll, A. McDonald, D. McMillan, R. Bakhshi, Offshore Wind Turbine Sub-Assembly Failure Rates Through Time, in: EWEA Annual Event, 2015.

[129] V. Slimacek, B. H. Lindqvist, Reliability of wind turbines modeled by a Poisson process with covariates, unobserved heterogeneity and seasonality, Wind Energy 19 (11) (2016) 1991–2002. `doi:10.1002/we.1964`.

[130] D. Mcmillan, G. Ault, Towards Quantification of Condition Monitoring Benefit for Wind Turbine Generators, Components 18 (2007) 1–11.

[131] D. McMillan, G. Ault, Specification of Reliability Benchmarks for Offshore Wind Farms, Tech. rep., Institute for Energy & Environment, Department of Electronics and Electrical Engineering, University of Strathclyde, Glasgow (2008).

[132] T. S. Selwyn, R. Kesavan, Reliability Measures of Constant Pitch Constant Speed Wind Turbine with Markov Analysis at High Uncertain Wind, Procedia Engineering 38 (2012) 932–938. `doi:10.1016/j.proeng.2012.06.117`.

[133] I. Dinwoodie, F. Quail, D. McMillan, Analysis of Offshore Wind Turbine Operation and Maintenance Using a Novel Time Domain Meteo-Ocean Modeling Approach, in: Proceedings of ASME Turbo Expo, 2012, p. 847. `doi:10.1115/GT2012-68985`.

[134] F. Besnard, K. Fischer, L. B. Tjernberg, A model for the optimization of the maintenance support organization for offshore wind farms, IEEE Transactions on Sustainable Energy 4 (2) (2013) 443–450. `doi:10.1109/TSTE.2012.2225454.`

[135] T. V. Tzioiutzias, A. N. Platis, V. P. Koutras, Markov Modeling of the Availability of a Wind Turbine Utilizing Failures and Real Weather Data, in: Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO), 2016, pp. 186–196. `doi:10.1109/SMRLO.2016.40.`

[136] J. Ribrant, Reliability performance and maintenance. A survey of failures in wind power systems, MSc Thesis, KTH School of Electrical Engineering (2006).

[137] H. Seyr, A. Barros, M. Muskulus, The Impact of Maintenance Duration on the Downtime of an Offshore Wind Farm - Alternating Renewal Process, in: Proceedings of the 30th International Congress & Exhibition on Condition Monitoring and Diagnostic Engineering Management, 2017, pp. 1–5.

[138] R. Jiang, R. Huang, C. Huang, Modeling the effect of environmental conditions on reliability of wind turbines, Journal of Shanghai Jiaotong University (Science) 21 (4) (2016) 462–466. `doi:10.1007/s12204-016-1747-7.`

[139] R. Moghaddass, C. Rudin, The latent state hazard model, with application to wind turbine reliability, Annals of Applied Statistics 9 (4) (2015) 1823–1863. `doi:10.1214/15-AOAS859.`

[140] P. Bangalore, Load and Risk Based Maintenance Management of Wind Turbines, Ph.D. thesis, Chalmers University of Technology, Gothenburg, Sweden (2016).

[141] P. Mazidi, M. Du, L. Bertling Tjernberg, M. A. Sanz Bobi, A health condition model for wind turbine monitoring through neural networks and proportional hazard models, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 231 (5) (2017) 481–494. `doi:10.1177/1748006X17707902.`

[142] G. Wilson, G. Ault, D. McMillan, Modelling the effects of the environment on wind turbine failure modes using neural networks, in: International Conference on Sustainable Power Generation and Supply (SUPERGEN 2012), Institution of Engineering and Technology, Hangzhou, China, 2012, pp. 42–42. `doi: 10.1049/cp.2012.1768`.

[143] G. Wilson, D. McMillan, Modeling the relationship between wind turbine failure modes and the environment, Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013 (2014) 801–809.

[144] International Electrotechnical Commission, IEC 61400-2: Wind turbines - Part 2: Power performance of electricity-producing wind turbines based on nacelle anemometry, Geneva, Switzerland (2013).

[145] D. J. Denis, Applied Univariate, Bivariate, and Multivariate Statistics, John Wiley & Sons Inc., Hoboken, New Jersey, 2016.

[146] W. H. Greene, Econometric Analysis, seventh ed Edition, Pearson Education Limited, Pearson Education Limited, 2012.

[147] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society, Series B 58 (1994) 267–288.

[148] M. J. Whittingham, P. A. Stephens, R. B. Bradury, R. P. Freckleton, Why do we still use stepwise modelling in ecology and behaviour?, Journal of Animal Ecology 75 (5) (2006) 1182–1189. `doi:10.1111/j.1365-2656.2006.01141.x`.

[149] A. N. Tikhonov, V. Y. Arsenin, Solutions of Ill-Posed Problems, Winston Washington, DC, 1977.

[150] J. Fan, R. Li, Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties, Journal of the American Statistical Association 96 (456) (2001) 1348–1360. `doi:10.1198/016214501753382273`.

[151] C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, The Annals of Statistics 38 (2) (2010) 894–942. `doi:10.1214/09-AOS729`.

[152] H. Zou, T. Hastie, Addendum: Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (5) (2005) 768–768. `doi:10.1111/j.1467-9868.2005.00527.x`.

[153] J. Huang, P. Breheny, S. Lee, S. Ma, C.-H. Zhang, The Mnet method for variable selection, Statistica Sinica 26 (402) (2016) 903–923. `doi:10.5705/ss.202014.0011`.

[154] H. Akaike, Information theory and an extension of the maximum likelihood principle, Tsahkadsor, Armenia, USSR, in: F. Petrov, B.N.; Csáki (Ed.), 2nd International Symposium on Information Theory, Budapest, 1973, pp. 267–281.

[155] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716–723. `doi:10.1109/TAC.1974.1100705`.

[156] C. M. Hurvich, C.-L. Tsai, Regression and Time Series Model Selection in Small Samples, Biometrika 76 (2) (1989) 297. `doi:10.2307/2336663`.

[157] M. L. McHugh, The Chi-square test of independence, Biochemia Medica 23 (2) (2013) 143–149. `doi:10.11613/BM.2013.018`.

[158] G. Glatting, P. Kletting, S. N. Reske, K. Hohl, C. Ring, Choosing the optimal fit function: Comparison of the Akaike information criterion and the F-test, Medical Physics 34 (11) (2007) 4285–4292. `doi:10.1118/1.2794176`.

[159] P. Kletting, G. Glatting, Model selection for time-activity curves: The corrected Akaike information criterion and the F-test, Zeitschrift für Medizinische Physik 19 (3) (2009) 200–206. `doi:10.1016/j.zemedi.2009.05.003`.

[160] J. W. Tuckey, Some Graphic and Semigraphic Displays, in: T. Bancroft (Ed.), Statistical Papers in Honor of George W. Snedecor, The Iowa State University Press, Ames, Iowa, 1972, pp. 293–316.

[161] C. Kleiber, A. Zeileis, Visualizing Count Data Regressions Using Rootograms, The American Statistician 70 (2013) (2016) 1–25. `doi:10.1080/00031305.2016.1173590`.

[162] B. Liu, Z. An, H. Kou, Wind turbine gearbox reliability analysis based on the system level stress-strength model, Journal of Shanghai Jiaotong University (Science) 21 (4) (2016) 484–488. `doi:10.1007/s12204-016-1751-y`.

[163] A. Ragheb, M. Ragheb, Wind turbine gearbox technologies, in: 2010 1st International Nuclear & Renewable Energy Conference (INREC), IEEE, 2010, pp. 1–8. `doi:10.1109/INREC.2010.5462549`.

[164] M.-H. Evans, White structure flaking (WSF) in wind turbine gearbox bearings: effects of 'butterflies' and white etching cracks (WECs), Materials Science and Technology 28 (1) (2012) 3–22. `doi:10.1179/026708311X13135950699254`.

[165] A. Greco, Material wear and fatigue in wind turbine systems, Wear 302 (1-2) (2013) 1583–1591. `doi:10.1016/j.wear.2013.01.060`.

[166] Society of Tribologists and Lubrication Engineers, Preprints Presented at the STLE 1993 Annual Meeting, no. 1 v. 48, pt. 1993, Society of Tribologists and Lubrication Engineers, 1993.

[167] P. Guo, N. Bai, Wind Turbine Gearbox Condition Monitoring with AAKR and Moving Window Statistic Methods, Energies 4 (11) (2011) 2077–2093. `doi:10.3390/en4112077`.

[168] R.-G. Cong, M. Brady, The Interdependence between Rainfall and Temperature: Copula Analyses, The Scientific World Journal 2012 (2012) 1–11. `doi:10.1100/2012/405675`.

[169] H. Zhang, The Optimality of Naive Bayes, Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference FLAIRS 2004 1 (2) (2004) 1 – 6. `doi:10.1016/j.patrec.2005.12.001`.

[170] I. Rish, An empirical study of the naive bayes classifier, Tech. rep. (2001).

[171] D. J. Hand, K. Yu, Idiot's Bayes – Not So Stupid After All?, International Statistical Review 69 (3) (2001) 385–398. `doi:10.1111/j.1751-5823.2001.tb00465.x`.

[172] S. Wan, M. Mak, Machine Learning for Protein Subcellular Localization Prediction, De Gruyter, 2015.

[173] P. Tsangaratos, I. Ilia, Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size, CATENA 145 (2016) 164–179. `doi:10.1016/j.catena.2016.06.004`.

[174] T. Wei, Wind Power Generation and Wind Turbine Design, WIT Press, Southampton, Boston, 2010.

[175] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, P. J. Tavner, Monitoring wind turbine gearboxes, Wind Energy 16 (5) (2013) 728–740. `doi:10.1002/we.1521`.

[176] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, Y. Qiu, Wind turbine condition monitoring: technical and commercial challenges, Wind Energy 17 (5) (2014) 673–693. `doi:10.1002/we.1508`.

[177] L. Colone, M. Reder, J. Tautz-Weinert, J. Melero, A. Natarajan, S. Watson, Optimisation of Data Acquisition in Wind Turbines with Data-Driven Conversion Functions for Sensor Measurements, Energy Procedia 137 (2017) 571–578. `doi:10.1016/j.egypro.2017.10.386`.

[178] A. Kusiak, Z. Zhang, Analysis of Wind Turbine Vibrations Based on SCADA Data, Journal of Solar Energy Engineering 132 (3) (2010) 031008. `doi:10.1115/1.4001461`.

[179] P. Lind, L. Vera-Tudela, M. Wächter, M. Kühn, J. Peinke, Normal Behaviour Models for Wind Turbine Vibrations: Comparison of Neural Networks and a Stochastic Approach, Energies 10 (12) (2017) 1944. `doi:10.3390/en10121944`.

[180] W. Yang, R. Court, J. Jiang, Wind turbine condition monitoring by the approach of SCADA data analysis, Renewable Energy 53 (2013) 365–376. `doi:10.1016/j.renene.2012.11.030`.

[181] A. Kusiak, W. Li, The prediction and diagnosis of wind turbine faults, Renewable Energy 36 (1) (2011) 16–23. `doi:DOI:10.1016/j.renene.2010.05.014`.

[182] B. Chen, P. C. Matthews, P. J. Tavner, Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS, Expert Systems with Applications 40 (17) (2013) 6863–6876. `doi:DOI:10.1016/j.eswa.2013.06.018`.

[183] E. Gonzalez, M. Reder, J. J. Melero, SCADA alarms processing for wind turbine component failure detection, Journal of Physics: Conference Series 753 (072019). `doi:DOI:10.1088/1742-6596/753/7/072019`.

[184] M. Reder, E. Gonzalez, J. J. Melero, Wind Turbine Failures - Tackling current Problems in Failure Data Analysis, Journal of Physics: Conference Series 753 (072027). `doi:10.1088/1742-6596/753/7/072027`.

[185] C. S. Gray, S. J. Watson, Physics of Failure approach to wind turbine condition based maintenance, Wind Energy 13 (5) (2010) 395–405. `doi:10.1002/we.360`.

[186] Y. Qiu, L. Chen, Y. Feng, Y. Xu, An Approach of Quantifying Gear Fatigue Life for Wind Turbine Gearboxes Using Supervisory Control and Data Acquisition Data, Energies 10 (8) (2017) 1084. `doi:10.3390/en10081084`.

[187] K. Kim, G. Parthasarathy, O. Uluyol, W. Foslien, S. Sheng, P. Fleming, Use of SCADA data for failure detection in wind turbines, in: ASME 5th International Conference on Energy Sustainability, 2011, pp. 2071–2079. `doi:10.1115/ES2011-54243`.

[188] M. Schlechtingen, I. F. Santos, S. Achiche, Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description, Applied Soft Computing 13 (1) (2013) 447–460. `doi:10.1016/j.asoc.2013.09.016`.

[189] P. Sun, J. Li, C. Wang, X. Lei, A generalized model for wind turbine anomaly identification based on SCADA data, Applied Energy 168 (2016) 550–567. `doi:10.1016/j.apenergy.2016.01.133`.

[190] P. Bangalore, M. Patriksson, Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines, Renewable Energy 115 (2018) 521–532. `doi:10.1016/j.renene.2017.08.073`.

[191] J. Tautz-Weinert, S. J. Watson, Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection, Journal of Physics: Conference Series 753 (2016) 072014. `doi:10.1088/1742-6596/753/7/072014`.

[192] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, M. Papaelias, Condition monitoring of wind turbines: Techniques and methods, Renewable Energy 46 (2012) 169–178. `doi:10.1016/j.renene.2012.03.003`.

[193] W. Qiao, D. Lu, A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis - Part II: Signals and Signal Processing Methods, IEEE Transactions on Industrial Electronics 62 (10) (2015) 6546–6557. `doi:10.1109/TIE.2015.2422394`.

[194] K. Fischer, D. Coronado, Condition Monitoring of Wind Turbines : State of the Art , User Experience and Recommendations, (2015), Tech. rep., Fraunhofer Institute for Wind Energy and Energy System Technology, Bremerhaven, Germany, available online, last accessed: 16/04/2018.
URL `https://www.vgb.org/vgbmultimedia/383_Final+report-p-9786.pdf`

[195] GL Renewables Certification, Rules and Guidelines, IV: Industrial Services; 4: Guideline for the certification of condition monitoring systems for wind turbines (2013).

[196] X. Gong, W. Qiao, Bearing Fault Diagnosis for Direct-Drive Wind Turbines via Current-Demodulated Signals, Industrial Electronics, IEEE Transactions on 60 (8) (2013) 3419–3428. `doi:10.1109/TIE.2013.2238871`.

[197] Y. Guo, X. Chen, S. Wang, R. Sun, Z. Zhao, Wind Turbine Diagnosis under Variable Speed Conditions Using a Single Sensor Based on the Synchrosqueezing Transform Method, Sensors 17 (6) (2017) 1149. `doi:10.3390/s17051149`.

[198] B. Geropp, Envelope Analysis - A Signal Analysis Technique for Early Detection and Isolation of Machine Faults, IFAC Proceedings Volumes 30 (18) (1997) 977–981. `doi:10.1016/S1474-6670(17)42527-4`.

[199] B. P. Bogert, J. R. Healy, J. W. Tukey, The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking, in: Proceedings of the Symposium on Time Series Analysis, 1963, pp. 209–243.

[200] M. P. Norton, D. G. Karczub, Fundamentals of Noise and Vibration Analysis for Engineers, Cambridge University Press, 2003.

[201] N. J. Wismer, Gearbox Analysis using Cepstrum Analysis and Comb Liftering, (2003), Tech. Rep. 403, Application Note Brüel & Kjaer, available online, last accessed: 16/04/2018.
URL `https://www.bksv.com/media/doc/bo0440.pdf`

[202] Specra Quest Tech Note, Rotating Machinery Faulty Diagnosis Techniques - Envelope and Cepstrum Analyses, (2006), Tech. rep., Spectra Quest, Inc., Richmond, VA , USA, available online, last accessed: 16/04/2018.
URL `http://spectraquest.com/technote_display/?technote_id=23`

[203] C. J. Crabtree, D. Zappalá, P. J. Tavner, Survey of Commercially Available Condition Monitoring Systems for Wind Turbines, (2014), Tech. rep., Durham University School of Engineering and Computing Sciences and the SUPERGEN Wind Energy Technologies Consortium, available online, last accessed: 16/04/2018.
URL `http://dro.dur.ac.uk/12497/`

[204] S. Koukoura, J. Carroll, A. Mcdonald, Wind Turbine Intelligent Gear Fault Identification, in: Annual Conference of the PHM Society, St. Petersburg, United States, 2017, pp. 1–7.

[205] M. Bach-Andersen, B. Rømer-Odgaard, O. Winther, Deep learning for automated drivetrain fault detection, Wind Energy 21 (1) (2018) 29–41. `doi:10.1002/we.2142`.

[206] J. M. Ha, H. Oh, J. Park, B. D. Youn, Classification of operating conditions of wind turbines for a class-wise condition monitoring strategy, Renewable Energy 103 (2017) 594–605. `doi:10.1016/j.renene.2016.10.071`.

[207] R. Uma Maheswari, R. Umamaheswari, Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train - A contemporary survey, Mechanical Systems and Signal Processing 85 (2017) 296–311. `doi:10.1016/j.ymssp.2016.07.046`.

[208] L. Rokach, O. Maimon, Clustering Methods, in: Data Mining and Knowledge Discovery Handbook, Springer-Verlag, New York, 2005, pp. 321–352.

[209] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing 26 (1) (1978) 43–49. `doi:10.1109/TASSP.1978.1163055`.

[210] J. Nelder, R. Wedderbu, Generalized Linear Models, Journal of the Royal Statistical Society, Series A-General 135 (3) (1972) 370–384. `doi:10.2307/2344614`.

[211] T. K. Ho, Random Decision Forests, in: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 1995, pp. 278–282.

[212] L. Breiman, Random Forests, (2001), Tech. rep., Statistics Department - University of California, Berkeley, CA, available online, last accessed: 02/05/2018.
URL `http://link.springer.com/10.1023/A:1010933404324`

[213] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140. `doi:10.1007/BF00058655`.

[214] J. H. Friedman, Greedy function approximation: A gradient boosting machine., Ann. Statist. 29 (5) (2001) 1189–1232. `doi:10.1214/aos/1013203451`.

[215] D. W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics 11 (2) (1963) 431–441.

[216] Y. Chauvin, D. E. Rumelhart (Eds.), Backpropagation: Theory, Architectures, and Applications, L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1995.

[217] R. Rojas, Neural Networks, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996. `doi:10.1007/978-3-642-61068-4`.

[218] T. D. Gedeon, Data Mining of Inputs: Analysing Magnitude and Functional Measures, International Journal of Neural Systems 08 (02) (1997) 209–218. `doi:10.1142/S0129065797000227`.

[219] N. V. Chawla, Data Mining for Imbalanced Datasets: An Overview, in: Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, 2009, pp. 875–886.

[220] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

Appendices

B Failure Modes for G $<$ 1 MW, G $\geq$ 1 MW, stall and pitch regulated turbines
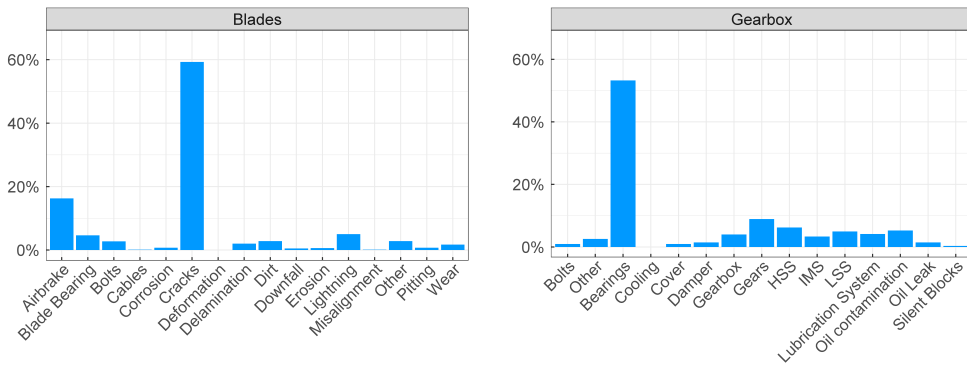
Figure B.1. Failure modes (G $<$ 1 MW turbines) for the blades and gearbox.
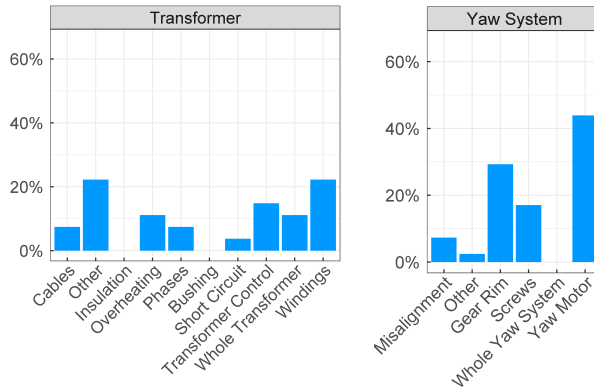
Figure B.2. Failure modes (G < 1 MW turbines) for the transformer and yaw system.
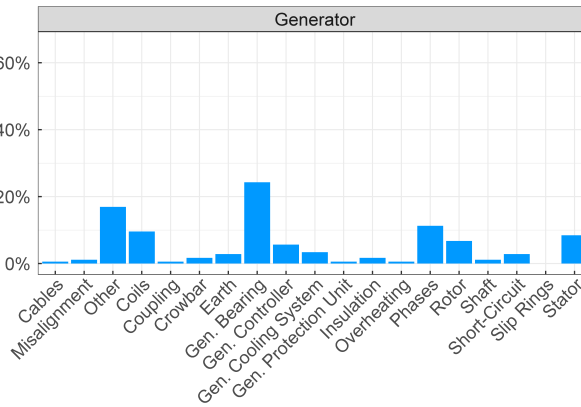
Figure B.3. Failure modes (G < 1 MW turbines) for the generator.
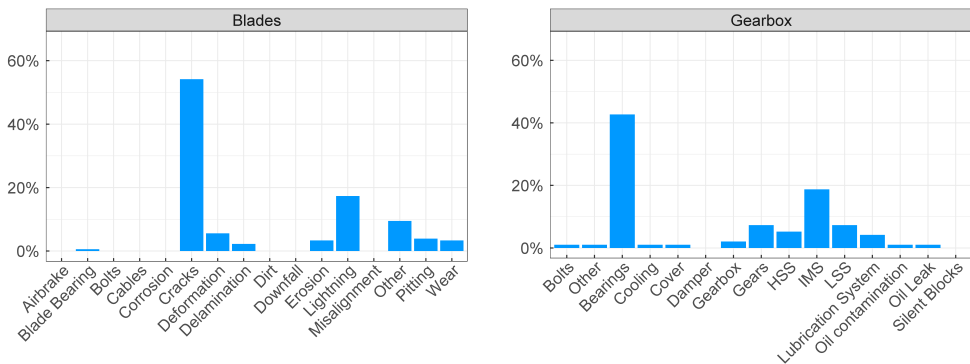
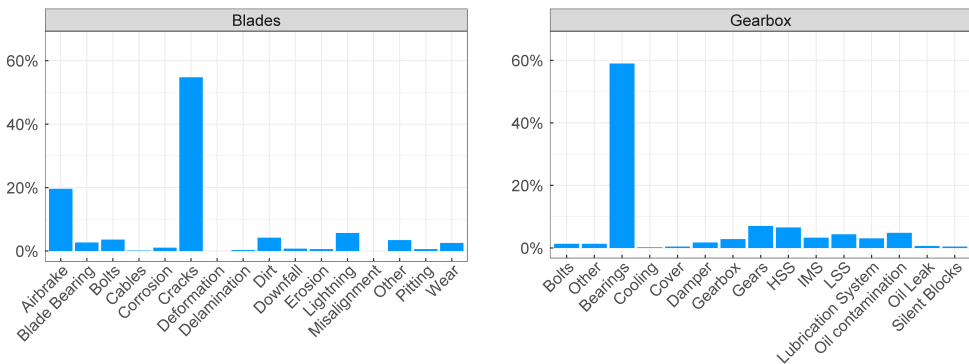

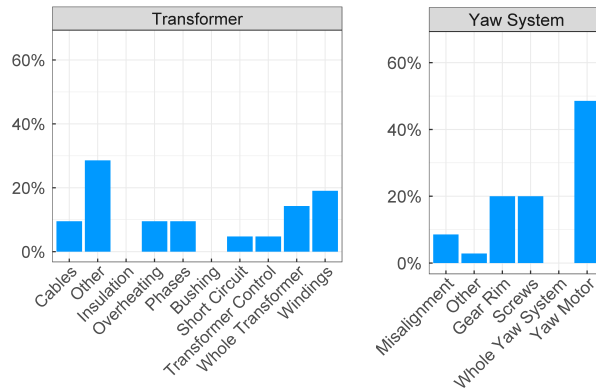**Figure B.4.** Failure modes (G ≥ 1 MW turbines) for the blades and gearbox.
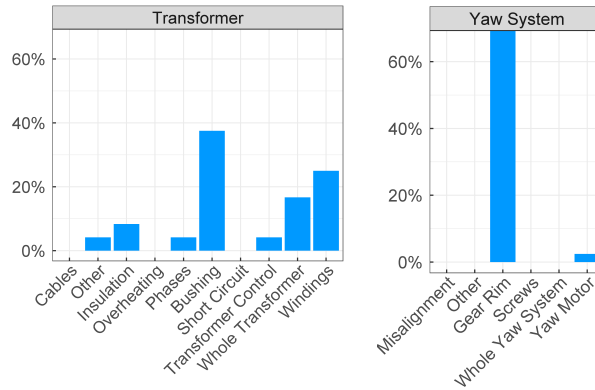
Figure B.5. Failure modes (G ≥ 1 MW turbines) for the transformer and yaw system.
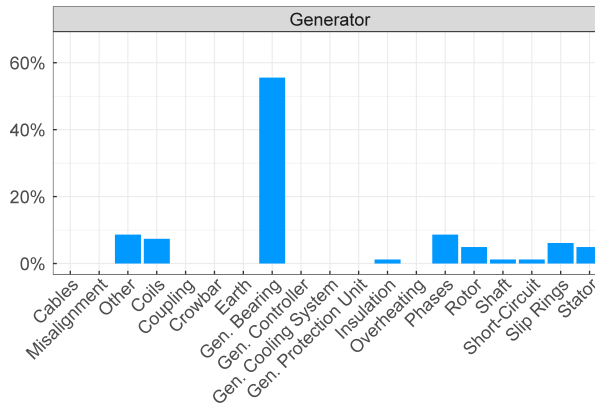
Figure B.6. Failure modes (G ≥ 1 MW turbines) for the generator.

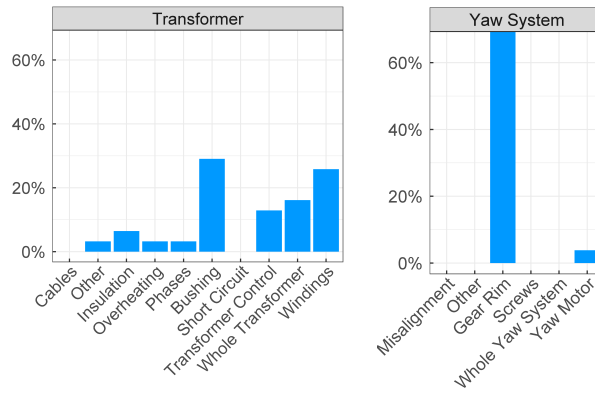**Figure B.7.** Failure modes (stall regulated turbines) for the blades and gearbox.

Figure B.8. Failure modes (stall regulated turbines) for the transformer and yaw system.
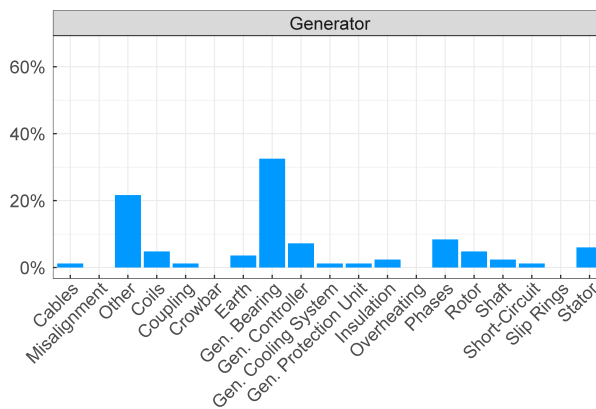
Figure B.9. Failure modes (stall regulated turbines) for the generator.
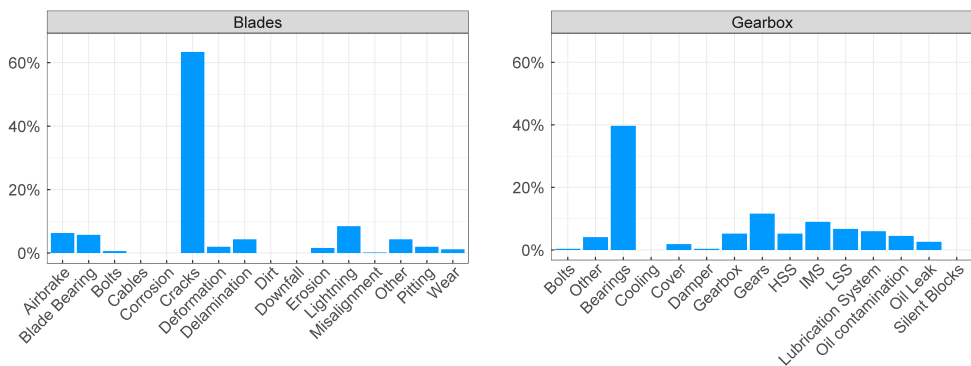
Figure B.10. Failure modes (pitch regulated turbines) for the blades and gearbox.

Figure B.11. Failure modes (pitch regulated turbines) for the transformer and yaw system.
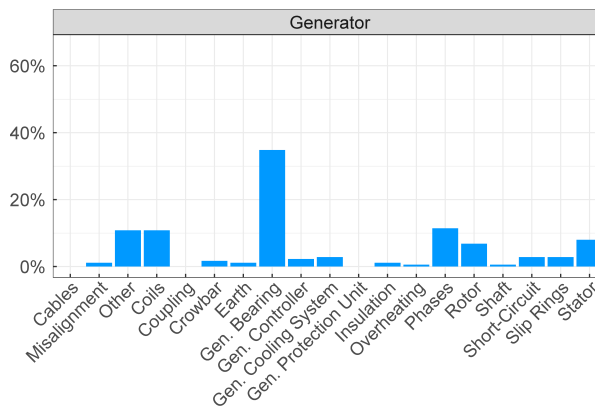

Figure B.12. Failure modes (pitch regulated turbines) for the generator.

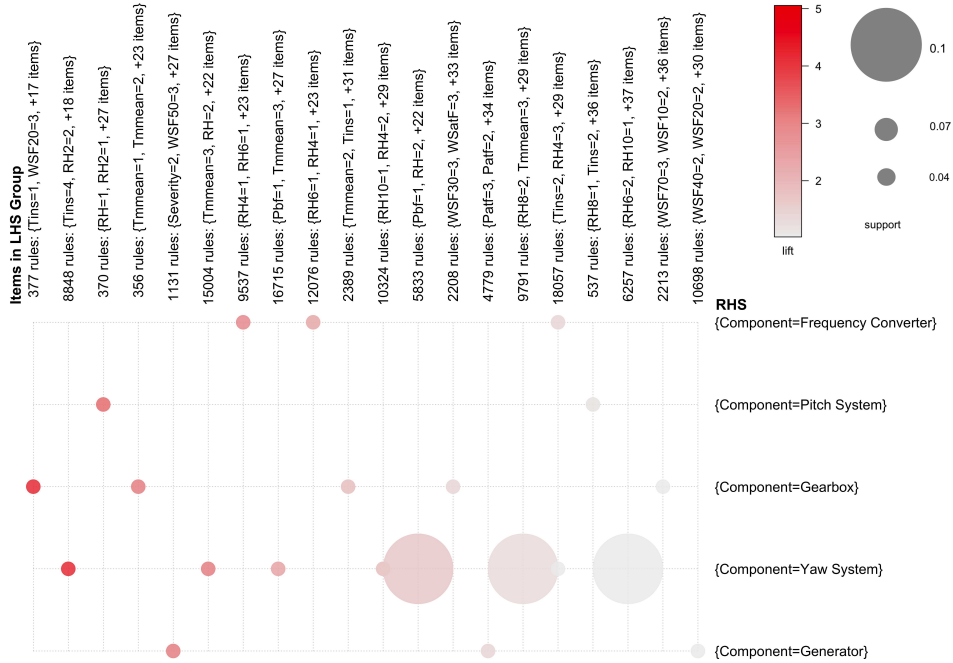# C Rule Matrices for 1-D and Multi-D clustering

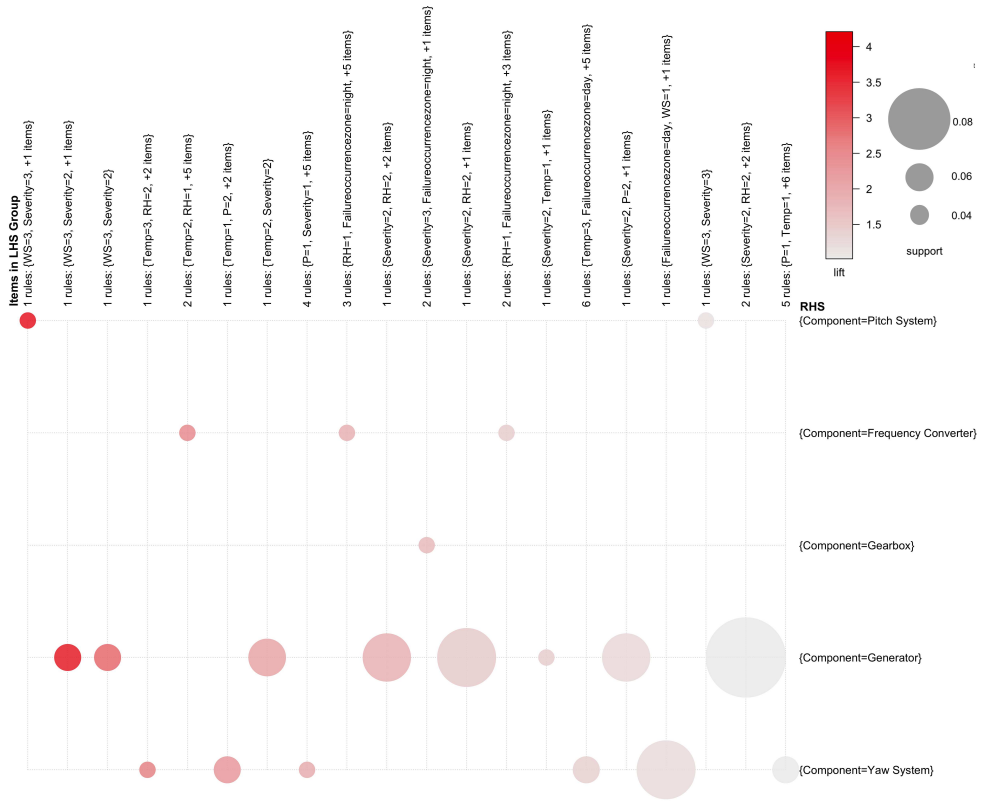**Figure C.13.** Grouped matrix for 1-D rules with a minimum *support* of 0.03.

Figure C.14. Grouped matrix for Multi-D rules with a minimum *support* of 0.03.

D Evaluation Metrics for the Predictions without Discretised Data Input

Table D.1. Results of the predictions using logistic regression without discretised input data.

Component	MCC	ACC	Sensitivity	Specificity	TP	FP	TN	FN
All Failures	0.275	0.639	0.789	0.471	15	9	8	4
Blades	0.666	0.861	0.800	0.885	8	3	23	2
Gearbox	0.172	0.750	0.400	0.806	2	6	25	3
Generator	0.213	0.861	0.250	0.938	1	2	30	3
Main Bearing	1.000	1.000	1.000	1.000	1	0	35	0
Pitch System	NA	NA	NA	NA	NA	NA	NA	NA
Yaw System	0	0.972	0	1.000	0	0	35	1

Table D.2. Results of the predictions using random forests without discretised input data.

Component	MCC	ACC	Sensitivity	Specificity	TP	FP	TN	FN
All Failures	0	0.5277778	1	0.0000000	19	17	0	0
Blades	-0.150	0.667	0	0.923	0	2	24	10
Gearbox	-0.068	0.833	0	0.968	0	1	30	5
Generator	0	0.889	0	1.000	0	0	32	4
Main Bearing	0	0.972	0	1.000	0	0	35	1
Pitch System	0	1.00	NA	1.000	0	0	36	0
Yaw System	0	0.972	0	1.000	0	0	35	1

In this dissertation reliability models and failure detection algorithms are presented that contribute to decreasing the overall cost of modern wind farms by enhancing conventional operation & maintenance through predictive modelling.