Julia Walgern

Field-data based reliability modelling of wind turbine subsystems

A thesis submitted for the degree of Doctor of Engineering

March 2025

Academic supervisor: Prof. Athanasios Kolios

Department of Naval Architecture, Ocean & Marine Engineering

University of Strathclyde, United Kingdom

Wind & Marine Energy Systems & Structures Centre for Doctoral Training

in collaboration with

Nothing in life is to be feared, it is only to be understood.

Marie Curie

Declaration of originality

I declare that I have composed this thesis myself and, except where otherwise noted, the work contained within it is entirely my own. Parts of this thesis have already been published, as noted in the text, but no part has been submitted previously for any other degree or professional qualification.

Julia Walgern

Abstract

Operations and maintenance (O&M) costs account for up to one third of the levelized cost of energy of wind farms. Wind turbine component failures lead to significant repair costs and revenue losses, making the use of operational knowledge crucial for reducing associated costs and risks. However, uncertainty persists due to a lack of quantified reliability of wind turbines and their components, particularly for newer turbine generations. This gap directly impacts O&M decision-making, as simulations and calculations depend on accurate reliability inputs. The increasing size and further evolving technology of wind turbines further complicate projections, which are essential for future wind farm planning and competitive auction bidding.

This thesis presents a series of analyses, which address the research question "How can operations and maintenance data be utilised more efficiently to further reduce costs and risks during wind farm planning as well as operation?" For this, a comprehensive review of existing reliability data, highlighting the shortcomings of previous studies, is provided and advantages and limitations of different reliability assessment methods are investigated. A detailed economic life cycle simulation and assessment framework is developed, integrating a cost-revenue model that accounts for CAPEX, OPEX, and revenue factors, as well as wake and blockage effects for offshore wind farms in the German North and Baltic Seas. A digitalisation workflow is introduced to transform unstructured, non-standardised maintenance reports into machine-readable data classifying components worked on during turbine visits. The feasibility of using text classifiers for preprocessing maintenance reports is evaluated, demonstrating their potential to reduce manual data processing efforts. Furthermore, the impact of classification methods on reliability key performance indicators is analysed.

The thesis utilises a unique dataset of 1335 onshore and offshore turbines with rated capacities of up to 9 MW, which covers maintenance records from 2006 to 2024, offering a highly diverse and recent data resource compared to previous studies. A thorough analysis of failure rates, repair times, and maintenance resource requirements is conducted, providing O&M simulation input for 29 subsystems, covering major component replacements, further corrective maintenance as well as preventive maintenance interventions. Failure behaviour over time for the entire wind turbine system and key subsystems is analysed using Nelson-Aalen plots, while the influence of covariates is assessed with a non-homogeneous Poisson process (NHPP) model. A comprehensive analysis of component failures within the pitch and converter subsystems is conducted comparing electrical and hydraulic pitch systems as well as low-voltage and medium-voltage power converters, respectively. Finally, the thesis compares the developed reliability modelling approaches against a previously published study and the impact of these models on O&M simulations is assessed, highlighting the limitations of average failure rates and the advantages of NHPP regression modelling.

The results indicate that although onshore wind turbines experience lower failure rates per turbine and year, their failure rates per megawatt of rated turbine capacity per year are higher than those of offshore turbines. The pitch, control, and converter subsystems are identified as the most critical with respect to high failure rates. The analysis reveals distinct reliability patterns across wind turbine subsystems over wind turbine operating age. While some subsystems follow a classical bathtub curve, others transition directly from early failures to deterioration, highlighting the need for time-dependent, subsystem-specific reliability modelling rather than assuming uniform failure behaviour. The results of NHPP regression in combination with a covariate selection process confirm that multiple factors significantly influence wind turbine and subsystem reliability. Newer turbine commissioning years generally enhance reliability, reflecting technological and design advancements. However, higher rated turbine capacity negatively impacts reliability, aligning with previous findings that larger turbines experience higher failure intensities. These opposing trends underscore the advantages of NHPP modelling in separating and quantifying individual covariate effects. Additionally, subsystem design choices are found to be a key determinant of reliability.

Keywords: wind turbines, operations and maintenance, reliability analysis, failure rate, corrective and preventive maintenance, maintenance reports, field data, failure data, reliability modelling, Nelson-Aalen plot, non-homogeneous Poisson process, digitalisation

Acknowledgements

The research for this Engineering Doctorate was mostly carried out within the three research projects “Reduction of uncertainties for continued operation of offshore wind farms combining reliability and yield analysis (RUN25+)”, “Digitalization of Maintenance Information (DigMa)”, and “Reliability and Monitoring of Power Converters in Wind Turbines (ReCoWind2)”, funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant numbers 03EE3106, 03EE2016A, and 03EI4066A, respectively. Further financial support was received by EPSRC through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1.

The successful completion of this thesis has been made possible through the contribution of many people who have provided motivation, valuable insights, critical discussions, and unwavering support. Especially, I would like to thank:

My academic supervisor Athanasios Kolios for encouraging me to pursue this degree and for supporting my participation in the Wind and Marine Energy Systems Centre for Doctoral Training. Your passion for wind energy reliability research in combination with pragmatism required in industry has been inspiring. I sincerely appreciate your guidance in navigating the process of journal publications and handling the challenges that come with peer review.

My EngD supervisor Katharina Fischer at Fraunhofer IWES for the scientific guidance, insightful discussions, and exceptional reviews. Your extensive expertise in various O&M-related topics has been invaluable throughout my EngD journey. I am especially grateful for your support as ally in writing research proposals, managing and executing both research and industry-funded projects, and leading our group “Technical Reliability” with dedication.

My co-authors for their collaboration on the research projects DigMa, RUN25+, ReCoWind2, and the BSH project. In particular, I would like to thank my working students for their invaluable support in preprocessing vast amounts of maintenance reports.

The CDT team, including the management, lecturers, and students from all cohorts, for creating an outstanding programme with insightful offshore wind energy modules and fostering a truly inspiring network of exceptional individuals.

All project partners from past and ongoing projects for providing valuable insights into asset management, particularly sharing maintenance data from numerous wind farms. This thesis would not have been possible without your data. I deeply appreciate the trustful collaboration.

My family and friends for your persistent encouragement, your patience in listening, and your endless support throughout this journey.

Table of contents

List of figures

List of tables

List of abbreviations

AEP	Annual energy production
BSH	Federal Maritime & Hydrographic Agency
CAPEX	Capital expenditure
CMS	condition monitoring systems
CNN	Convolutional neural network
CTV	Crew transfer vessel
DFIG	Doubly-fed induction generator
EESG	Electrically excited synchronous generator
EEZ	Exclusive Economic Zone
ELSA	Economic life cycle simulation and assessment (framework)
ERP	Enterprise resource planning
FMECA	Failure modes effects and criticality analysis
FTA	Fault tree analysis
IEA	International Energy Agency
IGBT	Insulated-gate bipolar transistor
IGCT	Integrated gate-commutated thyristors
JUV	Jack-up vessel
KPI	Key performance indicator
LCoE	Levelized cost of energy
LSC	Line-side converter
LV	Low voltage
MCR	Major component replacement
ML	Machine learning
MLE	Maximum likelihood estimation
MSC	Machine-side converter
MTBF	Mean time between failures
MTTF	Mean time to failure
MTTR	Meant time to repair

MV	Medium voltage
MW	Megawatt
NHPP	Non-homogeneous Poisson process
NLP	Natural language processing
O&M	Operations and maintenance
OCR	Optical character recognition
OEM	Original equipment manufacturer
OWF	Offshore wind farm
OPEX	Operational expenditure
PCA	Principal component analysis
PMSG	Permanent magnet synchronous generator
RAM	Reliability, availability, and maintainability
RDS-PP	Reference designation system for power plants
RNA	Rotor-nacelle assembly
SCADA	Supervisory control and data acquisition
SCIG	Squirrel-cage induction generator
SE	Service enterprise
SOV	Service operation vessel
SPARTA	System Performance, Availability and Reliability Trend Analysis
SVM	Support vector machine
TDoMM	Text description of the maintenance measure
TF-IDF	Term frequency-inverse document frequency
WACC	Weighted average cost of capital
WF	Wind farm
WT	Wind turbine
ZEUS	State event cause code

1 Introduction to the thesis

This chapter presents an introduction to the thesis covering background of onshore and offshore wind energy trends (Section 1.1.1) and operations and maintenance (Section 1.1.2), the motivation and challenges (Section 1.2) as well as the aim and objectives for this thesis (Section 1.3). The key scientific publications forming the foundation of this thesis are presented in Section 1.4 and a thesis outline is provided in Section 1.5.

1.1 Background

1.1.1 Onshore and offshore wind energy trends

New installations in the onshore and offshore wind market in 2023 led to passing the milestone of 1000 GW wind energy capacity installed worldwide (see **Figure 1**). While onshore wind energy is with 92.6% making up most of the installed capacity, offshore wind energy gains increasing importance. The global offshore market is projected to grow from 10.8 GW in 2023 to 37.1 GW by 2028, expanding its share of new global installations from the current 9% to 20% by 2028 [1]. In Europe, over 85 GW of offshore wind capacity is anticipated to be developed between 2024 and 2030 [2].

In 2023, wind energy accounted for 19% of the total electricity consumption in the EU-27. This share was considerably higher in several countries, reaching 56% in Denmark, 36% in Ireland, 31% in Germany, 29% in the UK, and 27% in both Spain and the Netherlands [2].

Figure 1. Newly installed and total capacity available worldwide for onshore and offshore wind energy (status 2023) [1]

The first offshore wind farm "Vindeby" was built in Denmark in 1991. Extreme cost reductions have been achieved since that time: While early offshore wind farms still received a fixed, state-financed feed-in tariff, there are currently wind farms under construction that manage completely without subsidies and even paid surcharges tendering for offshore areas. In Germany, for example, the Federal Network Agency awarded bids totalling EUR 12.6 billion for four areas with a total capacity of 7,000 MW in 2023 [3].

This was made possible by the immense technical development of wind turbines, among other things. While the 'Vindeby' wind farm was built with wind turbines with a nominal output of 0.45 MW, the models currently being installed in European waters already have a rated power of 15 MW. The trend of offshore and onshore turbine size from 1980 to 2030 are illustrated in **Figure 2**.

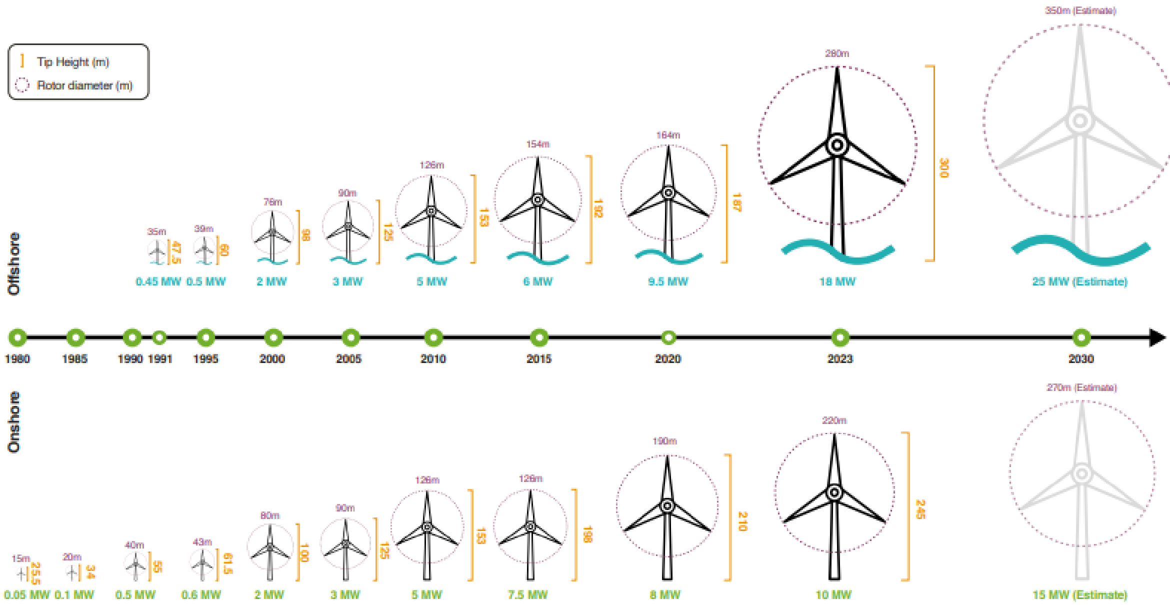

Figure 2. Trend of onshore and offshore turbine size from 1980 to 2030 [1]

Especially projections for future wind turbine sizes were repeatedly wrong. Siemens Gamesa is currently installing its SG DD-276 prototype of 21.5 MW rated capacity. Larger wind turbines deliver a higher energy yield. This means that more electricity can be generated with fewer turbines in a wind farm. Manufacturers are therefore building ever larger rotors. The largest wind turbine currently in use comes from the Chinese state-owned China Railway Rolling Stock Corporation (CRRC). It has a rated capacity of 20 MW and was recently put into operation in China. The pursuit of records is actively underway: the Chinese company Mingyang Smart Energy is building one with a rated power of 22 MW. Chinese manufacturers Goldwind, Shanghai Electric and China State Shipbuilding Corporation (CSSC) announced a wind turbine with 26 MW last autumn, while Dongfang Electric announced one with 26 MW as well [4].

While this turbine growth enables economies of scale for capital expenditure (CAPEX), operational expenditure (OPEX) for such big assets is difficult to predict. Currently, there remains uncertainty during the operations and maintenance (O&M) phase due to the unquantified reliability of wind turbines (WTs) and their components. This directly impacts the modelling of O&M processes, as field-based and WT technology-specific data is lacking, particularly for newer turbine generations. The increasing size of WTs and evolving platform technologies further complicate projections. Nevertheless, these projections are crucial for both, optimising O&M of already existing wind farms (WFs) and future WF planning to successfully prepare bids for upcoming auctions. The described situation provides the context for this thesis. Since OPEX make up for around 25% of levelized cost of energy (LCoE) [5], a better understanding of reliability is an opportunity for further cost and risk reduction.

1.1.2 Operations and maintenance

O&M is by far the longest phase in a wind energy project. While in the past WFs were designed for 20 years of operation, nowadays longer operation is planned for and lifetime extensions of up to 35 years are discussed (compare e.g. [6], [7]). Therefore, the project's economic success is also determined by O&M performance. NREL assumes a net OPEX reduction of 17% and 31% through wind farm economies of scale, turbine scaling, advanced O&M strategies, improved vessel accessibility, and remote maintenance strategies for offshore and onshore wind energy assets, respectively [5]. To achieve these numbers further research in the field of O&M is required. Especially tackling the point of improved O&M strategies, analysing existing WF data to support decision making plays an important role.

Operators of WFs collect different types of data during operation: operational data including turbine and meteorological parameters being organised within the supervisory control and data acquisition (SCADA) system and maintenance data in form of service reports or invoices. While SCADA data is more easily accessible and therefore often used in research [8], maintenance data is highly confidential and seldomly shared. This results in few publications which present real-world failure data (e.g. [9], [10], [11], [12], [13]).

Many O&M simulation tools have been developed to analyse and improve WFs' O&M (e.g. [14], [15], [16], [17], [18], [19]). Those come with two limitations: On the one hand, each tool is only applicable for scenarios which are already known and implemented. Each time new logistic concepts or O&M strategies are developed, an update of those tools is required. On the other hand, the simulation output is heavily dependent on the input data. As abovementioned failure data is typically quite old, the simulation results need to be interpreted with care.

1.2 Motivation and challenges

The main motivation for this thesis stems from insufficient and outdated publicly available failure data for OPEX modelling. IEA Wind Task 33 found widespread industry recognition of the importance of collecting and analysing reliability data to optimise both profit margins and LCoE [20]. The initiative draws the conclusion that the absence of standards for reliability data is hindering industry progress in addressing reliability challenges.

In the past, the main focus of WF developers was on the turbine supply agreement including logistics of installation and commissioning. A lot of effort was invested in synchronising the commissioning date of the turbines with grid availability to earn money with fed-in electricity from day one of operation. As most WF operators have concluded a full maintenance contract with the original equipment manufacturer (OEM) and therefore availability and warranties were contractually agreed on, reliability data was of minor interest during contract negotiations. However, more and more owners and operators of WFs realise that this information is highly relevant to become independent of the OEMs. While larger utilities start to maintain WFs with their own service teams, also smaller operators want to better understand their asset reliability and reasons for downtime. This is typically hindered by the lack of available and standardised reliability data.

Clifton et al. [21] elaborate that the digitalisation of the wind energy sector provides enhanced reliability, cost savings, new business models, and more cost-effective integration of wind energy as an energy source. The authors conclude, however, that digitalisation also presents three major challenges that must be addressed:

1. Data – creating FAIR data frameworks
2. Culture – connecting people and data to foster innovation
3. Coopetition – enabling collaboration and competition between organisations

This thesis will contribute to the three identified challenges by [21]. Specifically, it aims to provide reliability data to the public focusing on the FAIR principle, which stands for findable, accessible, interoperable and reusable. The aspects are addressed by publishing results in well-known open access wind energy journals, utilising existing standards and guidelines for data preprocessing, describing meta data, and discussing uncertainties and limitations. Consequently, people and data will be connected which supports innovation in the research community and coopetition is strengthened as industry stakeholders can benchmark reliability key performance indicators (KPIs) against their peers. Utilising operational knowledge of a large and diverse WT fleet will foster reliability improvements and optimisation of WFs' O&M strategies.

1.3 Aim and objectives

Based on the forementioned background (Section 1.1) as well as motivation and challenges (Section 1.2) regarding optimising O&M of onshore and especially offshore wind farms, this thesis aims to derive novel field-data based reliability models of wind turbine subsystems and components to support decisions for design optimisation, O&M strategies, and lifetime extension initiatives.

To achieve this overall aim, the following objectives are defined:

- Investigate and classify existing reliability figures for onshore and offshore wind turbines and derive suitable reliability assessment methods and metrics
- Develop a framework for economic feasibility studies of offshore wind farms considering relevant input parameters to quantify their impact on output
- Assess the challenges of deriving reliability metrics and evaluate the impact of differently applied methods for preprocessing and digitalisation of maintenance reports
- Develop reliability models of wind turbine subsystems and selected components based on real-world O&M data
- Evaluate the impact of using different reliability inputs for O&M simulations by comparing the developed reliability models with previous published ones

1.4 Scientific publications and collaborations

The thesis is composed of a portfolio of research works that have been published in peer-reviewed conference proceedings and scientific journals. The main publications which are underlying this thesis are listed below in chronological order:

- Marc-Alexander Lutz, Julia Walgern, Katharina Beckh, Juliane Schneider, Stefan Faulstich, Sebastian Pfaffel, 2022. “Digitalization Workflow for Automated Structuring and Standardization of Maintenance Information of Wind Turbines into Domain Standard as a Basis for Reliability KPI Calculation”. IOP Journal of Physics Conference Series (WindEurope Annual Event 2022), doi: 10.1088/1742-6596/2257/1/012004.
- Julia Walgern, Katharina Fischer, Paul Hentschel, Athanasios Kolios, 2023. “Reliability of electrical and hydraulic pitch systems in wind turbines based on field-data analysis”. Energy Reports, 9, 3273-3281, doi: 10.1016/j.egyr.2023.02.007.
- Julia Walgern, David Baumgärtner, Johannes Fricke, Niklas Requate, Athanasios Kolios, Martin Dörenkämper, Tobias Meyer, Lukas Vollmer, 2023. “Economic feasibility study for continued operation of German offshore wind farms”. IOP Journal of Physics Conference Series (EERA DeepWind conference 2023), doi: 10.1088/1742-6596/2626/1/012031.
- Julia Walgern, Katharina Beckh, Neele Hannes, Martin Horn, Marc-Alexander Lutz, Katharina Fischer, Athanasios Kolios, 2024. “Impact of using text classifiers for standardising maintenance data of wind turbines on reliability calculations”. IET Renewable Power Generation, 18(15), 3463-3479, https://doi.org/10.1049/rpg2.13151
- Julia Walgern, Nils Stratmann, Martin Horn, Nathalene Then Wei Ying, Moritz Menzel, Fraser Anderson, Athanasios Kolios, Katharina Fischer, 2025. “Reliability and O&M key performance indicators of onshore and offshore wind turbines based on field-data analysis”, submitted to Wind Energy for publication.
- Katharina Fischer, Fraser Anderson, Julia Walgern, 2025. „Medium-Voltage versus Low-Voltage Converter Reliability in Wind Turbines: A Field-Data Based Study” PCIM2025.
- Julia Walgern, Fraser Anderson, Athanasios Kolios, Katharina Fischer, 2025. “Field-data based wind turbine reliability modelling: Quantifying effects of operating age, design and technological development”, submitted to Wind Energy for publication.

At the beginning of each chapter, it is indicated on which publication the chapter is based.

In addition, parts of the research work were presented at different conferences. Conference contributions and further involvement in publications and collaborative works during the research phase for the EngD degree which were not directly used for this thesis are listed in Appendix A.

1.5 Thesis outline

The rest of this thesis is structured as follows:

Chapter 2 presents a literature review on O&M modelling (Section 2.1) as well as reliability modelling and existing input figures for O&M modelling (Section 2.2). Both parts are summarised, and research gaps are identified (Section 2.3). This chapter addresses the first part of objective 1 of the thesis.

Chapter 3 presents the methodology approach followed within the thesis (Section 3.1) and introduces the dataset utilised for the overall thesis (Section 3.2). The chapter covers the second part of objective 1 of the thesis.

Chapter 4 introduces an economic life cycle simulation and assessment framework called ELSA. It is applied to a cumulative scenario that categorises all existing offshore wind farms in Germany based on size and key dimensions, in order to inform decisions regarding the business case for continued operation beyond their nominal service life. This chapter supports objective 2 of the thesis.

Chapter 5 addresses the fact that due to the lack of structured and standardised data, maintenance data is often underutilised or requires significant manual effort. To tackle this, a digitalisation workflow is proposed and applied to real-world wind turbine service reports and invoices to streamline preprocessing of maintenance data.

Chapter 6 examines the effectiveness of text classifiers in automating categorisation compared to manual labelling of maintenance reports. Discrepancies in failure rate KPIs from manual versus classifier-processed data and resulting uncertainties of both methods are analysed. Both Chapters 5 and 6 cover objective 3 of the thesis.

Chapter 7 presents average failure rates of all subsystems of a wind turbine and compares reliability figures of onshore and offshore wind turbines. Additionally, corrective and preventive maintenance interventions, major component replacements, average repair times, and average number of required technicians for different maintenance interventions are evaluated.

Chapter 8 seeks to determine the failure rates of two pitch system concepts – electrical and hydraulic – through statistical analysis of a large sample of onshore assets. The findings are classified based on turbine rating, seasonal effects, and the reliability performance of different manufacturers.

Chapter 9 digs deeper into a study comparing medium-voltage and low-voltage power converter reliability making use of failure rates, visual analysis of seasonal patterns, and Nelson-Aalen plots to characterise the failure behaviour through time.

Chapter 10 presents more detailed reliability analyses which explore temporal trends in failure behaviour and the effect of potentially influential factors (or so called "covariates") on reliability. Nelson-Aalen plots are utilised to establish trends in failure behaviour through time and the non-homogenous Poisson process (NHPP) quantifies the effect of relevant covariates. Chapters 7, 8, 9, and 10 all address objective 4 of the thesis.

Chapter 11 evaluates the impact of different reliability modelling approaches as input for O&M simulations. The two models developed within this thesis – average failure rates per MW per year and advanced multivariate reliability models – are compared to the most widely used reliability model to date. The advantages and limitations of each approach are outlined. Chapter 11, along with Section 10.3.3, addresses objective 5 of this thesis.

Chapter 12 summarises all chapters of the thesis (Section 12.1) and discusses the contribution of the thesis to knowledge, research, and industry (Section 12.2). The thesis concludes with future work and an outlook (Section 12.3) and provides concluding remarks (Section 12.4).

2 Literature review

This chapter presents a general literature review dedicated to O&M strategies and modelling (Section 2.1) and published input figures for O&M modelling focusing especially on reliability data (Section 2.2). A summary and identification of research gaps is provided (Section 2.3). More tailored literature reviews for specific topics can also be found in the respective sections of the following chapters (see Sections 4.1, 5.2, ,6.2, 7.2, and 8.1).

2.1 O&M strategies and modelling

WFs are operated using two main maintenance strategies: preventive maintenance and corrective maintenance. Preventive maintenance is further divided into time-based (scheduled) and condition-based maintenance, the former typically conducted annually in summer to prevent failures. Corrective maintenance is performed when failures occur, which can lead to long downtimes depending on timing and conditions. To reduce corrective maintenance and associated downtimes, operators employ predictive maintenance by analysing data from SCADA systems or condition monitoring systems (CMS). CMS offer more detailed signals acquired by means of additional sensors but come with additional costs, making their economic feasibility dependent on balancing early fault detection against investment and monitoring-service costs.

During the operational phase of a WF, the choice of a maintenance strategy is primarily driven by the trade-off between lost revenue and increasing maintenance costs. Lost revenue is directly linked to the WF's availability or to be precise its downtime. Downtime is driven by both, the actual repair time and the waiting period before maintenance can commence. The latter varies considerably between onshore and offshore maintenance [22]. For offshore wind farms (OWFs), both the preparation time – including staff, vessel and spare parts allocation – and the waiting time for favourable weather conditions can be significantly longer. Consequently, a comprehensive O&M strategy is essential, encompassing O&M facilities and ports, vessels and equipment, maintenance schedules and methodologies, replacement of critical components, and the supply of consumable such as oil, grease and filters. Additionally, effective monitoring of turbine blades, foundations, scour protection, and cables are critical components of a robust O&M framework.

The optimisation of O&M strategies to minimise downtime and reduce costs is a highly relevant research topic. Existing literature explores various approaches of quantitative methods typically involving either mathematical optimisation or event-based stochastic approaches. Several simulation models have been developed to evaluate different aspects of O&M or O&M strategies as a whole (e.g. [14], [15], [16], [17], [18], [19]). Bendlin et al. review and classify various O&M tools based on their functions, including failure analysis, weather assessment, routing, scheduling, and economic modelling [23]. Commonly employed approaches include Monte Carlo methods, Weibull distributions, Markov chains, and Poisson processes. Next to academic tools, Shoreline provides a commercial simulation tool which is widely recognised within the industry [24]. All these models and corresponding studies depend on reliable input data. A review of input figures is presented in the next section.

2.2 Input figures for O&M modelling

Next to power curves, weather data, vessel characteristics and hourly rates, failure data is decisive input for simulating O&M activities. While the former are typically available in the public domain, failure data in form of failure rates or more advanced reliability models are rarely published. This is due to strict confidentiality of maintenance data, the difficult question of data ownership between operators and OEMs and the fear of OEMs and operators to lose competitive advantage when sharing this information. The more data an organisation can analyse, the better assumptions can be derived for the development of future wind farms which can be decisive in competitive auctions. Nonetheless, a few studies about WT reliability have been published in the past:

Tavner et al. and Spinato et al. analysed data of 6000 onshore WTs from Denmark and Germany recorded for 11 years before 2008 as part of the DOWEC project ([25], [9]). This data stems from two different initiatives, namely the Windstats survey in Denmark and Germany (often referred to as WSDK and WSD) [26] and a survey conducted by the Landwirtschaftskammer (LWK) in Schleswig Holstein in Germany [27]. Even though these studies are based on data from 1994 till 2008, due to the amount of collected data and the fact that failure rates per turbine and year as well as hours lost per failure are presented, these pioneering figures are still used regularly in the field of reliability research.

The ReliaWind project assessed data of around 350 “modern turbines” being defined as WTs with a rated capacity >850 kW back in 2011 [10]. In contrast to the DOWEC project, ReliaWind published failure data in form of normalised failure rates and normalised hours lost for different subsystems. That allowed to understand which subsystems were failing most often but no exact numbers were provided, which could have been utilised for O&M simulations.

Faulstich et al. published results of the Scientific Measurement and Evaluation Programme (WMEP) which covers data of 1500 onshore WTs collected from 1989 to 2006 [11]. The authors presented annual failure rates per turbine and year differentiated in minor and major failures accompanied by mean annual downtime per failure category.

Carroll et al. were the ones presenting offshore failure rates for the first time [12]. For this study around 350 WTs of one OEM with a range of 2 to 4 MW rated capacity covering 1769 WT years of operation were utilised. Next to failure rates per turbine and year, also repair times and spare part costs were published, which has made this publication the to date most prominent input for O&M simulations.

Reder et al. presented a reliability analysis based on 4300 onshore WTs with rated capacities between 300 kW and 3 MW. They introduced an own taxonomy to cluster components and displayed normalised failure rates and downtimes for WTs smaller than and above 1 MW as well as for direct drive turbines [28].

The SPARTA initiative being sponsored by the Crown Estate and the Offshore Renewable Energy (ORE) Catapult is collecting performance and reliability KPIs of offshore assets and publishes reports with aggregated figures on a regular basis. Not each report is covering failure data. The Portfolio Review 2016 presented monthly repair rates per turbine [29] and also the last published review of 2020/21 showed average monthly component failures per turbine and per MW [13]. The studies are based on data of 1378 and 1505 WTs, respectively.

Anderson et al. present failure rates per turbine and year for different failure definitions based on data of one offshore wind farm covering roughly 600 turbine operational years [30].

Further detailed reviews of published failure rate statistics are provided by Pfaffel et al. [31] and Cevasco et al [32].

Next to these relatively famous but older studies, there are also more recent studies available presenting failure data for specific countries. Artigao et al. present normalised failure rates and downtime for onshore WTs in Spain [33]. The study is based on a dataset covering 75 WTs over a period of 11 years. Even though published in 2019, the data itself was recorded as early as 2001-2011. Sarma et al. analyse data of one Turkish onshore WF. The study is based solely on SCADA data recorded from 2017 till 2019. Therefore, only distributions of downtime are evaluated [34]. Moreover, a few publications focusing on specific subsystems are available (e.g. [35], [36]) which support root cause analysis but are of limited help when looking for O&M simulation tool input.

2.3 Summary and identified research gaps

While many O&M simulation tools are available – commercial and academic-driven ones – there is constant need to develop those further to incorporate newest technologies, strategies and research findings. All tools have in common that they rely heavily on available input figures. The output of each tool can only be as good as the utilised input is.

Different reviews (e.g. [31], [32], [33], [34]) conclude that roughly 20 different initiatives were observed in the past publishing reliability statistics. The Windstats survey stands out due to the number of turbines included but is based on turbine technology which was installed before 2008. The reliability study of Carroll et al. captivates with a multitude of details which are helpful for O&M modelling but is limited to small offshore WTs of one OEM. To the author's best knowledge SPARTA is the only initiative still running and which includes newer turbine technologies. However, the last publication is from 2022 and figures of only few selected subsystems are presented. Therefore, all currently available studies are either old and thus based on outdated turbine technologies or provide limited possibilities to utilise the figures for O&M simulations as only normalised failure rates are provided to the public, not all subsystems of a turbine are covered, or the underlying dataset is too small and specific for generalising it to new use cases. Additionally, a comparison in magnitude of failure rates is difficult as no standardised taxonomy or failure rate definition is utilised [30]. Furthermore, most failure rates are presented per turbine and year making it difficult to apply those to WTs with higher rated capacities. While average failure rates in general are easy to interpret and utilise for O&M modelling, they come at the cost of detail. Reliability differs with age of assets (e.g. [11]) – normally represented by a bathtub curve [37] – and can be affected by different designs (e.g. [28]) or operation (e.g. [38], [39]). All these aspects are not covered in simple average failure rates.

Therefore, this thesis aims for collecting and analysing failure data of a diverse and recent WT fleet with sufficient size to generalise findings. Moreover, failure rates will be presented per WT and year, but also per WT rated capacity (in MW) and year to ease the process of extrapolation for future WT generations. Next to providing failure rates, more advanced reliability models will be developed to address (i) temporal trends in failure behaviour and (ii) the effect of potentially influential factors on failure behaviour. To assure applicability and interpretability common standards will be applied for data preprocessing. Additionally, a digitalisation and classification workflow will be introduced to speed up the process between recording data and publishing failure data as a lengthy labelling process hinders providing statistics of up-to-date turbine technology.

3 Methodology approach and utilised datasets

This chapter presents the methodology approach for reliability modelling (Section 3.1) and utilised datasets for this thesis (Section 3.2). The dataset described within this chapter provides an overview of the overall dataset used for the thesis and subsets for specific research are further described in respective chapters (see Sections 5.3.2, 6.3.1, 7.3.2, 8.2.2, 9.2, 0). Similarly, the methodology outlined in this chapter covers only the generic approach utilised for this thesis and further details are provided in the following chapters (see Sections 4.2, 5.3, 6.3.2, 7.3.1, 8.2.1, 9.3, 10.2.1).

3.1 Methodology approach for reliability modelling

In order to understand what the real problems of WF operators regarding reliability are and to provide as realistic reliability KPIs as possible, a field-data based approach is followed within this thesis. The term “field data” describes O&M data directly recorded within or documented for a WF. This can be either operational data in form of time-series data or maintenance reports. In this thesis, the focus is on the latter. From maintenance reports, it can be derived when and what has happened on a WT and which measures have been taken, which is important and necessary information for reliability modelling.

Figure 3 displays the workflow which is utilised for this thesis. First, the data collection needs to be initiated and organised. In general, specific projects - either publicly or industry-funded ones - are required to establish a cooperation with WF operators who can provide failure data. As failure data is highly confidential, a trustworthy relationship needs to be built first. As soon as both parties are aligned which data sources of which WF shall be shared, typically a platform is provided to the operator to upload data. Within this thesis project, data of different operators and WFs was collected which is further described in Section 3.2.

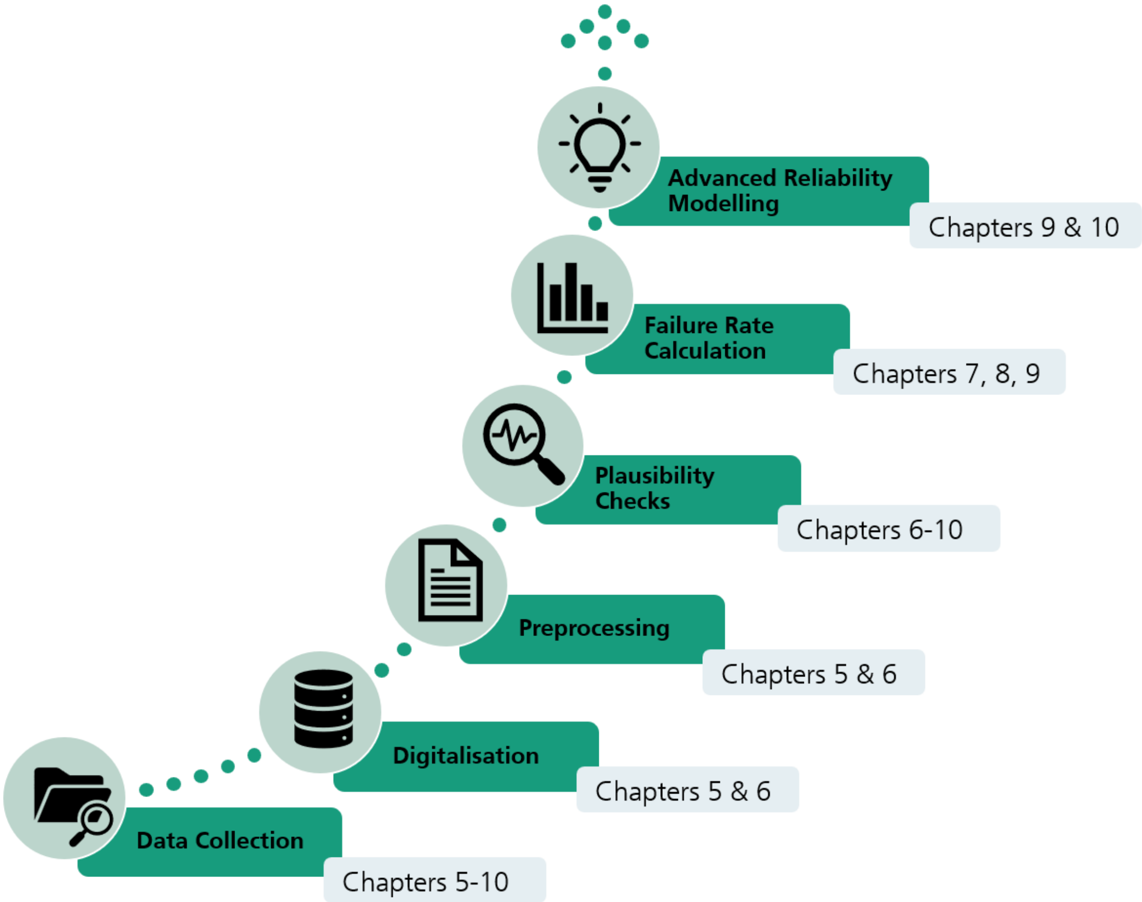

Figure 3. Workflow of the methodological approach

Maintenance or failure data comes typically in the form of Excel sheets or PDFs. Therefore, a second step is to digitalise especially content of PDFs and bring all different data sources in one common format. Information such as work order description, spare parts, affected WT, the date, time and duration of the maintenance action, number of technicians involved, and follow-up tasks / recommendations are transferred to a standardised template which can be fed into the reliability data base. Details about digitalisation of maintenance reports can be found in Chapter 5.

Third, preprocessing of data is a huge task. The main aim is to convert typically free text of work order descriptions, recommendations and spare parts in machine readable and standardised codes which can be utilised for data analyses. Within this thesis, two standards are applied for labelling the datasets to guarantee interoperable and reusable results: On the one hand, the reference designation system for power plants RDS-PP (Application Guideline Part 32: Wind Power Plants) [40] is utilised. It allows to clearly identify components independent of turbine type and technology as well as terminology used by technicians. RDS-PP codes are hierarchically organised, which supports clustering components in different subsystems, and follow the general rule “from large to small”. Therefore, codes become longer the more information of the component is available. On the other hand, it needs to be identified and labelled if the maintenance activity is corrective or preventive and which maintenance action has been taken. For this, the State-Event-Cause-Code ZEUS [41] is applied. It is also organised in a hierarchical way so that different levels of detail describing the maintenance action can be captured. Both standards make data and corresponding analysis results of different assets comparable. For further information regarding the preprocessing see also Chapters 5 and 6.

Fourth, plausibility checks are required after the preprocessing is finalised. Especially completeness and correctness of the dataset is evaluated. The following questions are typically guiding the assessment:

- Is data for each WT of the WF available?
- Are there data gaps between first and last service report? Can these be explained?
- Are all subsystems / component categories covered by the dataset?
- Does the dataset contain all severity levels of maintenance actions?
- Are some reports saved twice / referring to the same maintenance intervention?
- Does the recorded duration of a WT visit make sense?
- Are information like time and coordinates in a common format (summer/wintertime; WGS vs. UTM)?

In some cases, queries must be clarified with the data provider before starting the data analysis itself.

Fifth, failure rate calculations are a simple approach to obtain a first overview of the dataset as part of the reliability analysis (see Section 7.3.1.2 for utilised equations, etc.). Next to calculating failure rates per WT and year as done in most available publications, within this thesis failure rates per WT rated power (in MW) and year are assessed. This facilitates the comparison of different assets with WTs of different sizes as well as inter- and extrapolating reliability figures. Additionally, failure rates per month to examine seasonal patterns or failure rates per year of operation to evaluate changing failure behaviour over time can be calculated. Moreover, comparing failure rates of different data subsets can reveal influences of design parameters.

Failure rate analysis is a straightforward and easily interpretable method; however, it primarily relies on comparative evaluation of results and visual analysis for drawing conclusions. Therefore, the last step of the workflow covers methods of advanced multivariate reliability modelling, which enable separating and quantifying effects of e.g. design, technological development, operating age and environmental and load conditions on WT reliability.

An important aspect of analysis is the examination of failure patterns over time. While in reliability modelling often a Weibull analysis is referred to, within this thesis a non-homogeneous Poisson process (NHPP) regression is utilised. The primary reason for this limitation is that Weibull analysis is only appropriate for evaluating non-repairable systems or components. Such analysis requires knowledge of the component's age at the time of failure, which is typically unknown beyond the first replacement. It is also unknown in cases when the maintenance history available for analysis starts some time after the wind turbine's commissioning, i.e. in case of left-censored data. In contrast, NHPP regression models are well-suited for analysing repairable systems, where failures are treated as recurrent events followed by system repairs. This method effectively handles left-censored data as knowing the WT's operating age at the time of failure is sufficient to draw conclusions about failure intensity over time.

In reliability analyses of repairable systems, failure intensity varies over time, typically following the characteristic pattern of a bathtub curve [37]. **Figure 4** illustrates the three distinct phases of reliability trends, which collectively form the characteristic shape of the bathtub curve of repairable systems:

- Early failures, which are characterised by a decreasing failure rate (δ<1)
- Constant failures, which are described by a constant failure rate (δ=1)
- Deterioration failures, which are defined by an increasing failure rate (δ>1)

The parameter δ is referred to as the shape parameter and is derived from the power-law process governing the failure intensity λ_0 [42]:

$$\lambda_0 = \left(\frac{\delta}{\upsilon}\right)\left(\frac{t}{\upsilon}\right)^{\delta-1} \tag{3.1}$$

Herein, δ>0 determines the phase of the bathtub curve, while υ>0 represents the scale parameter.

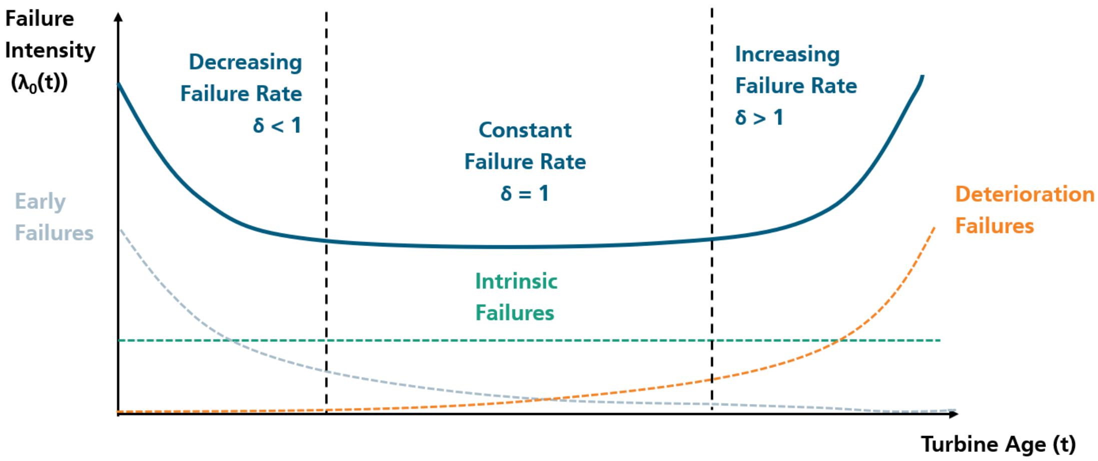

Figure 4. Bathtub curve for repairable systems describing failure intensity over time [43]

Trends in failure behaviour through time are assessed by utilising the Nelson-Aalen estimator, which is a non-parametric estimator of the cumulative intensity function Λ_0 in case of censored or incomplete data [42]:

$$\Lambda_0 = \sum_{t_i \leq t} \frac{d_i}{n_i} \tag{3.2}$$

Herein, d_i is the number of failure events at time t_i and n_i is the total number of turbines at risk at t_i. The log-log plot of the cumulative intensity function versus time – referred to as Nelson-Aalen plot – is utilised to identify the shape parameter δ in a power-law process. The value of δ corresponds to the slope of the resulting line (cf. **Figure 5**).

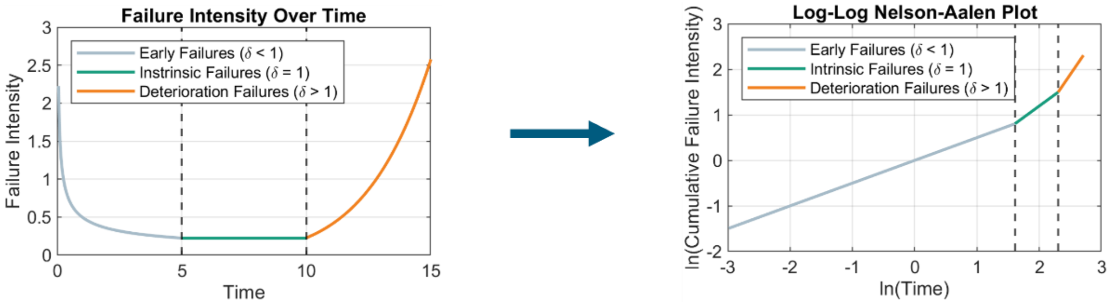

Figure 5. Derivation of the shape parameter δ utilising Nelson-Aalen plots [44]

Utilisation of the Nelson-Aalen estimator has the following advantages:

- Non-parametric: In comparison to parametric methods (e.g. Weibull, Exponential) the Nelson-Aalen estimator does not assume any specific distribution for failure time stamps, making it flexible for real-world datasets for which failure distributions are unknown or complex.
- Censored data: Can incorporate left-censored (e.g. data is not available from commissioning onwards) or right-censored data (e.g. turbines that have not failed yet) without biasing the cumulative intensity function.
- Interpretation: Provides a stepwise estimate of the cumulative intensity function over time, which is intuitive and easy to visualise via Nelson-Aalen plot.
- Exploratory analysis: Allows for quick comparison of cumulative failure intensity across subsystems, turbine types, or operational conditions before moving to regression models.

Another important aspect of failure analysis is the identification of relevant influences on reliability, so called covariates. For this, a covariate vector x is included in the equation describing the failure intensity of the NHPP [45]:

$$\lambda(t) = z\, \lambda_0(t) \exp\left(\beta_1 x_1 + \cdots + \beta_n x_n\right) \tag{3.3}$$

Next to the baseline failure intensity $\lambda_0(t)$ and the factors influencing reliability, z is accounting for heterogeneity that cannot be explained by the set of observable covariates. The parameters β_i, δ, υ, and z are estimated using the maximum likelihood method. For the covariate selection procedure, likelihood ratio statistics and a subsampling procedure are utilised and correlated covariates are handled by principle component analysis (PCA) as described in [46] and [47].

In order to analyse reliability of current technologies being present in WFs, the described workflow needs to be repeated regularly so that always recent data is included in the results and respective conclusions can be drawn.

3.2 Dataset description

The largest part of the dataset utilised for this thesis has been collected and preprocessed during the last five years and is continuously growing. It comprises in total 1335 onshore and offshore wind turbines covering 5539 WT operational years. Key characteristics of the dataset are visualised in **Figure 6** and presented in **Table 1**.

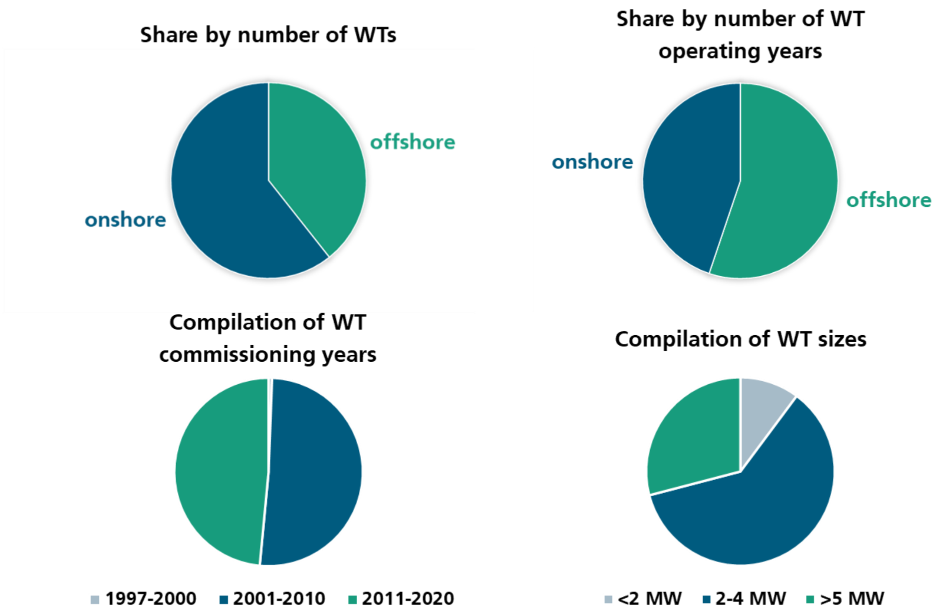

Figure 6. Key characteristics of the dataset

The distribution of onshore and offshore data is presented both in terms of the number of WTs and total WT operating years. While the dataset includes a greater number of onshore WTs, it encompasses more offshore WT operating years. The turbines in this dataset were commissioned between 1997 and 2020, with the majority representing modern turbine technology. This is reflected in the distribution of WT sizes: Only 10% of the WTs have a rated capacity below 2 MW, most fall within the 2-4 MW range, and 30% exceed 5 MW. Since the dataset is derived from maintenance reports, only WT generations that have been in operation for at least two years are included, limiting the maximum rated capacity to 9 MW. The evaluated failure data spans from 2006 to 2024. The offshore subset includes turbines from four different OEMs, while the onshore subset covers nine OEMs. The dataset originates from wind farms across seven different countries.

Table 1. Key characteristics of the dataset

	Offshore	Onshore
WT operational years considered	3056	2483
Number of WT OEMs covered	4	9
Rated capacity considered	Up to 9 MW	
Available data period	2006-2024	
Number of countries covered	7	

Compared to datasets used in previous reliability studies (summarised in Section 2.2), this dataset provides unparalleled diversity, scale, and recency. Various data subsets are employed in each chapter, corresponding to different studies. Detailed descriptions of the respective dataset utilised are provided in the subsequent chapters.

4 Economic life cycle simulation and assessment framework: case study of the economic feasibility for continued operation of German offshore wind farms

With a large number of wind farms already deployed in German waters and aiming to achieve a minimum of 70 GW of offshore wind capacity by 2045, investigating the potential for extended operation of existing assets is an important task. This chapter documents the development of a life cycle cost/revenue framework capable of incorporating CAPEX and OPEX related elements, revenue factors, and deployment location specific aspects, in order to support decisions on the business case for continued operation beyond the nominal service life. The framework called ELSA (Economic Life cycle Simulation and Assessment) is developed and applied to a cumulative scenario, which classifies all existing offshore wind farms in Germany with respect to size and key dimensions. Outcomes of the analysis support the case for extended operation, while highlighting the importance of wake effects to AEP, the magnitude and variability of O&M costs and finally the influence of CAPEX and financial modelling. The material of this chapter has been peer reviewed and published in [1].

4.1 Introduction

The German government plans for accelerated expansion of offshore wind energy. In order to meet set climate objectives, Germany is aiming for a rated capacity of offshore wind turbines (WTs) of at least 30 GW by 2030 and at least 40 GW by 2035. In 2045, a minimum of 70 GW of generation are targeted [48]. To achieve those targets, next to building new offshore wind farms (OWFs) it will be crucial to keep existing OWFs in operation beyond their design lifetime of typically 25 years. This will bridge the time required for strengthening existing supply chains for meeting the set capacity targets. The newest amendment to the Wind Energy at Sea Act (Windenergie-auf-See-Gesetz (WindSeeG)) allows a one-time extension of the permit period by a maximum of ten years under special conditions, provided that the immediate subsequent use of the wind farm area is compatible with the site development plan published by the Federal Maritime and Hydrographic Agency (BSH) (cf. [49], [7], [6]). For further planning purposes, it is essential to understand if continued operation of German OWFs is economically viable since permit extensions will only be pursued if they are financially justified. Therefore, this work presents a feasibility study analysing the economic situation of German OWFs, aiming to provide decision support for planning the future use of currently occupied sites and informing the development of the site development plan and associated timelines.

Several operations and maintenance (O&M) simulation tools have been presented in the past. The Operation and Maintenance Cost Estimator (OMCE) of the Energy Research Centre of the Netherlands (ECN) focusses on calculating future O&M costs for offshore wind farms during

---

[1] Julia Walgern, David Baumgärtner, Johannes Fricke, Niklas Requate, Athanasios Kolios, Martin Dörenkämper, Tobias Meyer, Lukas Vollmer, 2023. „Economic feasibility study for continued operation of German offshore wind farms". IOP Journal of Physics Conference Series (EERA DeepWind conference 2023), doi: 10.1088/1742-6596/2626/1/012031 [212]

the operational phase [14]. The Norwegian offshore wind cost and benefit model (NOWIcob) [15] and Offshore TIMES developed by Fraunhofer IWES [17] are O&M logistics models predicting availability, costs and revenue utilising an event-based Monte Carlo simulation. However, all of these tools do not take CAPEX and financing related elements into account and are therefore not sufficient to determine the overall economic feasibility of an offshore wind farm. In comparison, Shafiee et al. [18] have developed a whole life cost analysis framework for offshore wind farms considering, in addition to the O&M phase, the development phase, installation and commissioning as well as decommissioning. Utilising historical cost figures as a baseline scenario, a combined multivariate regression / neural network model predicts costs and identifies key cost drivers of future projects. A deterministic approach is followed. Similarly, Ioannou et al. [19] have published a parametric life cycle techno-economic model considering all phases of offshore wind farms using the ECN's OMCE for O&M cost predictions. Results show cumulative cost return profiles and identified break-even points comparing different investor strategies. None of the publications above consider wake losses in their economic studies. Therefore, this work develops an economic life cycle simulation and assessment (ELSA) framework incorporating the O&M simulation tool Offshore TIMES and sophisticated wake loss calculations.

The chapter is outlined as follows: First, an introduction to the different models contributing to the ELSA framework for the economic feasibility study is given and required input parameters and related assumptions are described (Section 4.2). Afterwards, the profitability of each German OWF is analysed and generalised, and anonymised results are discussed and presented (Section 4.3). Last, a summary of main conclusions as well as an outlook to future work are given (Section 4.4).

4.2 Methodology

Within this study, the economic feasibility analysis is based on two major models contributing to the ELSA framework (see **Figure 7**): The O&M cost model "Offshore TIMES" [17] is utilised to simulate each OWF. It requires inputs such as reliability and O&M figures, the logistics concept, weather data and WT and wind farm (WF) information, to compute operational expenditure (OPEX) and annual energy production (AEP). Together with further input parameters such as the chosen feed-in tariff, available electricity price, capital expenditure (CAPEX), weighted average cost of capital (WACC) and repayment plan, yearly total project costs and total project revenue are estimated within the life cycle cost and revenue model which was developed within this thesis. Comparing those two quantities, the profitability of an OWF can be assessed:

$$profitability_i = revenue_i - OPEX_i - (residual\ debt_i + interest_i) \qquad (4.1)$$

Herein, all quantities are given in Euros and i denotes the time step. While revenue and OPEX are calculated with hourly resolution in the first place, for the profitability calculation both quantities are aggregated to yearly sums as residual debt and interest are derived from CAPEX on a yearly basis. Total project costs and revenue are derived using a net-present value approach.

In order to consider not only AEP from an O&M perspective, wake losses are calculated with a numerical weather model and are further input parameters for the life cycle cost and revenue model.

Additionally, the maximum design life of the wind turbines is estimated by assessing fatigue loads of the rotor-nacelle-assembly (RNA) using generic turbine models and by analysing design-related reserves of the foundations. This provides technical boundaries for the economic assessment.

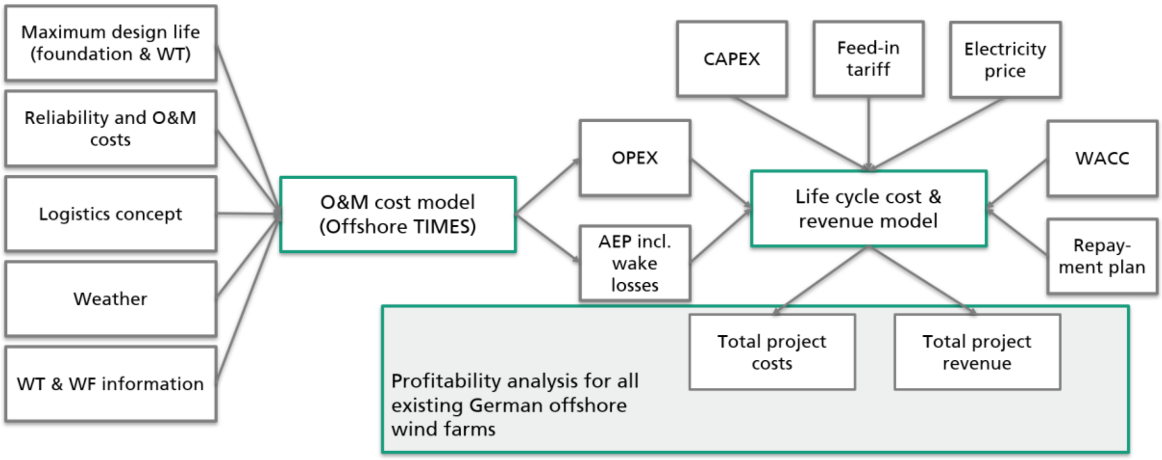

Figure 7. Economic feasibility analysis: workflow, inputs and outputs of the Economic Life cycle Simulation and Assessment (ELSA) framework

Different publications have been used as a basis for required input parameters (see [50], [12], [51]). Any inputs have been discussed with several OWF operators for verification and have been adapted where necessary. For this purpose, structured expert elicitation similar as in [52] was applied. Approximately ten stakeholder interviews with different entities were conducted using a previously defined questionnaire with around 50 questions.

All utilised input parameters for the analysis and relevant assumptions are presented and discussed in the following subsections.

4.2.1 Wind turbine and wind farm information

Currently in 2022, there are 28 OWFs including 1537 WTs installed in Germany. All relevant information required for the analysis is presented in Subsection 4.2.1.1In order to perform an economic feasibility analysis for each existing OWF but to reduce the amount of computational effort for determining the maximum design life by means of fatigue loads and dealing with the limited amount of publicly available information required as inputs, the WFs and turbine types are classified into three generic WT types. The clustering process and maximum design life considerations are outlined in Appendix B. The long-term yield potential of each OWF is investigated with the numerical Weather Research and Forecasting model (WRF) [53], which maps both the site-specific and the large-scale meteorological influences on the yield potential. This is important to understand how the WTs and WFs impact each other and how much AEP is reduced due to wake losses (cf. Subsection 4.2.1.3).

4.2.1.1 Considered offshore wind farms

Within this study all German OWFs are considered which were installed by 2022. An overview of the OWFs and their key characteristics can be found in **Table 2**.

The table is divided into three categories depending on the location of the OWF. This is indicated by the column "ROP Area" referring to the area defined by the maritime spatial plan (ROP) [54]: OWFs which are built in the German Exclusive Economic Zone (EEZ) of the North Sea (N-…) or Baltic Sea (O-…), and OWFs which are not defined by the ROP (indicated as "ROP area: none"). Furthermore, the commissioning date, the installed capacity of the WF, the WT rated power, the respective number of turbines and the WT manufacturer and type are summarised.

4.2.1.2 Generic wind turbine types and maximum design life

In total 1537 WTs are installed in German OWFs, with rated capacities ranging from 2.3 MW to 9.0 MW. A total of 13 different power classes are installed, in some of which there are further differences in terms of rotor diameters and hub heights.

Table 2. Overview of German offshore wind farms installed by 2022

Wind farm	ROP area	Year of Commis-sioning	Installed capacity in MW	WT rated power in MW	No of WTs	WT manu-facturer	WT type
alpha ventus	N-2	2010	60	5	12	Adwen & REPower	AD 5-116 & 5M
BARD Offshore 1	N-6	2013	400	5	80	BARD	Bard 5.0
Dan Tysk	N-5	2014	288	3.6	80	Siemens	SWT-3.6-120
Meerwind	N-4	2014	288	3.6	80	Siemens	SWT-3.6-120
Amrumbank West	N-4	2015	288	3.6	80	Siemens	SWT-3.6-120
Borkum Riffgrund 1	N-2	2015	312	4	78	Siemens	SWT-4.0-120
Nordsee Ost	N-4	2015	295.2	6.15	48	Senvion	6.2M126
Trianel WF Borkum I	N-2	2015	200	5	40	Adwen	AD 5-116
Global Tech 1	N-8	2017	400	5	80	Adwen	AD 5-116
Gode Wind 01	N-3	2017	330	6	55	Siemens	SWT-6.0-154
Gode Wind 02	N-3	2017	252	6	42	Siemens	SWT-6.0-154
Nordsee One	N-3	2017	332.1	6.15	54	Senvion	6.2M126
Sandbank	N-5	2017	288	4	72	Siemens	SWT-4.0-130
Veja Mate	N-6	2017	402	6	67	Siemens	SWT-6.0-154
Borkum Riffgrund 2	N-2	2019	448	8	56	MHI Vestas	V164
Deutsche Bucht	N-6	2019	260.4	8.4	31	MHI Vestas	V164
Hohe See	N-8	2019	497	7	71	Siemens	SWT-7.0-154
Merkur Offshore	N-2	2019	396	6	66	GE	Haliade 150-6MW
Albatros	N-8	2020	112	7	16	Siemens	SWT-7.0-154
Trianel WF Borkum II	N-2	2020	202.56	6.33	32	Senvion	6.2M152
Kaskasi	N-4	2022	342	9	38	Siemens	SG 8.0-167 DD
Baltic 1	O-4	2011	48.3	2.3	21	Siemens	SWT-2.3-93
Baltic 2	O-3	2015	288	3.6	80	Siemens	SWT-3.6-120
Wikinger	O-1	2017	350	5	70	Adwen	AD 5-135
Arkona Becken Südost	O-1	2019	384	6.4	60	Siemens	SWT-6.0-154

Borkum Riffgat	none	2014	113.4	3.6	30	Siemens	SWT-3.6-120
Butendiek	none	2015	288	3.6	80	Siemens	SWT-3.6-120
Nordergründe	none	2017	110.7	6.15	18	Senvion	6.2M126

In order to estimate the economic feasibility of all German OWFs with reasonable effort, the existing WTs installed in the WFs are assigned to three representative generic WT models. A description of the clustering process is presented in Appendix B.

Within the project, the generic turbine models were used to investigate the technical feasibility of a service-life extension with respect to fatigue loads of the RNA based on selected load cases. In addition, reserves in the remaining lifetime of the foundations were assessed, mainly based on comparing new and old design-standards. Various aspects of and decision bases for lifetime extension are part of current research (see e.g. [55], [56], [57]). While the presented procedure is not suitable for final evaluation of lifetime extension, it gives a rough estimation of the fatigue reserves related to the known design assumptions and the assumption on offshore wind conditions. This allows for simple assessment of technical maximum design life, which is utilised as boundary condition for the economic analysis. Further details fall outside the scope of this thesis but can be found in [58]. The key finding is that operation beyond 25 years is technically likely if economic conditions are feasible, including the need for potential additional replacement or maintenance costs.

4.2.1.3 Future yield potential and wake effects

The generated energy of each OWF is computed using the respective power curve and meteorological ERA5 reanalysis data [59]. However, the future yield of most existing German OWFs will decrease due to the wake effects from newly installed wind farms. The estimation of the development of wake losses within the service life of the WFs analysed in this study is based on two states of offshore wind deployment in different years: the year 2021 as representation of the current state of deployment and the year 2031 as representation of the full deployment of wind energy within the vicinity of the existing WFs. For the estimation of wake losses, data was available from a simulation in [58] describing the full deployment outlined in [6] to be realised approximately in 2040. This scenario is used as representative for the wake losses for the existing WFs in 2031, as WFs planned to start operation after 2031 will be located so far away that their influence can be neglected. The simulations in [58] are conducted with version 4.3 of the WRF model [53] using the Fitch wind farm parametrisation [60] for estimating the energy yield and the influence of the WFs on the wind field. **Figure 8** visualises the reduction of wind speed in the two selected deployment states compared to a simulation without any wind energy deployment. Wake losses are calculated for these two states of deployment relative to a gross production estimate, which is calculated with the wake engineering model suite FOXES [61], using the modelled wind fields without any wind farm deployment as input. Further information about the assumptions made about the state of the OWF deployment in 2031 and beyond and the model details can be found in [58]. The simulation results used for estimating the wakes for the state of deployment in 2031 are identical to Scenario 09 described in [58].

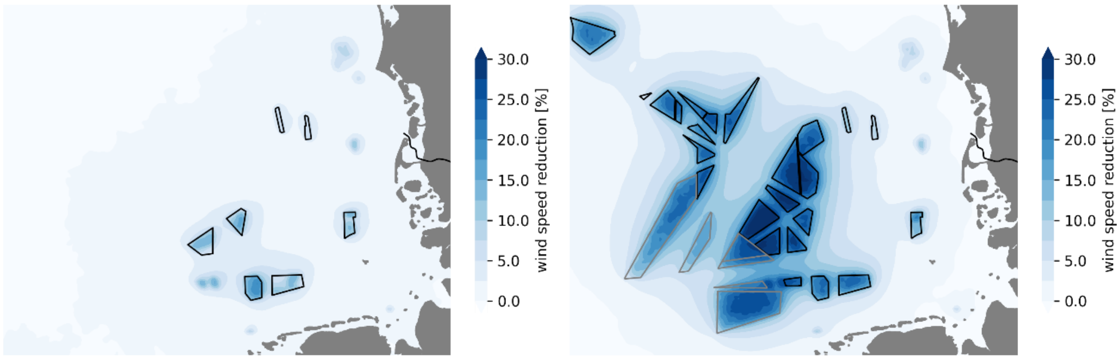

Figure 8. Simulated mean wind speed reduction in the German Bight under current wind farm deployment (2021, left) and for the future scenario of deployment used to project the wake losses (approx. 2040, right)

4.2.2 Operations and maintenance costs

Next to the investment costs, the O&M costs account for around 30% of the levelized cost of energy (LCOE) (cf. [19], [18], [50]). These consist to a large extent of corrective maintenance and annual planned maintenance. The required input for both maintenance measures cannot be broken down on a wind farm-specific basis, but approximations can be made based on turbine class and size.

The Offshore TIMES software developed by Fraunhofer IWES is used to determine the related O&M costs. It is a holistic, time series-based software for the investigation and planning of OWFs. It simulates the performance of maintenance tasks and the associated logistics of an OWF over its entire lifetime in order to determine important performance indicators such as the availability of the WTs or the O&M costs. The failure of WT's subsystems has a major impact on the maintenance work to be carried out on a WT. The reliability of these systems is simulated stochastically in Offshore TIMES. This means that the failure of a subsystem occurs with a certain probability depending on the type of failure. For this reason, the Offshore TIMES model uses a Monte Carlo simulation technique based on time steps, in which the maintenance and logistics of an OWF are simulated over several years of operation at variable time resolution (e.g., hourly). A simulation scenario is iterated in several Monte Carlo runs in order to be able to make a statistically significant evaluation across all simulation runs in later analyses. Offshore TIMES distinguishes between costs for technicians, vessels and repair costs. Required inputs, related assumptions and insights from stakeholder interviews are presented in the following subsections.

4.2.2.1 Vessels and technicians

Two exemplary logistics strategies were proposed as a basis for discussion for the stakeholder interviews and agreed on for the analysis. These are listed in **Table 3**. In Concept I, two crew transfer vessels (CTVs) per WF are used, which can be chartered in the home port as needed. In Concept II, a service operation vessel (SOV) is permanently stationed at the OWF instead, in order to directly handle work that arises. A jack-up vessel (JUV) typically applied for major component changes is used in both concepts.

Table 3. Logistics concepts

Vessel type	Concept I	Concept II
CTV	2	-
SOV	-	1
JUV	1	1

The following values for personnel and vessel costs as well as vessel characteristics were used for the analysis after discussion in the stakeholder interviews (cf. **Table 4** and **Table 5**). A day rate covers a 12-hours technician shift.

Table 4. Vessel and technician costs in Euro

Cost type	Day rate	Annual rate	Mobilisation cost
Technician	500	-	-
CTV	3,500	-	-
SOV	-	7,000,000	-
JUV	320,000 / 75,000	-	1,500,000 / 750,000

Table 5. Vessel characteristics

Vessel type	Velocity	Capacity for technicians	Maximum wave height	Maximum wind speed	Mobilisation time	Maximum time offshore
CTV	20 knots	12	1.5 m	15 m/s	1 day	1 day
SOV	20 knots	50	2.5 m	15 m/s	-	unlimited
JUV	11 knots	20	2.0 m	10 m/s	2 weeks	2 weeks

Most inputs regarding the logistics concepts were evaluated as suitable during the stakeholder interviews. The proposed logistics concepts were implemented in this way, with the WFs close to the coast being maintained using CTVs (Concept I) and those further away using an SOV (Concept II). In addition, the choice of JUV was differentiated according to turbine size. For repairs of the 3.6 MW turbine, a smaller JUV with lower cost rates (75,000 € per day) is sufficient. Furthermore, the interviews revealed that work during the night shift is not used in practice. It is therefore excluded in the study.

4.2.2.2 Corrective maintenance

Based on a field data study for OWTs by Carroll et al. [12] on mean turbine subsystem failure rates, average repair times, average material costs and number of technicians required per maintenance measure, trends for different WT generations were captured in the stakeholder interviews where quantifiable. Subsequently, assumptions were made for the inputs to the O&M cost model based on the available information. An initial overview of generic inputs that formed the basis for discussion in the interviews can be found in [12].

Based on the stakeholder interviews, the respective inputs were verified and adjusted as necessary. The main findings from the interviews and any adjustments to the input parameters are summarised below:

First, the assumptions regarding annual average failure rates of WT subsystems were found to be appropriate. In order to better adapt the reliability models of the generic turbines to the real WTs, a distinction was made between turbines with gearbox (generic turbines with 3.6 MW and 5 MW) and direct drive (generic turbine with 7.5 MW). In addition, the failure rate per WT and year for the gearbox in newer turbines (5 MW) was reduced (-50%) based on information from the stakeholder interviews. In addition, the replacement of the entire "hub" subsystem was removed as a possible corrective maintenance.

Second, the assumptions for active repair times at the WT were evaluated as roughly realistic. Even if some operators consider shorter repair times to be possible, the majority of interviewed operators agreed with these assumptions, so that they were retained for all subsystems. Also, for components of the largest generic turbine the repair times themselves remain in the same order of magnitude. However, fewer suitable weather time windows and a lower availability of the necessary ships – and with that longer waiting times – can be observed for larger components due to the required logistic.

Third, the material costs were evaluated as appropriate for older turbines (corresponding to the generic 3.6 MW class). For larger turbine types, higher material costs are often observed, and corresponding input parameters were adjusted (5 MW class: + 50%; 7.5 MW class: +75%). Likewise, there are large differences depending on the turbine OEM and availability in practice, which, however, could not be included in the simulation in the generic consideration of the present study.

4.2.2.3 Planned maintenance

Next to corrective maintenance, also the input parameters for annual planned maintenance have been discussed within the stakeholder interviews and have been defined (see **Table 6**).

Table 6. Input parameters for annual planned maintenance

Description	Value
Required vessel type	CTV / SOV
Required number of technicians	8 / 6
Required maintenance time per WT	24 h / 48 h
Material costs	20,000 €

Based on the feedback obtained in the stakeholder interviews, also the parameters for annual maintenance have been differentiated by WT size. The larger turbines (5 and 7.5 MW class) are maintained with more modern and efficient maintenance campaigns (i.e. with eight technicians in 24 h) compared to the 3.6 MW turbines (i.e. with six technicians in 48 h).

Additionally, most of the existing OWFs have full maintenance contracts with the OEMs for the first three to five years covering all occurring maintenance measures for a fixed price. An order of magnitude of 90,000 € per MW per year [50] has been estimated as realistic, although this figure can vary greatly depending on the WF and the portfolio size of a developer.

4.2.3 Remuneration

In order to estimate the revenue of each OWF, a differentiation between two time periods for the commissioning date of the WFs and the associated remuneration is necessary:

- Commissioning date up to and including 2020 with fixed remuneration per MWh
- Commissioning date from 2021 onwards with remuneration according to tender

The respective remuneration assumptions are outlined in the following subsections.

4.2.3.1 Subsidies

The remuneration according to the Renewable Energies Act (EEG) for OWFs with commissioning dates up to 2020 depends on four factors [62]:

- Year of commissioning
- Water depth at the location of the respective plant
- Distance to the German administrative territory
- Selected subsidy model (basic model or compression model; for WFs commissioned in 2020 only the basic model can be selected [62])

The remuneration for the compression model for a WF is shown in **Figure 9** in green. For the same WF in blue the respective remuneration is shown if the base model is chosen instead. Such time series have been created for each WT and then aggregated for each WF. While the compression model offers a higher initial remuneration of 19.4 ct/kWh for a shorter period, the basic model provides 15.4 ct/kWh over a longer duration. After the initial period, both models receive the statutory 3.9 ct/kWh. Consequently, WF operators must decide between higher short-term subsidies or lower long-term subsidies.

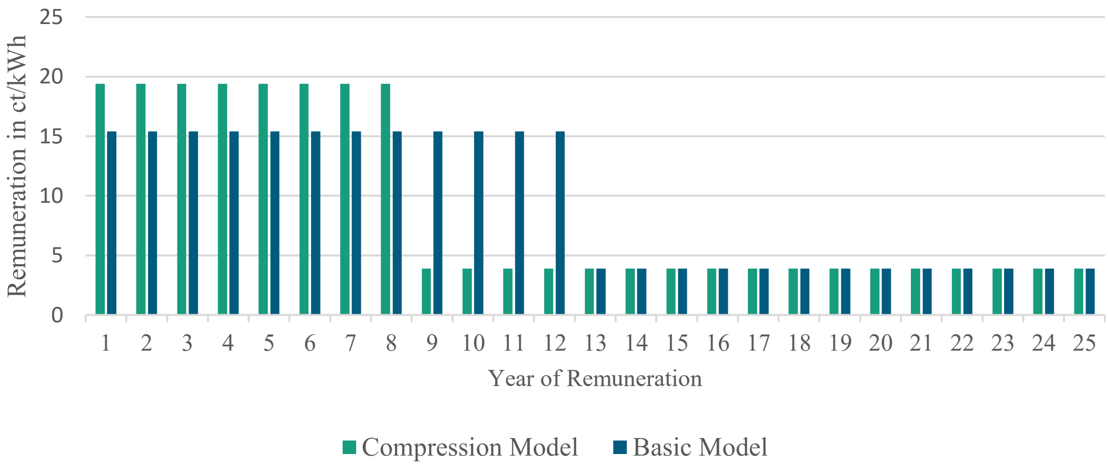

Figure 9. Exemplary remuneration of a WF with compression and basic model

Both funding models include an extension of the funding period based on water depth and distance to the administrative territory of the Federal Republic, which is derived from the following equation:

$$t_{extended\ subsidy\ duration} = (water\ depth_{>20\ m} \cdot 1.7 + distance_{administrative\ border} \cdot 0.5)\ months \qquad (4.2)$$

Herein, the water depth is specified in meters and the distance to the administrative border is specified in sea miles. This extended support period is always granted with the increased initial remuneration of the basic model, regardless of whether the basic or compression model was chosen.

The geo-positions of all WTs (at sea) have been obtained from the market master data register [63]. In addition, also the commissioning dates for the respective turbines are documented there. Where necessary, this list was corrected in consultation with the WF operators. In order to expand it to include the water depth of each individual WT, the water depths for all geo-positions were determined from the BSH's GeoSeaPortal [64]. A map with the administrative borders of the Federal Republic of Germany was obtained from the Federal Agency for Cartography and Geodesy [65] and used to determine the shortest distance to the geo-position of each individual WT. The support models chosen by the WF operators are not known. However, it can be assumed that most operators have chosen the compression model, as this allows a sooner repayment of the loans and thus results in a reduction of interest costs.

The remuneration for OWFs with commissioning from 2021 onwards results only from the surcharge value of the respective auction of the WF.

In addition to the remuneration through a fixed support model or the value of the tender, it is always possible to sell the electricity to the electricity market. It is generally assumed that the electricity will be sold to the electricity market if the electricity price exceeds the current price of the subsidy.

4.2.3.2 Electricity price forecast

Hourly electricity prices for the years 2010-2018 from [66] are used to model the electricity price. For a forecast into the future, these electricity prices are used repeatedly from 2019 onwards to continue the time series. In addition, the values in the years from 2019 onwards are multiplied by an annual factor so that they represent the average electricity prices from a study of Patzack et al. [67]. For the transition period 2019 to 2024, which are not examined in the study, a linear progression of the factor is assumed between year 2018 and the first forecast year of the study 2025. This approach results in the preservation of the hourly volatility of electricity prices from 2010-2018, but at the same time a prediction can be made for the future on the development of the annual price level. The annual mean of the electricity prices is shown in **Figure 10**.

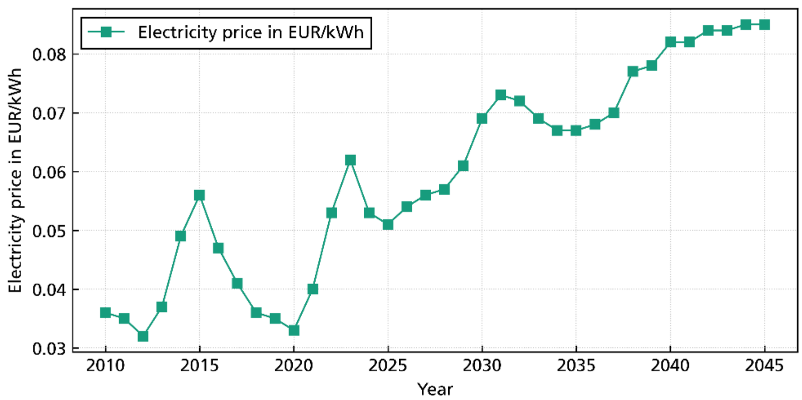

Figure 10. Yearly average of the electricity price time series used between 2010 and 2045

4.2.4 Capital expenditure

The level of investment costs is an important parameter that significantly contributes to whether a WF can be operated economically. The investment costs of OWFs depend on many factors, such as the size and location of the WF, the competitive situation, the contractual conditions or the exchange rate. Due to the large number of influencing factors and the high sensitivity of such data, an individual definition for each existing German OWF in the North Sea and Baltic Sea has not been possible. Instead, an estimate has been made based on a publication by BVGA, The Crown Estate and ORE Catapult, which break down a wide range of items for CAPEX for a British OWF [50]. The cost breakdown relates to a generic UK OWF with a rated capacity of 1,000 MW and a commissioning date in 2022. The costs quantified include development and project management, wind turbine, balance of plant, installation, and commissioning. In total, the investment costs for this generic wind farm amount to £2,370,000/MW. Based on these key figures, trends for different WF generations were to be captured in stakeholder interviews. While in principle 2,800,000 €/MW was considered appropriate for many and especially younger WFs, there were only few comments on older WFs. These comments were very specific and therefore cannot be generalised to all older WFs. Thus, in the further course of the economic analysis, uniform investment costs per MW have been assumed for all existing WFs. Due to similar reasons, a uniform and constant WACC of 5% as published by [50] and judged as realistic during the stakeholder interviews has been applied to all OWFs.

4.3 Results and discussion

The results of the profitability analysis of all German offshore wind farms show that the economic viability depends strongly on the electricity price that can be achieved after the period with guaranteed remuneration according to the EEG. Preliminary simulations with historical (i.e., low) electricity prices have shown that in such a scenario there would be no economic viability after the expiry of the EEG subsidy. Simulations with electricity price time series taking into account forecasts with continuously rising electricity prices usually lead to economic business cases beyond the approval period of 25 years of operation. Thus, it can be assumed that most wind farm operators will not only aim for 25 years of operation, but also for up to 10 years of continued operation according to the WindSeeG, as long as the electricity price rises sufficiently. It should be mentioned that the assumptions regarding the electricity price increase have been made in a way that leads to a conservative assessment of the economic viability of wind farms.

It must be considered that it can make economic sense not to repair a WT in the event of damage to major components towards the end of its service life, but to continue operating the WF with a reduced number of turbines and power output. Especially the lack of certain WT spare parts can become a major challenge. In the following subsections, some of the partial results of the profitability analysis are discussed in more detail.

4.3.1 Future yield potential and wake effects

Depending on the location of the existing WF, wake effects have a significant influence on the annual yield. While in 2022 wake losses in a range of approx. 7% to 31.5% can be observed for different existing WFs, in 2031, considering planned expansion scenarios, wake losses of up to 50% can be expected for a few existing WFs. Strongly varying effects can be observed: While some existing WFs are only slightly affected by the expansion, other areas will be confronted with doubling of the wake effects. **Figure 11** gives an overview of mean annual full load hours expected in 2031 compared to the current scenario in 2021.

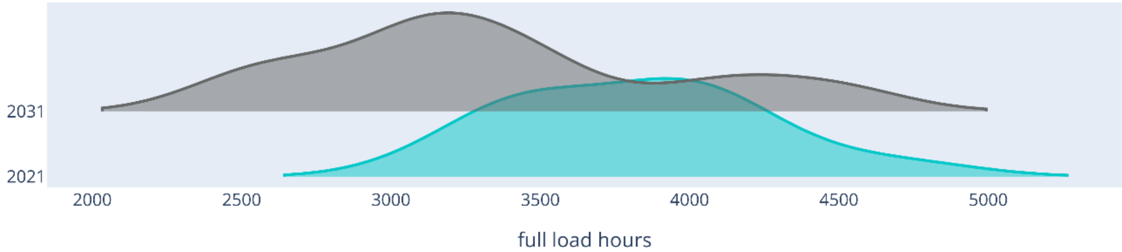

Figure 11. Distribution of mean annual full load hours of the analysed OWF (assuming 100% availability) in the two simulated expansion scenarios in 2021 and 2031

Note that the yearly full load hours that contribute to the feasibility study may vary due to different wind conditions in each year. Depending on the cost situation and financing, this reduction can have a strong influence on the economic viability of the WFs.

4.3.2 O&M costs

The economic viability also depends decisively on the O&M costs incurred. These in turn vary depending on the reliability of the WTs and their subsystems, the logistics concept, but also the distance of the WF from the coast. In the analysis results, average O&M costs per year and MW of around 50,000 € to 220,000 € can be observed. These results are in a similar order of magnitude as the assumptions from the study by [50] and the analysis of [68]. As O&M costs vary significantly for different OWFs, it should be taken into account that especially for older and smaller WFs as well as for WFs located further away from the coast, higher O&M costs are to be expected compared to WFs with a more recent commissioning date typically having a larger WF and WT nominal capacity. Additionally, differences in O&M costs per MW can be explained by different turbine technologies evaluated within this study.

To eliminate a potential variation of O&M costs resulting from a limited number of Monte Carlo iterations, it was investigated how many runs are required to obtain stable, meaningful results. **Figure 12** illustrates how availability and O&M costs change with increasing number of Monte Carlo runs. Starting at around 200 runs both values start to converge and do not change any more relative to the values computed with 300 runs. With the aim of acquiring results of good accuracy with a reasonable computational effort, 300 runs have been determined as sufficient in the present study.

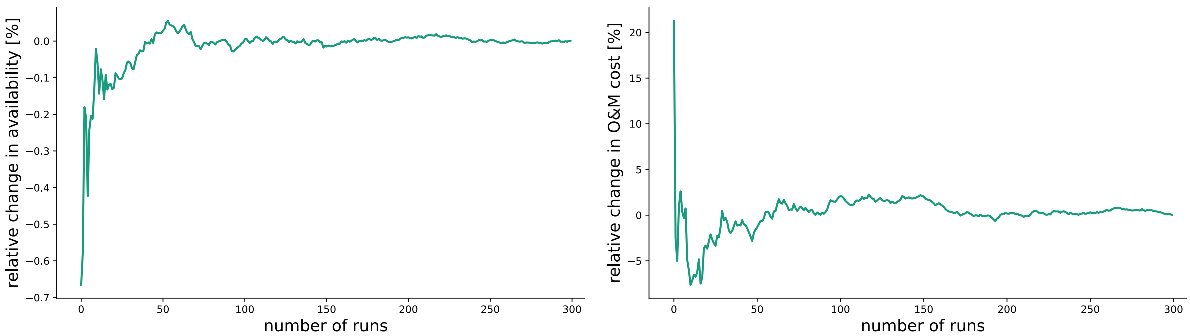

Figure 12. Relative change in averaged availability (left) and O&M costs (right) dependent on number of Monte Carlo iterations

Particularly for smaller WFs, independently performed maintenance cannot be carried out economically in some cases. To shed more light on this aspect, scenarios with costs for corrective maintenance and costs for full maintenance contracts in the first five years of operation have been developed and compared. While for some existing WFs independently organised maintenance is more cost-effective, other WFs, especially older and smaller ones, benefit from full maintenance contracts. **Figure 13** gives an overview of the expected profitability of existing German OWFs considering independently organised maintenance and full maintenance contracts, respectively. While for the scenario with independently organised maintenance, 82.1% of OWFs are profitable for at least 25 years, this share increases to 89.3% in case of applying full maintenance contracts. This results in only 17.9% in the one and 10.7% of German OWFs in the other scenario for which continued operation is not economically viable. Given the results presented here, but also others that cannot be disclosed for confidentiality reasons, it is clear that continued operation is economically attractive for most German OWFs (cf. Figure 13).

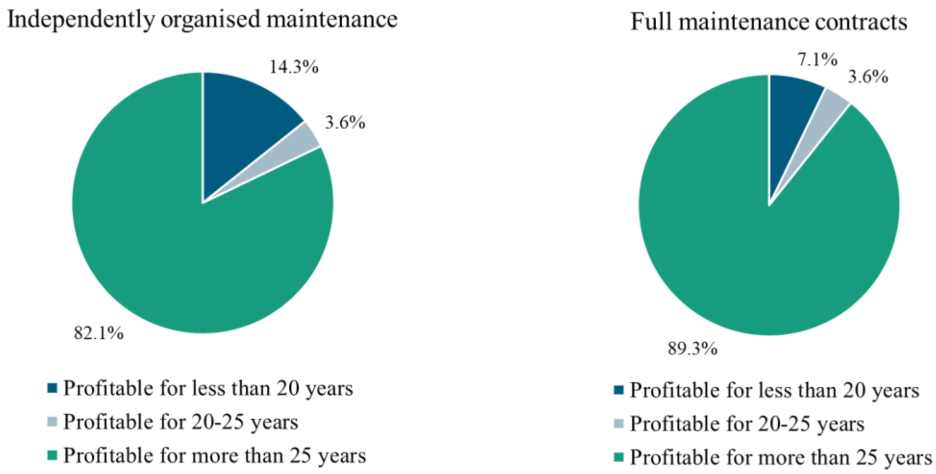

Figure 13. Expected profitability of existing OWFs in case of independently organised maintenance (left) and full maintenance contracts (right)

4.3.3 CAPEX and financing model

Finally, the influence of CAPEX and financing models should be discussed more deeply. Depending on the existing WF, initial investment costs can vary greatly and how the WF is financed also plays a decisive role in the profitability analysis. However, as only few interviewees commented on these aspects, uniform investment costs and WACC assumptions were used (cf. Section 4.2.4). Nevertheless, these assumptions together with all the simulation results provide a coherent overall result, as continued operation beyond the approval period of 25 years was also described as conceivable and financially attractive in the stakeholder interviews. Thus, based on both the simulation results and the stakeholder interviews, the conclusion can be drawn that continued operation is an attractive option for most existing German OWFs.

4.4 Conclusions and outlook

In this chapter, the economic feasibility for extended operation of German offshore wind farms has been investigated. A comprehensive economic life cycle simulation and assessment (ELSA) framework has been developed for this purpose that can be applied in different portfolios of wind power assets. It covers a cost/revenue model which incorporates CAPEX and OPEX related elements, revenue factors, and deployment location specific aspects. To limit the level of granularity in the present study, a classification of all existing wind farms in German waters with respect to size and key dimensions has been used. As a key conclusion, the results support the case for extended operation for most German OWFs, while highlighting the importance of wake effects to AEP, the magnitude and variability of O&M costs and finally the influence of CAPEX and financial modelling.

The outcomes of this study provide promising conclusions towards obtaining additional value from existing assets by means of service-life extension. However, certain assumptions should be addressed in order to increase confidence and further quantify the outcomes. More specifically:

- The conclusions obtained rely heavily on the inputs provided. In this chapter, inputs have been mainly obtained from literature and interviews with stakeholders. More quantitative inputs from real wind farms including details about O&M activities and cost figures would provide more reliable results.
- Reliability data is scarce in the literature or the public domain, while they rely significantly on the type of technology, the maturity of turbine concepts and the period of operation. The development and utilisation of improved reliability models taking into consideration also the effect of component age and loads would considerably reduce uncertainties in the analysis.
- At the same time, maintenance logistics inputs are affected by the deployment location and the maturity of the supply chain. Continuous development and adaption of Offshore TIMES are essential to cover newest logistic concepts within the simulations.
- Within this study the nominal service life of wind turbines is defined regardless of the nominal service life of major components such as blades, the gearbox or the generator. In case of a required replacement close to the end of service life, the decision strategy within the framework is not adjusted. Consequently, unrealistic costs can be included in the analysis when having a break-down in the last days of operation which would not occur in real life. To this end, higher fidelity models could be employed once relevant input data become available, investigating the effect of major replacements to the business case of service life extension.

4.5 Acknowledgements

Fraunhofer IWES was commissioned by the Federal Maritime and Hydrographic Agency (BSH) for the “Further Development of the Framework Conditions for the Planning of Offshore Wind Turbines and Grid Connection Systems” in 2021. The presented work was carried out within this advisory project. The discussions with the BSH as well as interviewees from the German offshore wind industry within the project are gratefully acknowledged. ERA5 reanalysis data were obtained from the Copernicus Climate Change and Atmosphere Monitoring Services.

5 Digitalisation and preprocessing of O&M data under consideration of standards and guidelines

Maintenance data of wind turbines is an important information source for calculating key performance indicators. Also, it can be used for developing models for early fault detection. Both activities aim for supporting informed decisions in operations and maintenance. However, such data is rarely available in a structured and standardised format which hinders the interoperability of different enterprises. Consequently, maintenance information is often unused or only usable with considerable personnel effort. To digitalise wind farm maintenance, a digitalisation workflow is developed and presented in this chapter. The workflow consists of the steps optical character recognition, information extraction and text classification. The workflow is applied on real-world wind turbine service reports and invoices. First results for each step show good performance metrics and potential for further real-world application of the proposed method. The material of this chapter has been peer reviewed and published in [2].

5.1 Introduction

Up to 30% of the levelized cost of energy (LCOE) of wind turbines (WTs) are driven by the cost of operations and maintenance (O&M) [69]. In order to further decrease this share and to be able to perform informed and strategic decisions in the maintenance of WTs, key performance indicators (KPIs) based on maintenance information are crucial. Modern enterprises are embedded in a fast-paced environment which is dependent on data. In order to use this data and to ensure interoperability of different enterprises, a common structure and standards are necessary. Even though there are already standards and guidelines available for the communication of maintenance data and to classify affected components of a WT during a maintenance measure, such as Reference Designation System for Power Plants® (RDS-PP®) [40] or Reference Designation System for Power Systems (RDS-PS) [70], they have only been applied to a limited extent to date. This results in a variety of maintenance reports and formats and hinders communication between enterprises. Consequently, maintenance information is heterogeneous, unstructured and rarely standardised. Since utilisation of maintenance information is only possible with considerable personnel effort, KPI-driven maintenance optimisation or developing models for early fault detection is hindered and scarce within the wind energy domain. Even if standards are implemented at some point, historical data cannot be used for further analysis or reliability KPI calculation yet. To be able to use historical data and to speed up the usage of standards in the maintenance of WTs, a workflow for digitalisation of maintenance information (digitalisation workflow) (see **Figure 14.**) is proposed. The digitalisation workflow consists of the three steps optical character recognition, information extraction and text classification. It produces standardised, machine-readable maintenance

[2] Marc-Alexander Lutz, Julia Walgern, Katharina Beckh, Juliane Schneider, Stefan Faulstich, Sebastian Pfaffel, 2022. „Digitalization Workflow for Automated Structuring and Standardization of Maintenance Information of Wind Turbines into Domain Standard as a Basis for Reliability KPI Calculation". IOP Journal of Physics Conference Series (WindEurope Annual Event 2022), doi: 10.1088/1742-6596/2257/1/012004 [107]

information, in which WT components are classified according to the RDS-PP scheme. To validate the workflow, real-world WT service reports and invoices are used.

The outline of the chapter is as follows: After the introduction (Section 5.1) the state of the art on digitalisation methods including optical character recognition, information extraction and classification is presented in Section 5.2. Furthermore, an overview of the challenges with current maintenance data of WTs and of available standards is given. In Section 5.3 the methods used for comparing WT service reports and for implementing the digitalisation workflow are outlined. Also, the dataset utilised for validating the digitalisation workflow is described. Afterwards, the results for the aforementioned methods are presented and briefly discussed (Section 5.4). In Section 5.5 the findings of the chapter are summarised and an outlook is given.

5.2 State of the art on digitalisation methods and existing standards

After describing the current situation on how maintenance information is recorded and stored within the wind industry, different approaches for digitalising data are reviewed. Afterwards, the following sub-chapters revise the state of the art of the individual steps of the proposed digitalisation workflow.

5.2.1 State of the art on recording maintenance information

Maintenance information are available in numerous maintenance documents. File formats can vary from documents which are structured already (Excel, CSV) to scans of letters in pdf or image format. Dependent on the document type, the depth of information about the maintenance measure differs: While maintenance reports or turbine logs can provide detailed failure and measure descriptions, invoices often give limited technical insights. A large heterogeneity of documents complicates the data processing and not from all documents the same information content can be extracted. Moreover, even within one type of document the depth of information can vary significantly. This is mainly related to data sovereignty: Dependent on which service enterprise (SE) is performing the maintenance service at the WT and which documentation has been agreed on, the information content differs substantially. A reduced information content hampers labelling the maintenance measure and the related WT component. Standardisation of different document types and their content is crucial to evaluate and communicate the data in an automated manner. As the majority of maintenance information is described in free text, text classification is required to label the affected component consistently and to make the information comparable (cf. [71]).

5.2.2 State of research on approaches for digitalisation of maintenance information

Recommendations on how to collect data for the purpose of reliability assessment are summarised in [20]. However, real-world data is often not collected according to this recommendation. Therefore, solutions to make use of this data are necessary. Approaches that aim to extract knowledge from natural language in maintenance data of different domains are summarised under the term Technical Language Processing [72]. Discovering KPIs from natural language in maintenance work orders is shown in [73]. Gao et al. focus on the extraction of performance metrics, e.g. mean time to failure, and therefore implement a pipeline for machine reading of unstructured maintenance work orders [74].

Approaches that are specific to the wind energy domain can be seen in ([75], [76], [77]). Blanco-M. et al. show a text mining approach to assess the failure condition of WTs by using the maintenance history [75]. A domain ontology for wind energy is presented in [78]. Other methods combine WT sensor data and maintenance reports for the purpose of predictive maintenance [77]. Further approaches focus on textual information of accidents at WTs to generate knowledge ([79], [80]). However, none of the before mentioned approaches consider domain standards like RDS-PP [40] or RDS-PS [70] for describing the affected components in a uniform manner. Furthermore, these methodologies have not yet been adopted within the wind energy industry.

5.2.3 Optical character recognition

Optical character recognition (OCR) is a well-researched field for more than six decades ([81], [82]). Several reviews and books within this domain are available ([82], [83]). An evaluation of the performance metrics of various OCR approaches is provided in [84]. Different OCR tools such as ABBYY FineReader [85], OmniPage [86] or Tesseract [87] exist. A comparison of Tesseract to other tools is shown in [88].

5.2.4 Information extraction

Information extraction aims at creating a structured representation of selected information out of text [89]. Information extraction is a major task in computer linguistics that is subdivided in several areas, e.g. template filling, named entity recognition, identification of relations and event detection [90]. An overview of the domain-specific information extraction is given in ([89], [91]). The various applications are shown in [92]. Approaches to extract content and to derive relations out of PDF documents are shown in ([93], [94]).

5.2.5 Classification

Classification in the machine learning sense is the task to automatically assign categories to data points. Depending on the task, it can be distinguished between assigning one class to a data point (multi-class) and assigning several classes (multi-label). The domain standard that is considered in this work, RDS-PP®, contains a number of categories up to the four-digit range. Thus, relevant classification methods need to be able to tackle this amount of categories. Two types of classification models are suitable. First, support vector machine (SVM) is a one-vs-all approach that uses one classifier for each class ([95], [96]). Originally intended to perform binary classification (assigning one out of two classes) the SVM can be employed for multi-label as well. Second, word embedding approaches map high dimensional and sparse co-occurrence matrices based on the words that occur in the training data to a low dimensional latent vector space [97]. Most current approaches make use of this method. Language models such as BERT and other transformer variants ([98], [99], [100]) use pre-trained models and fine-tune on given data. They can be used, among other applications, to perform classification tasks.

5.3 Methodology

In this section the methodology for each step within the digitalisation workflow and the datasets used for validation are described. First, it is explained how wind turbine service reports are compared (Subsection 5.3.1). Second, the datasets used for validating the different steps of the proposed digitalisation workflow (Subsection 5.3.2) are described. Third, an overview of the implementation of the digitalisation workflow is given (Subsection 5.3.3).

5.3.1 Comparison of existing field data: wind turbine service reports

An overview of initiatives that collected maintenance data of WTs for the purpose of reliability studies is shown in ([31], [101], [102]). Acquisition of data but especially evaluation is most efficient if same formats and structures are used. This can be the case for maintenance data being collected and structured by initiatives or for maintenance data recorded by the same SEs. As stated before, in real world a variety of different templates is used. In order to understand which information content is available and needs to be considered within the digitalisation workflow, several service reports of WTs are compared and the available information content provided by different SEs is clustered into categories such as site name, WT-type or the text description of the maintenance measure (TDoMM). The dataset for this comparison consists of five different types of service reports, which have been recorded by five different SEs.

5.3.2 Datasets used for the validation of the digitalisation workflow

Two different datasets are used to evaluate the steps of the digitalisation workflow. The first dataset "invoices" is used for the validation of the steps OCR and information extraction. It consists of 152 invoices that have been issued by one SE to an operator of WTs. The invoices are available as scanned images. For the purpose of enterprise resource planning (ERP) the relevant information of the invoices, such as the issued date, the invoice number and the invoice total, are currently transferred manually into the ERP system. Next to the invoices, this information is also available. Therefore, the manually inserted information from the ERP system can be used as a ground truth to evaluate the performance of the first two steps of the digitalisation workflow.

The second dataset "service reports" is used for the steps of information extraction and classification. It comprises of 4000 offshore WT service reports covering 240 operational turbine years. The service reports are provided in PDF format and contain the date of the measure, the number of technicians, the working hours, the materials used as well as the TDoMM. Initially, the TDoMM is not labelled with a domain standard such as RDS-PP. As a first step, the labelling according to RDS-PP is performed manually by a domain expert. Each TDoMM is labelled with one or more RDS-PP component categories depending on how many measures are undertaken at different WT components. Once all available TDoMMs within the WT service reports are labelled according to RDS-PP, a basis for training and testing the classifier is available.

5.3.3 Digitalisation workflow

The digitalisation workflow, depicted in **Figure 14.**, consists of the three steps: OCR, information extraction and text classification. The different steps of the digitalisation workflow are theoretically described in Lutz et al. [71] and are briefly outlined again.

At different stages of the digitalisation workflow, structured information can be archived. This can be the case following the steps of OCR and information extraction or after the classification step, which even allows for structured and standardised information being stored. Since archiving can be easily automated, this step is illustrated in **Figure 14.** as part of the digitalisation workflow but is not explained in the following. If maintenance information is present in the form of pictures or scans, at first it is necessary to apply OCR. By doing so, scans are converted, the output thereafter is machine-readable text. Next to providing semi-structured files like .csv or .txt, the output of the OCR can serve as input for the step of information extraction. The purpose of information extraction is to convert multiple different files into one common structure where in same columns same information categories can be found, e.g. the date of the maintenance measure or the TDoMM. Thus, many different service reports can be converted into a single tabular form, e.g. a spread sheet. With the aforementioned steps, structured information is available. Since some information categories, such as the TDoMM, are often given as free-text information, those information still need to be standardised utilising predefined categories. By using text classification methods, this text can be classified into categories which are given by a domain standard (e.g. RDS-PP [40], RDS-PS [70] or State-Event-Cause-Code (ZEUS) [41]). This results in structured and standardised maintenance information being available and usable for further analyses.

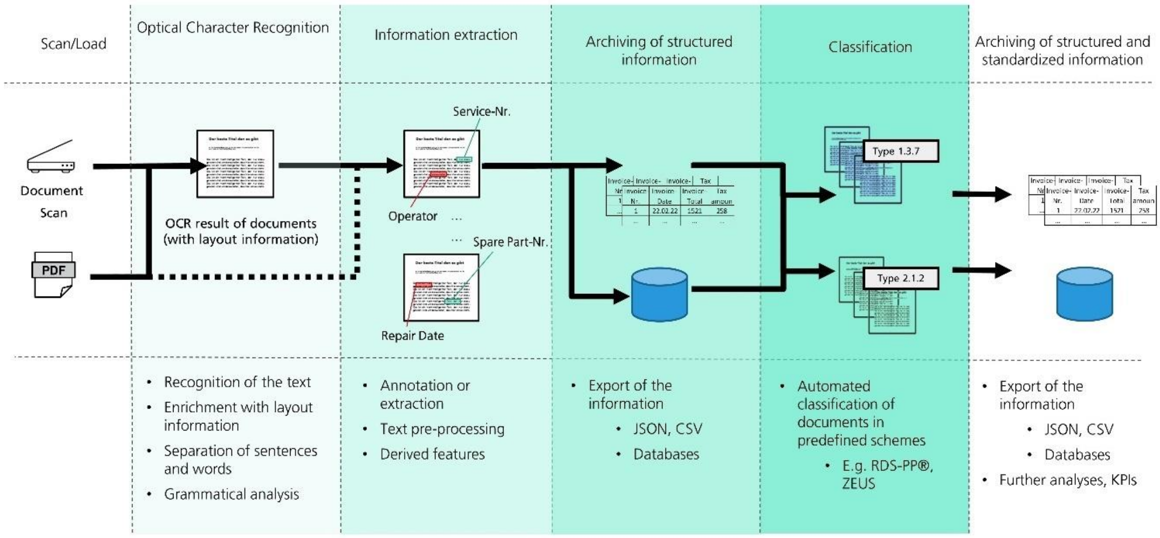

Figure 14. Overview of the proposed digitalisation workflow

5.3.3.1 Optical Character Recognition

Maintenance data can be available in different formats. If data is available as scanned images or pictures, information can neither be extracted nor can the content of the data be searched. To be able to access the information, OCR is applied which converts images into text. The text can then be searched and further used. The OCR output can be used as input for the next digitalisation workflow step. As stated in Section 5.2, OCR is a well-researched domain. Therefore, OCR can easily be implemented into the digitalization workflow. In this work the authors use the OCR open-source engine Tesseract [87] as it provides a good performance in comparison to other tools [88].

5.3.3.2 Information Extraction

Besides the output of the OCR, semi-structured files like .csv, .pdf or .txt can be used as input for the step of information extraction. In comparison to the definition of information extraction within the field of computer linguistics (see Section 5.2), information extraction within the proposed digitalisation workflow has the purpose of converting multiple different single files into one common structure in which same information categories can be found in same columns, e.g. the date of the maintenance measure or the TDoMM. Information extraction within the proposed workflow comprises of the sub-steps template filling and text preprocessing. The sub-step of text preprocessing is similar as outlined in [103]. Within the sub-step of template filling each part of the text is assigned to an information category, which allows for transferring information in the same structure. With the aforementioned sub-steps, structured information is available. However, some information categories still need to be standardised, for instance the TDoMM. This can be achieved by using text classification methods, which are described in the next Section.

5.3.3.3 Classification

The data basis used for the classification is the dataset "service reports" containing around 4000 service reports of WTs, which are manually labelled according to RDS-PP by a domain expert. RDS-PP is hierarchically structured (see Appendix C, **Figure 55**). An example for a possible classification of the TDoMM in WT service reports to RDS-PP is given in **Table 24** in Appendix C. By analysing the dataset, it is found that on RDS-PP level two, around 55% of all service reports are represented by only three RDS-PP labels (see **Figure 15.**). Therefore, the dataset is imbalanced.

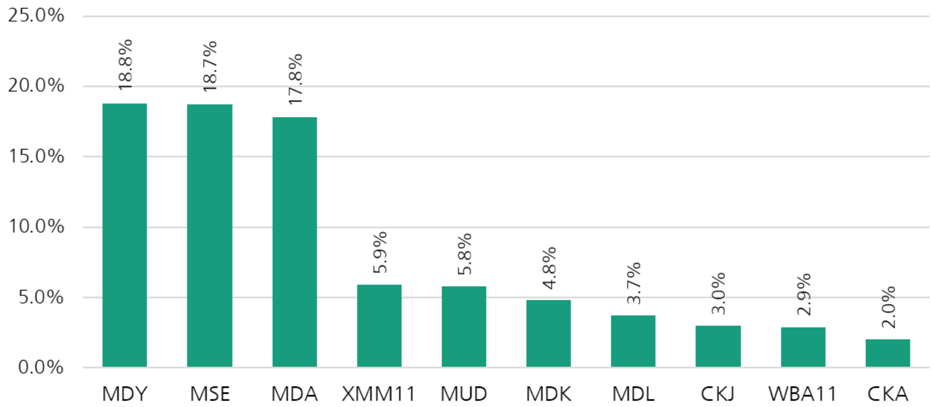

Figure 15. Distribution of the ten most frequent RDS-PP labels on level 2 within the dataset “service reports”

While on RDS-PP level one, 16 different categories are theoretically available, on level two 62 RDS-PP categories can be used. Therefore, typical classifier choices are established methods such as SVM [95] and transformer variants such as BERT [98]. Transformer variants require a large corpus of domain-specific text which is not available for WTs. Consequently, SVM is used as classification method. Within the dataset each TDoMM is labelled with one or more RDS-PP components, which makes it a multi-label classification.

In general, the text classification consists of four steps:

- Text cleaning: E.g. removing special characters and lowercasing
- Vectorisation: Each report is represented by a vector
- Training: Splitting the data in train and test set and training a classifier on the train set to learn a mapping from text to RDS-PP label(s)
- Evaluation: Test the classifier on the remaining test set

Each TDoMM is transformed with TF-IDF (term frequency-inverse document frequency) [103], which assigns a value of information content to each word. This results in low values for frequent words, e.g. "the", and high values for relatively infrequent words, e.g. "nacelle". In particular, for each RDS-PP level one SVM is trained such that performance metrics for each level can be compared. In addition, one classifier is trained which is referred to as final. For the final classifier all given labels on all respective levels are used for training and prediction.

This classifier allows for labelling all TDoMMs without knowing which RDS-PP level is searched for and primarily serves as a reference point for now.

Evaluation is performed with F1-scores. F1 is calculated from recall and precision as follows

$$F_1 = 2 * \frac{precision * recall}{precision + recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{5.1}$$

Herein, TP, FP and FN refer to true positives, false positives and false negatives, respectively. F1 is a metric that is more robust towards class imbalance. Thus, it is the preferred choice over other metrics such as accuracy. A distinction can be made between micro F1, which aggregates contributions of all classes and is therefore weighted by label frequencies, and macro F1, which computes the metric independently for each class an then averages the results, thereby treating all labels equally regardless of their prevalence. Micro F1 is particularly suitable for evaluating classifier performance on imbalanced datasets where the frequent labels dominate, while macro F1 is more appropriate when equal importance should be given to all classes, including rare ones.

5.4 Results and discussion

5.4.1 Comparison of existing field Data: wind turbine service reports

Different service reports of five SEs are compared. Results can be seen in **Table 7**. It can be noticed that some information categories are available in all service reports, e.g. the type and the identifier of the WT, the date and the TDoMM. However, the description of the TDoMM is not documented in a standardised way, for instance by using RDS-PP [40], RDS-PS [70] or ZEUS [41]. Instead, free text is used. Different information categories can be seen in service reports of different SEs. Also, the type of documentation format varies. The service reports are available as scanned images, Word documents, XML or PDF-files. The results of this comparison (see **Table 7**) support the assumption that maintenance data provided by different SEs differ significantly regarding the structure and information content. Furthermore, standards for classifying maintenance activities are rarely used. This raises the challenge as well as the demand of an automated digitalisation workflow.

Table 7. Available information categories in service reports of different SEs

Service Enterprise	Site Name	Costumer	WT Type	WT Number	Service Number	Energy Produced	Operational Time	Maintenance Reason	Wind Speed at arrival	TDoMM	Material Consumption	Time of Maintenance	Open Tasks	Comments	Technicians	Working Hours	Date Next Measure	Status of Work Order	Commuting Distance	ID Work Vehicle	Contract Number
1	X	X	X	X	X	X	X			X	X	X		X	X	X	X	X			
2	X		X	X	X				X	X			X	X	X						
3	X	X	X	X	X			X		X	X	X			X	X					
4	X	X	X	X	X			X		X	X	X			X	X					
5	X	X	X	X	X	X		X		X	X	X				X			X	X	X

5.4.2 Digitalisation workflow

In the following sections results for the different steps of the digitalisation workflow are presented.

5.4.2.1 Optical character recognition

OCR is conducted on the first dataset "invoices". As described in Section 5.3.2, 152 invoices are available as scanned images. After using OCR all invoices are converted to machine-readable text. Since an automated evaluation of the feasibility of this process is only possible after the step of information extraction, results are described in the following Section 5.4.2.2.

5.4.2.2 Information extraction

After the step of OCR, the dataset "invoices" is further processed using the step of information extraction. As a result, all invoices are stored in the same structured format. This allows for comparing and validating the extracted information categories with the ground truth data exported from the ERP-system (see Section 5.3.2). The results can be seen in **Table 8**.

Table 8. Results of OCR and information extraction for the first dataset “invoices”

Number of invoices evaluated	152
Share of correct entries [%]	94.5

Next to the number of invoices evaluated, the share of correct entries is shown. One entry is defined as the content of a certain information category in one invoice such as the issuing date of the invoice or the amount of cost stated in the invoice. Out of all available entries being extracted a share of 94.5% is similar as listed in the ground truth and is therefore evaluated as "correct".

Furthermore, information extraction is performed on the dataset "service reports". All different information categories in the WT service reports of the SE could be extracted and summarised in a common table. The content of each report is represented in a row of that common table. Same information categories are found in same columns. No ground truth data is available to validate automatically if all data is extracted correctly. Nevertheless, after carefully comparing the content of many service reports with the structured result manually, the authors believe that all service reports are extracted correctly.

5.4.2.3 Classification

For the TDoMMs of the dataset "service reports" single-label and multi-label classifiers for different RDS-PP levels are trained and tested. Classification results are depicted in **Figure 16**. Each level represents a level in the RDS-PP hierarchy. **Figure 16(a)** and **(b)** show results for multi-label and **Figure 16(c)** and **(d)** for single-label classification. **Figure 16(a)** and **(c)** refer to micro F1 while **Figure 16(b)** and **(d)** depict macro F1-scores. In both cases, multi-label and single-label, macro F1-scores are lower than micro F1-scores. This can be explained by the label imbalance in the dataset, i.e. that some labels occur more frequently than others (also see **Figure 15.**).

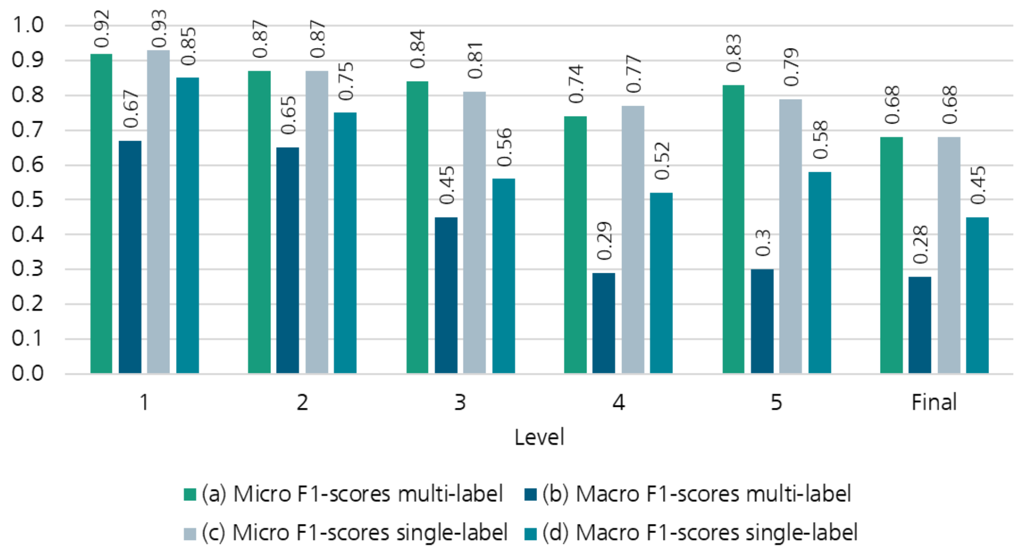

Figure 16. Micro and macro F1-scores of single-label and multi-label classifiers for different RDS-PP levels

As a result, infrequent labels cannot be captured well by the classifier and the micro F1-scores are higher than the macro F1-scores. In the multi-label as well as in the single-label setting it can be seen that with ascending levels (from one to five) F1-scores decrease. This is expected since the number of data points that belong to a certain level decrease from level one to level five, while, at the same time, the level of annotation detail is increasing. Only classifier results for level five show outliers in this descending trend. Even though more RDS-PP labels are theoretically available for labelling the TDoMM in WT service reports, only 16 different labels are used for level five within the dataset. In comparison, over 70 labels are used for labelling on level four as information depth is rarely precise enough to label until level five. Therefore, the actual label space in level five is smaller than e.g. for level four leading to better classifier performance for level five. Inspecting the macro F1-scores, the results show higher scores for the single-label than for the multi-label setting. This can be explained by the fact that the single-label setting reduces the problem to predicting only one affected WT component because multi-label cases are not included in the training and test set. This effect cannot be observed when comparing the micro F1-scores which needs to be analysed further. Overall, the results indicate that a classification system can provide decision support in the annotation process and for certain RDS-PP levels an automatic labelling procedure is possible.

5.5 Conclusion and outlook

WT service reports are often heterogeneous. Their information content and information depth vary. Also, no uniform structure or standards are used for documentation. Data preprocessing as well as determining the maintained component are therefore essential steps for enabling further analyses. At the same time, these steps can only be performed manually so far, thus are time-consuming, and often require the knowledge of a domain expert.

This chapter proposes a digitalisation workflow for maintenance information of WTs including the steps of OCR, information extraction and classification. Since OCR and information extraction show good results with an accuracy of around 95% or above depending on the dataset evaluated, the focus lies on classification of maintained components according to the standard RDS-PP. The process uses a SVM for text classification on a real-world dataset. First classification results look promising with micro F1-scores showing high values. However, as not every component is maintained in the same frequency, the RDS-PP label distribution in the dataset is imbalanced. Thus, higher represented labels are captured better by the classifier. A lower macro F1-score is therefore seen. Beside the label distribution, investigations show that the level depth of assigned RDS-PP labels also has an impact on the classification performance. The digitalisation workflow allows utilising operational knowledge which was not accessible in the past. Furthermore, the results indicate that automated digitalisation of maintenance reports of WTs is possible and the usage of standards can be accelerated. Thereafter, KPIs can be calculated automatically which supports data-driven decision making. Further analyses and uncertainty evaluations for different datasets are planned to drive decisions for productive use. Also, classification for other guidelines and standards, e.g. ZEUS or RDS-PS, will be tested. The comparison of different classifiers and associated insights will be utilised to develop classifiers further considering application-specific requirements.

5.6 Acknowledgements

This research is to a large part carried out within the research project "Digitalization of Maintenance Information" (DigMa) funded by the German Federal Ministry of Economic Affairs and Climate Action (BMWK), grant number 03EE2016A. The authors thank the project partners for providing field data and for sharing their requirements and experience.

6 Impact of using text classifiers for standardising maintenance data of wind turbines on reliability calculations

This chapter delves into the challenge of efficiently digitalising wind turbine maintenance data, traditionally hindered by non-standardised formats necessitating manual, expert intervention. Highlighting the discrepancies in past reliability studies based on different key performance indicators (KPIs), the chapter underscores the importance of consistent standards, like RDS-PP, for maintenance data categorisation. Leveraging on established digitalisation workflows, we investigate the efficacy of text classifiers in automating the categorisation process against conventional manual labelling. Results indicate that while classifiers exhibit high performance for specific datasets, their general applicability across diverse wind farms is limited at the present stage. Furthermore, differences in failure rate KPIs derived from manual vs. classifier-processed data reveal uncertainties in both methods. The study suggests that enhanced clarity in maintenance reporting and refined designation systems can lead to more accurate KPIs. The material of this chapter has been peer reviewed and published in [3].

6.1 Introduction

Maintenance data of wind turbines is essential for analysing operations and maintenance (O&M) activities and for calculating related key performance indicators (KPIs). Corresponding data can facilitate the optimisation of O&M through logistic concept improvements or the implementation of preventive maintenance strategies, reducing the levelized cost of energy (LCoE). However, maintenance data in the wind industry is seldom available in a machine-readable and standardised format. As a result, this data is either overlooked or requires significant manual effort of a domain expert to process the information content.

Numerous reliability studies based on manually labelled maintenance reports have been published to understand which components undergo maintenance. A comprehensive review of available reliability data is given in [32] and [31]. Some publications provide KPIs for all wind turbine subsystems (see e.g. [10], [12] and [29]), while others focus on specific subsystem such as the power converter (see e.g. [36], [104]), the pitch system (e.g. [35], [105]) or the main bearing (e.g. [106]). Direct comparisons between such studies can be challenging due to the different categorisation systems and variations in provided KPI definitions. For instance, Carroll et al. reported reliability figures from about 350 offshore wind turbines, identifying the "pitch / hydraulics" subsystem as the most frequently failing one [12]. In contrast, the System Performance, Availability and Reliability Trend Analysis (SPARTA) initiative noted the blade adjustment system as having the second highest monthly repair rate based on an analysis of 1045 offshore wind turbines located in UK waters [29]. Given that Carroll et al. use annual

[3] Julia Walgern, Katharina Beckh, Neele Hannes, Martin Horn, Marc-Alexander Lutz, Katharina Fischer, Athanasios Kolios, 2024. „Impact of Using Text Classifiers for Standardising Maintenance Data of Wind Turbines on Reliability Calculations". IET Renewable Power Generation, 18(15), 3463-3479, https://doi.org/10.1049/rpg2.13151 [198]

failure rates while SPARTA uses monthly repair rates, the studies base their findings on different KPI definitions. Moreover, the components included in the defined subsystems likely differ, complicating KPI comparisons. Anderson et al. highlight that even the definition of "failure" can vary in field-data based studies, impacting the KPI values [30]. One solution for uniformly defining subsystems and components of wind turbines is to employ standards and guidelines like the reference designation system RDS-PP for wind turbines [40] or RDS-PS [70]. Regrettably, these standards have not gained wide acceptance in the wind energy industry yet. Instead, many proprietary classification systems are in use which do not easily translate to the mentioned standards.

Most of the reliability studies mentioned above rely on data recorded before 2015. The extensive manual effort required to preprocess maintenance information often results in significant delays between documenting site visits and publishing data-based findings.

On the one hand, [107] developed a digitalisation workflow to standardise wind turbine maintenance information. This process involves optical character recognition, information extraction and text classification. After reviewing various classifier methods, they employed support vector machine (SVM) approach to train and test a text classifier to label service reports with RDS-PP components. As RDS-PP is organised in a hierarchical structure (cf. [107]), classification results for different levels have been presented and compared using F1 scores. While initial classification results displayed promising micro F1 scores, these were not explored further for productive application. Notably, training text classifiers for isolated RDS-PP levels is of limited practical relevance, as only the combined levels provide insights into the affected components and subsystems.

On the other hand, [108] analysed three different methods of labelling service reports, namely expert labelling, text classification and AI-assisted tagging in combination with a rule-based approach, to differentiate between predictive and corrective maintenance work orders. Afterwards, failure rates for the overall wind turbine system derived from the differently preprocessed datasets – making use of the simple categorisation of predictive and corrective activities – were compared. Results show that the AI-assisted tagging approach reduces data preparation time significantly, however, calculated KPIs are not reliable [108]. These findings are based on data of a single wind farm.

In contrast, this chapter addresses the challenge of efficiently digitalising wind turbine maintenance data, traditionally hindered by non-standardised formats requiring manual expert intervention. This study investigates the efficacy of various text classifiers in automating the categorisation process of maintenance data against conventional manual labelling. The novelty of this research lies in the comprehensive evaluation of different text classifiers trained on diverse datasets and their impact on reliability KPI calculations. By comparing manual labelling with classifier-processed data, this chapter reveals the uncertainties in both methods and suggests improvements in maintenance reporting and designation systems to achieve more accurate KPIs.

Within this chapter, different text classifiers, which classify the text descriptions of wind turbines' maintenance measures into RDS-PP categories, and thus different wind turbine components and subsystems, are analysed and implications for real-world application are assessed. Uncertainty resulting from using text classifiers based on different training datasets varying in size and homogeneity is evaluated. Additionally, it is analysed how reliability KPIs differ depending on the chosen preprocessing method comparing manual labelling against different text classifier results. Our findings offer recommendations for digitising wind turbine maintenance reports, making this study invaluable for researchers and practitioners processing text-based service reports to derive reliability figures or understand spare parts usage.

The chapter is structured as follows: First, the state-of-the-art literature on text classification is discussed (Section 6.2). Second, an introduction of the used methods for classification but also for comparing classification results is given and the analysed dataset is described (Section 6.3). Afterwards, several text classifiers based on different training datasets are evaluated and compared and industry perspectives are presented based on conducted interviews. Next to the classifier performance itself, the impact on reliability KPI calculation and corresponding uncertainties are analysed and barriers to the adoption of text classifiers in the wind energy sector discussed (Section 6.4). Last, main conclusions are summarised and an outlook to future work is given (Section 6.5).

6.2 State of the art literature on text classification

Text classification, a fundamental task in natural language processing (NLP), involves assigning predefined categories to text data ([109], [110]). Over the past few decades, this field has witnessed significant advancements, driven by the evolution of machine learning and deep learning techniques. This literature review highlights the key developments and state-of-the-art approaches in text classification.

The initial methods for text classification relied heavily on traditional machine learning algorithms such as Naive Bayes, k-nearest neighbours (k-NN), and support vector machines (SVM). These algorithms typically used bag-of-words or term frequency-inverse document frequency (TF-IDF) representations of text. For instance, [95] demonstrated the effectiveness of SVMs for text categorisation, showing superior performance compared to other methods at the time due to its ability to handle high-dimensional data, while recently, [111] have shown the impactful application of such methods for fault diagnosis for control of critical infrastructure.

To improve classification performance, extensive feature engineering was employed. Techniques like n-grams, part-of-speech tagging, and named entity recognition were used to extract meaningful features from text ([112], [113]). Ensemble methods, which combine multiple classifiers, were also explored [114]. The Random Forest algorithm, an ensemble of decision trees, proved effective for various text classification tasks due to its robustness and ability to handle large feature spaces [115].

The advent of deep learning marked a significant shift in text classification. Neural networks, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrated remarkable capabilities in capturing complex patterns in text data ([116], [117]). [118] introduced a CNN model for sentence classification that outperformed traditional methods by leveraging pre-trained word embeddings and convolutional filters to capture local dependencies in text.

RNNs, especially long short-term memory (LSTM) networks [119] were effective in handling sequential data, making them suitable for text classification tasks. LSTMs addressed the vanishing gradient problem, enabling the capture of long-range dependencies [120]. This made them particularly useful for tasks like sentiment analysis and document classification, where context plays a crucial role.

The introduction of attention mechanisms further revolutionised text classification. Attention allows models to focus on relevant parts of the input sequence, enhancing their ability to capture context [121]. The Transformer model, introduced by [122] utilised self-attention mechanisms to process entire sequences simultaneously, rather than sequentially as in RNNs. This innovation led to significant improvements in both training efficiency and classification performance.

BERT (Bidirectional Encoder Representations from Transformers), developed by [98], built on the transformer architecture and introduced bidirectional context understanding. BERT achieved state-of-the-art results on various NLP benchmarks by pre-training on a large corpus and fine-tuning on specific tasks. Its ability to understand context in both directions of a text sequence made it particularly powerful for text classification [123].

Recent developments in large language models, such as GPT-3 [124] and GPT-4, have further advanced text classification. These models, with billions of parameters, are pre-trained on diverse datasets and can be fine-tuned for specific tasks with minimal additional training [125]. Their deep contextual understanding and ability to generate coherent text have set new benchmarks in text classification performance.

Transfer learning, where pre-trained models are fine-tuned on specific tasks, has become a dominant approach in text classification. Models like BERT, RoBERTa [126], and T5 [127] exemplify this trend. Fine-tuning these models on domain-specific data leads to substantial improvements in performance, as they leverage the rich knowledge gained during pre-training.

The integration of text with other modalities, such as images and audio, has opened new avenues for text classification [128]. Multimodal models that combine textual and visual data are being explored for tasks like social media analysis and sentiment classification [129]. Hybrid approaches that combine rule-based systems with machine learning are also gaining traction, offering a balance between interpretability and performance.

Text classification has found applications across various domains, including sentiment analysis, spam detection, topic labelling, and more [130]. The ongoing research focuses on improving model interpretability, handling low-resource languages, and reducing biases in text classification models [131]. Future directions include the development of more efficient models that require less computational power and the exploration of unsupervised and semi-supervised learning techniques to leverage unlabelled data.

In addition to text classifiers, the use of multimodal knowledge graph (KG) databases presents a promising approach for managing maintenance data of wind turbines. Knowledge graphs integrate heterogeneous data sources, including structured data (sensor readings, operational logs) and unstructured data (maintenance reports, technical manuals), enabling a holistic representation of information [132]. KGs can enhance data interoperability, facilitate advanced analytics, and improve decision-making processes by connecting related entities and capturing complex relationships [133]. For instance, in the healthcare domain, KGs have been used to integrate clinical data and literature, aiding in diagnosis and treatment planning [134]. Similarly, in wind turbine maintenance, a KG could unify data from various sources, providing a comprehensive view of turbine health and maintenance needs. A comparative analysis of text classifiers and KGs could reveal synergies, such as using classifiers to populate KGs, ultimately improving data utilisation and operational efficiency in wind energy systems.

This study advances the state of the art by integrating advanced text classification techniques with domain-specific fine-tuning to automate the categorisation of wind turbine maintenance data, a task traditionally requiring extensive manual effort. Unlike previous methods, our approach leverages the hierarchical structure of RDS-PP for precise component-level categorisation, enhancing the reliability and accuracy of maintenance logs. Additionally, by comparing classifier performance across diverse datasets and exploring the integration of large language models, we address scalability and adaptability challenges, providing a robust framework for standardising maintenance data and improving operational efficiency in the wind energy sector.

6.3 Methodology and datasets

The methodology for this study involves a structured approach with distinct steps to ensure clarity and reproducibility (see **Figure 17**). The text classifiers are built using maintenance reports collected from wind turbines, which are initially available in various formats including text files and PDFs. The research employs natural language processing (NLP) to automate the categorisation of wind turbine maintenance logs, enhancing reliability assessments. Initially, maintenance reports are digitised using optical character recognition (OCR), followed by text preprocessing to standardise formats. Term frequency-inverse document frequency (TF-IDF) vectorisation converts text data into numerical features for model training. Support vector machine (SVM) classifiers, optimised with Platt scaling for probabilistic outputs, are trained on manually labelled data adhering to the Reference Designation System for Power Plants (RDS-PP). The classifiers undergo rigorous evaluation through five-times four-fold cross-validation, ensuring robust performance metrics, including precision, recall, and F1 scores. The classifiers then categorise maintenance activities, facilitating the calculation of failure rates and other key performance indicators (KPIs). Comparative analyses between classifier and manually derived KPIs reveal the models' efficacy. Industry feedback is incorporated to tailor classifier configurations, aiming for seamless integration into maintenance workflows, thus improving data-driven decision-making in wind turbine operations.

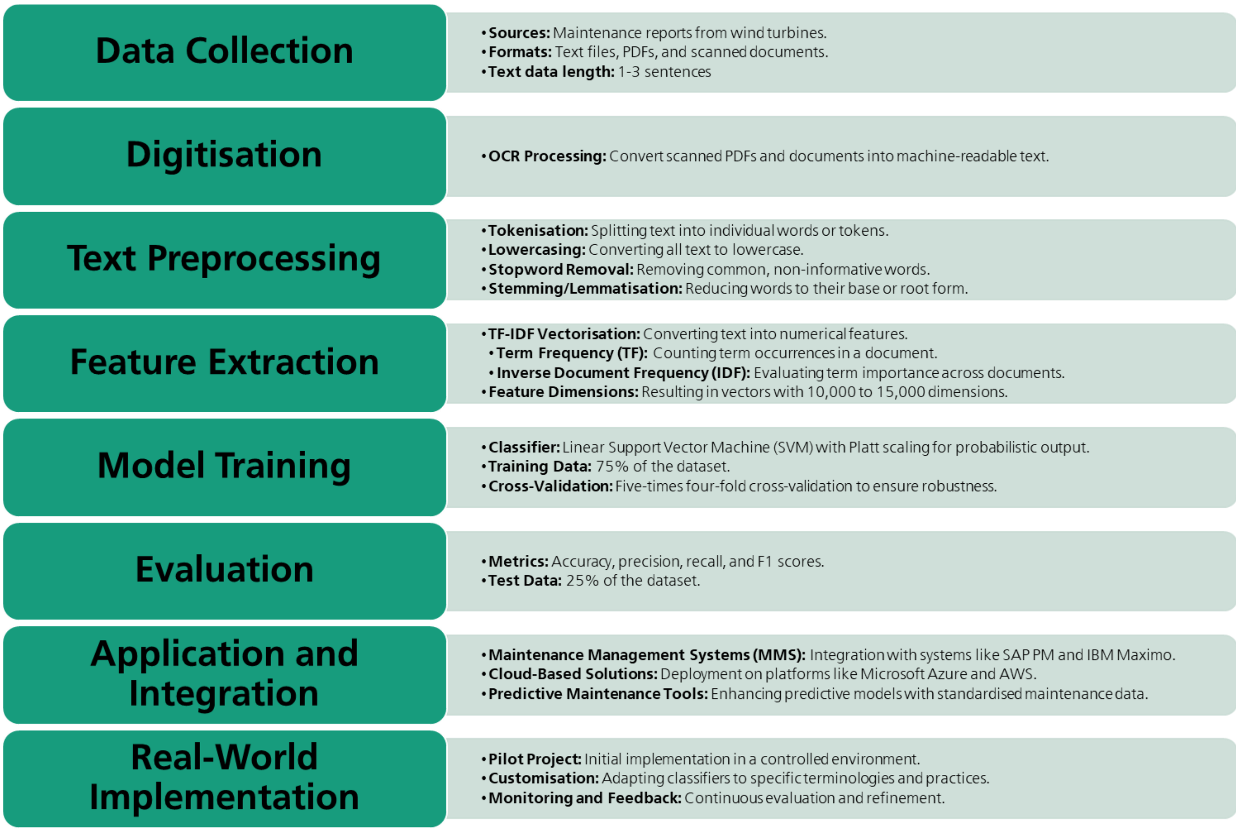

Figure 17. Workflow of the methodological framework

6.3.1 Datasets

The datasets employed in this analysis originate from 15,000 maintenance reports spanning 342 wind turbines, both onshore and offshore. These reports represent approximately 800 operational turbine years and are sources from two distinct operators. As respective wind turbines are located in different countries and maintenance is performed by different companies, the reports are generally available in English but in rare examples German and French language is used as well. Additionally, different turbine types are included in the analysis which naturally leads to terminology variance. This is further exacerbated by the documentation from different companies. Despite these variations, the reports generally follow a standard structure that includes crucial sections and details necessary for accurate classification.

- Header Information: The header typically contains metadata about the maintenance intervention, such as the date, time, wind turbine identifier, and the technicians' names. This section provides contextual information but is often not directly used for text classification.
- Summary of Maintenance Activity: This section briefly overviews the maintenance activity performed. It may include a high-level description of the issue addressed and the actions taken. For example, a summary might state, "Replaced faulty pitch motor in turbine T123."
- Detailed Description: The detailed description is the core of the maintenance report. It includes a step-by-step account of the maintenance process, components, tools and materials, and any observations or measurements taken. This section can

vary in length from a few sentences to several paragraphs, depending on the complexity of the task and the organisation. Detailed descriptions are crucial for text classifiers as they contain the technical terms and context needed for accurate categorisation.

- Parts and Materials Used: This section lists all the parts and materials used during maintenance. It typically includes part numbers, quantities, and sometimes supplier information. This structured data can be cross-referenced with the textual descriptions to enhance classification accuracy.
- Recommendations and Next Steps: Maintenance reports often conclude with recommendations for future actions or follow-up maintenance tasks. This section may also note any potential issues that need monitoring. Although this information is valuable for ongoing maintenance planning, it is secondary for the initial classification of the report.
- Signatures and Approvals: The report may include signatures from the technician and supervisory personnel, indicating that the maintenance activity has been reviewed and approved. This section is typically irrelevant for text classification but ensures the report's validity and compliance.

For every maintenance activity, the relevant components are categorised manually using the reference designation system RDS-PP for wind turbines [40]. For this, mainly the summary of maintenance activity and detailed description was utilised as other information categories were not available within all maintenance reports. This labelled dataset forms the foundation for training and testing text classifiers.

6.3.2 Methodology

6.3.2.1 Text classification and corresponding metrics

The chosen classification method is support vector machines (SVM). More specifically, a linear support vector classification [135] is used and a probabilistic output is achieved with Platt scaling [136]. Text data is transformed with TF-IDF vectorisation method as outlined in [137], [138], and [135]). This straightforward approach is inspired by [139] who contrasted the language model BERT [98] and linear SVM [140] for text classification tasks, emphasising the trade-off between enhanced performance and computational expense.

For this study, various text classifiers were implemented to classify maintenance measure descriptors into RDS-PP categories. In this realm of machine learning, such label categorisation is termed as "predictions". These text classifiers do not differ in their classification method but in their training dataset and the detail of the RDS-PP predictions. A comprehensive overview of all scenarios, including training and test set specifications, is provided in **Table 9**.

The scenario description is defined as follows: First, the dataset the classifier is trained on is identified, whether it covers turbines of different original equipment manufacturers (OEMs) or solely from one OEM. Second, the type of predictions made by the classifier is specified. Scenarios 1 to 22 pertain to the varied hierarchy levels within RDS-PP, addressing all subsystems and component categories defined by RDS-PP. Higher levels offer more intricate component description. Scenarios 23 to 26 only focus on a subset of the principle subsystems that fail most frequently, with all other subsystems consolidated into an "other" category. Third, the proportion of training data used (expressed as a percentage) is specified in the scenario description whenever scenarios are differentiated by the amount of training data.

All scenarios utilise single-label classifiers, excluding maintenance reports that have multi label cases. This exclusion ensures a clear assignment of component categories during classifier training. Hence, the filtered dataset contained only reports documenting a single component category. In each scenario, the first step is the selection of the respective subset from the whole dataset, e.g., in scenario 1 the relevant subset is the maintenance reports from operator 1. For the experiments, four-fold cross-validation was used resulting in splits of 75% training data and 25% test data. To provide a consistent comparison of different classifiers, the training and test set sizes remain constant across respective scenarios, ensuring only one variable is evaluated simultaneously. In scenarios 7-10, which investigate the effect of training size, the data is sampled from the training data according to the indicated fraction. Scenarios 1 to 10 and 23 to 26 can be directly compared, while scenarios 11 to 16 and 17 to 22 should be examined independently. These scenarios aid in contrasting the efficacy of multiple classifiers and assessing the influence of the training set size and homogeneity.

Table 9. Overview of all presented classifier test scenarios. Scenario description includes: (1) training dataset composition (single vs. multiple OEMs), (2) prediction type (RDS-PP hierarchy levels or focus on key subsystems), and (3) proportion of training data used.

Scenario	Scenario description	Training set	Test set
1	Operator1_up_to_level2_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
2	Operator1_up_to_level3_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
3	Operator1_up_to_level4_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
4	Operator2_up_to_level2_100%	1787 maintenance reports of operator 2	595 maintenance reports of operator 1
5	Operator2_up_to_level3_100%	1787 maintenance reports of operator 2	595 maintenance reports of operator 1
6	Operator2_up_to_level4_100%	1787 maintenance reports of operator 2	595 maintenance reports of operator 1
7	Operator1_up_to_level2_10%	179 maintenance reports of operator 1	595 maintenance reports of operator 1
8	Operator1_up_to_level2_25%	447 maintenance reports of operator 1	595 maintenance reports of operator 1
9	Operator1_up_to_level2_50%	894 maintenance reports of operator 1	595 maintenance reports of operator 1
10	Operator1_up_to_level2_75%	1340 maintenance reports of operator 1	595 maintenance reports of operator 1
11	OEM1_up_to_level2	1148 maintenance reports of operator 1 only including OEM1 reports	383 maintenance reports of operator 1 only including OEM1 reports

12	OEM1_up_to_level3	1148 maintenance reports of operator 1 only including OEM1 reports	383 maintenance reports of operator 1 only including OEM1 reports
13	OEM1_up_to_level4	1148 maintenance reports of operator 1 only including OEM1 reports	383 maintenance reports of operator 1 only including OEM1 reports
14	Operator1_up_to_level2	1148 maintenance reports of operator 1	383 maintenance reports of operator 1 only including OEM1 reports
15	Operator1_up_to_level3	1148 maintenance reports of operator 1	383 maintenance reports of operator 1 only including OEM1 reports
16	Operator1_up_to_level4	1148 maintenance reports of operator 1	383 maintenance reports of operator 1 only including OEM1 reports
17	OEM2_up_to_level2	638 maintenance reports of operator 1 only including OEM2 reports	213 maintenance reports of operator 1 only including OEM2 reports
18	OEM2_up_to_level3	638 maintenance reports of operator 1 only including OEM2 reports	213 maintenance reports of operator 1 only including OEM2 reports
19	OEM2_up_to_level4	638 maintenance reports of operator 1 only including OEM2 reports	213 maintenance reports of operator 1 only including OEM2 reports
20	Operator1_up_to_level2	638 maintenance reports of operator 1	213 maintenance reports of operator 1 only including OEM2 reports
21	Operator1_up_to_level3	638 maintenance reports of operator 1	213 maintenance reports of operator 1 only including OEM2 reports
22	Operator1_up_to_level4	638 maintenance reports of operator 1	213 maintenance reports of operator 1 only including OEM2 reports
23	Operator1_7categories_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
24	Operator1_5categories_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
25	Operator1_4categories_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1
26	Operator1_3categories_100%	1787 maintenance reports of operator 1	595 maintenance reports of operator 1

To evaluate the performance of the text classifiers, F1 scores are utilised. These scores are derived from precision and recall [135]:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{6.1}$$

The F1 score can range between 0 and 1. An F1 score of 1 indicates perfect precision and recall. Distinctions have been made between micro and macro F1 scores. Micro F1 scores account for label imbalance, while macro F1 scores are determined using unweighted means for each label [135]. For multi-class classification, where each sample has only one valid classification result, micro F1 scores are equivalent to accuracy measures. In order to conduct a statical analysis, a five-times four-fold cross-validation is employed, repeating the classifier evaluation 20 times with text classifiers trained on randomly sorted training and test datasets. Subsequently, mean values for macro and micro F1 scores are computed and compared for different test scenarios. Boxplot analysis was conducted to measure the variance of the computed metrics due to the dataset size and imbalance. The manually labelled maintenance reports serve as ground truth.

While F1 score, which harmonises precision and recall, is widely used, additional metrics like precision-recall (PR) curves and area under the receiver operating characteristic curve (AUC-ROC) offer nuanced insights.

- Precision-Recall Curves: PR curves are particularly valuable for imbalanced datasets, highlighting the trade-off between precision and recall across different thresholds. They provide a clearer picture of classifier performance when positive classes are rare [141].
- AUC-ROC: AUC-ROC evaluates the overall ability of the model to discriminate between classes, plotting true positive rate against false positive rate. It is robust to class imbalance and offers a single scalar value summarising performance across all thresholds [142].
- Recent Advances: Recent studies advocate combining these metrics for a more holistic evaluation. For instance, [143] emphasize the importance of using multiple metrics to avoid misleading conclusions in model performance evaluation.

The classification into component labels is based on the hierarchical structure of RDS-PP; meaning it organises information at multiple levels, with each subsequent level providing more detailed specifications. The codes represent broad categories of components or systems at the highest levels, while the lower levels refer to specific parts and their functionalities. This hierarchical structure allows for detailed, systematic categorisation that facilitates better management and analysis of maintenance data.

RDS-PP codes are alphanumeric combinations where each code segment conveys specific information about the component's function, type and location within the overall system. For instance, a code like "MSE10 KF001" is a precise identifier: "MSE" denotes the converter system, "10" specifies an overall subsystem within the converter, and "KF001" indicates a particular control system component within that subsystem. This structured naming approach ensures that every part of a wind turbine is uniquely and consistently identified. Each RDS-PP code segment builds upon the previous one, offering increasing detail.

Evaluations are split between "fully correct" labels, where the text classifier's prediction matches the manually preprocessed label, and "soft correct" labels. The latter occurs when a text classifier predicts a label higher up in the RDS-PP hierarchy compared to the manual label, therefore, not being wrong but more generic than possible. **Table 10** offers three illustrative examples: "MSE10 KF001" represents "Control System Converter System Overall", whereas "MSE10" is a broader descriptor of the "Converter System Overall". A false prediction is exemplified where "MDA11" signifies "Rotor Blade System 1". Given this dual evaluation approach, each F1 score (macro and micro) is calculated for both "fully correct" and "soft correct" evaluation, respectively.

Table 10. Example evaluation of text classifiers' prediction into RDS-PP labels

Maintenance description	Model prediction	True label	Evaluation
Converter control board exchanged	MSE10 KF001	MSE10 KF001	fully correct
Converter control board exchanged	MSE10	MSE10 KF001	soft correct
Converter control board exchanged	MDA11	MSE10 KF001	false

6.3.2.2 Failure rate calculation

In the subsequent phase of this study, average failure rates per wind turbine per year are calculated using differently preprocessed datasets, aiming to quantify their impact on reliability KPIs. The average failure rate f of a specific subsystem is expressed as the ratio of the sum of all failures N of that subsystem over a given time frame to the total number of operational wind turbine years observed within this period T:

$$f = \frac{\sum_{i=1}^{I} N_i}{\sum_{i=1}^{I} X_i T_i} = \frac{N}{T} \tag{6.2}$$

Herein, N_i denotes the number of failures of the analysed subsystem in the time interval i, X_i represents the count of wind turbines examined during this interval and T_i is the span of the time interval.

Contrasting with the initial segment of uncertainty analysis, wherein random maintenance reports were selected to establish the training and test datasets, this phase uses continuous maintenance reports series in chronological order as test sets to ensure the derived KPIs are meaningful.

6.4 Results and discussion

6.4.1 Performance comparison of text classifiers based on different models

At first, preliminary experiments with a SVM, a CNN [118] with pre-trained word embeddings and a fine-tuned Transformer variant XLM-RoBERTa [144] were performed. A requirement for the experiments was to have access to the model which excluded proprietary models such as ChatGPT. In addition, the utility of open-source models for classification tasks when dealing with technical language is so far lacking [145]. The experiments were performed with an 80-20 train-test split and labels on the most precise level. **Table 11** reports the performance of each model and shows that the linear SVM outperformed the other models. Moreover, traditional methods like TF-IDF and SVM are less computationally intensive and can be more cost-effective for specific datasets and tasks. For instance, in our scenarios where the volume of text data is manageable and the complexity of the language is not exceedingly high, these methods can provide competitive performance with significantly lower computational overhead. Therefore, all further experiments were conducted following the SVM approach.

Table 11. Comparison of three model architectures, SVM, CNN and XLM-RoBERTa

	Linear SVM	CNN	XLM-RoBERTa
Macro F1	0.41	0.34	0.34
Accuracy	0.71	0.67	0.68

6.4.2 Performance comparison of text classifiers trained on different datasets

6.4.2.1 How well does a classifier perform for different levels of detail?

Figure 18 presents the F1 scores for test scenarios 1, 2 and 3, comparing text classifiers. Although all are trained and tested on datasets of the same size, they predict components with varying degrees of detail. An up-to-level-4 classifier can precisely predict a label when the maintenance description provides ample information. However, if the maintenance report is not detailed, it might opt for more generalised labels from RDS-PP hierarchy levels two and three. In comparison, an up-to-level-2 classifier always predicts the broader subsystem, corresponding to RDS-PP hierarchy level 2, even when the maintenance reports are more informative.

The findings indicate that as the granularity of the target label increases, classifier performance decreases. This outcome is intuitive, as the more nuanced predictions classifiers can make, they are faced with a greater variety of potential classification categories, intensifying the challenge. A subsequent boxplot analysis revealed a minor fluctuation in in micro F1 scores. In contrast, macro F1 scores experienced more significant variability. This disparity might be attributed to classifications of component categories that are infrequent in a dataset. Dependent on the train-test division, these rare categories might be predicted less accurately, which is more reflected in the macro F1 scores than in the micro F1 scores.

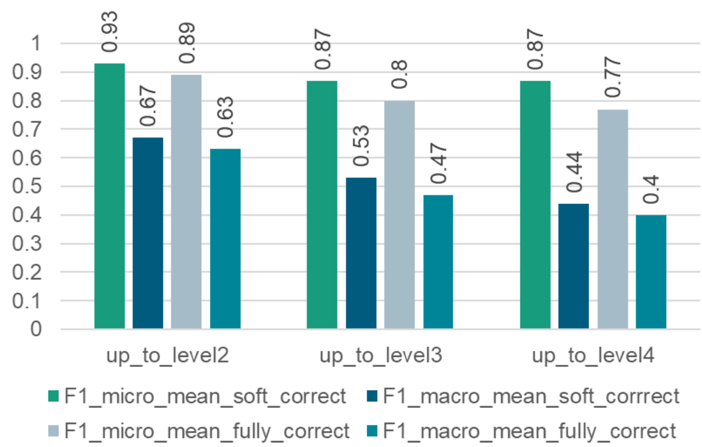

Figure 18. Comparison of F1 scores for test scenarios 1 to 3

6.4.2.2 How many data points are required for training to see sufficient classification results?

Manual labelling of service reports is a meticulous and time-intensive task. Given this, the study assessed the efficacy of classifiers trained on smaller datasets. **Figure 19** delineates results for test scenarios 7 to 10. These scenarios involve classifiers trained on 75%, 50%, 25% and 10% of the original training dataset size of test scenario 1, respectively.

A general trend is apparent: as the size of the training dataset reduces, the attainable F1 scores deteriorate. Nevertheless, a comparative assessment between scenario 1 (using the original training dataset size) with scenario 10 (utilising 75% of original training dataset size) reveals only marginal performance declines. Remarkably, even with a substantially truncated training dataset, as in scenario 7 (10% of original training dataset size), micro F1 scores are still quite competitive. In contrast, macro F1 scores decrease substantially when decreasing the training dataset size. Depending on the focus of analysis, the laborious manual effort expended on labelling to devise a training dataset could be considerably reduced if the post-labelling analysis is principally concerned with frequently occurring components or subsystems.

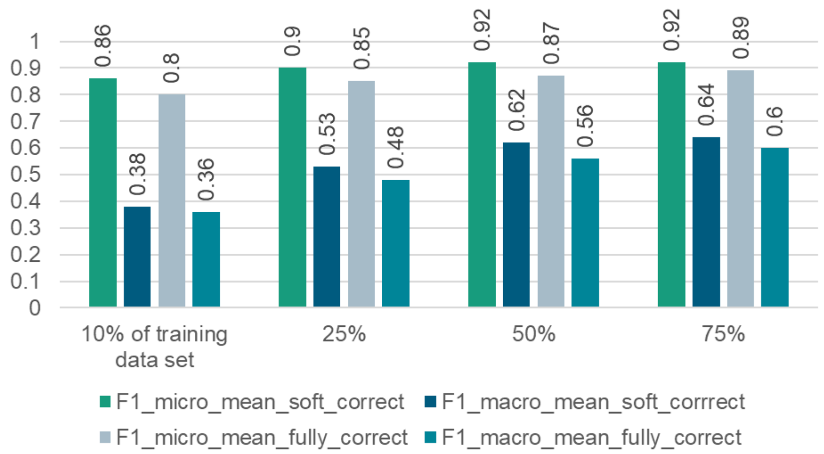

Figure 19. Comparison of F1 scores for test scenarios 7 to 10

From a machine-learning standpoint, these findings are somewhat unexpected. Conventionally, one would anticipate significantly enhanced classifier performance with more extensive training datasets. One plausible explanation for the reduced impact of training dataset size on results might be inconsistencies inherent within the training dataset. Such inconsistencies can diminish the advantages offered by larger datasets.

Examining the manual labelling process that employs RDS-PP for component categorisation supports this hypothesis. Classifying based on RDS-PP is not always straightforward for specific wind turbine components. Challenges arise due to ambiguities in distinguishing between different technical setups within RDS-PP. Moreover, RDS-PP guidelines might lack explicit definitions for certain component categories. This ambiguity leaves it to experts to decide the most fitting category, potentially leading to inconsistencies in the labelling process.

6.4.2.3 Does the classification result improve when the training set is more specific?

The results of previously presented test scenarios suggest that larger training datasets do not offer significant advantages in terms of text classifier performance. Consequently, scenarios 11 to 22 were established to assess whether the classification results enhance when the training datasets are more tailored. From an engineering standpoint, the available service reports were scrutinised for pronounced differences. A significant observation was the variance in component naming conventions across different OEMs. As such, test scenarios 11 to 13 and 17 to 19 trained specific text classifiers based on data of two distinct OEMs. To draw a comparison with the comprehensive operator classifiers without confounding various factors, scenarios 14 to 16 and 20 to 22 used classifiers trained on randomly selected data points from operator 1, comprising OEM1 and OEM2 data. However, the training dataset size was consistent with the OEM-specific scenarios. The performance evaluation of these classifiers was executed on OEM-centric data.

The outcomes of some of these test scenarios are illustrated in **Figure 20**. When comparing F1 scores for up-to-level-2 (scenario 11) and up-to-level-3 (scenario 12) predictions for OEM1 with predictions from classifiers trained on the more generic operator 1 data (scenarios 14 and 15, respectively), the classifiers using the more tailored data exhibit marginally better results across all F1 metrics, albeit the differences are not significant. Similar patterns emerged for up-to-level-4 classifiers (scenarios 13 and 16) and classifiers tested on OEM2 data (scenarios 17 to 22). It is noteworthy, known from manual labelling, that variations in terminology exist for identical component categories, contingent on the OEM. These discrepancies can be attributed to factors such as distinct component suppliers, wind turbine manuals and O&M procedure description semantics. Service technicians, influenced by these factors, employ varied terminology in describing and documenting their tasks. As a result, it might be advantageous to utilise smaller, yet more specific datasets for training text classifiers, especially when the goal is categorising components of a specific technology group, like a particular OEM.

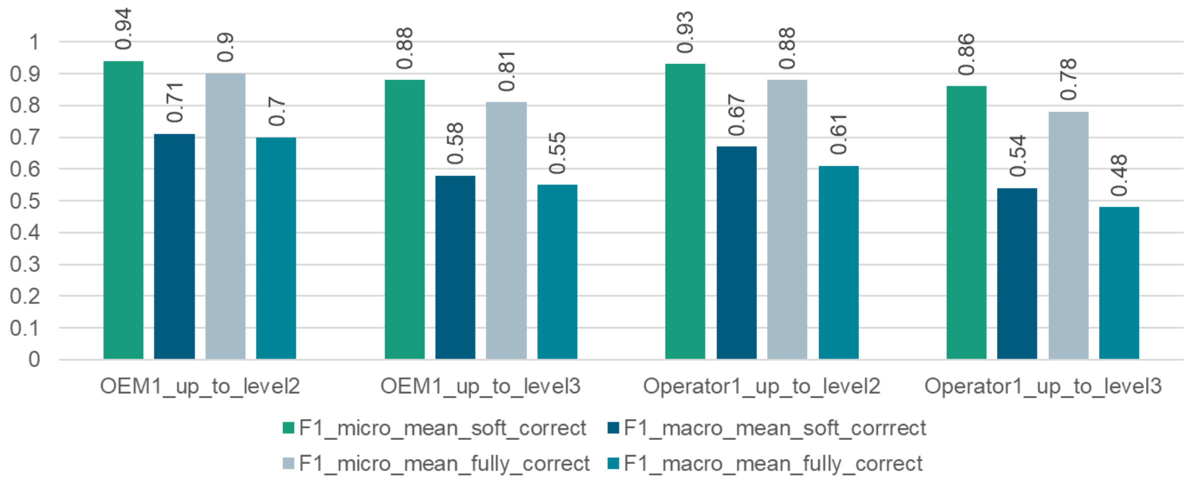

Figure 20. Comparison of F1 scores for test scenarios 11, 12, 14 and 15

6.4.2.4 How much does the classification result improve when less label categories need to be predicted?

Depending on the information that needs to be extracted from the labelled maintenance reports and the focal points of the data analysis, varying levels of detail in the labelling process become necessary. For analyses targeting the subsystem level, the RDS-PP level 2 would be sufficient. Often, operators are mainly concerned with the most critical subsystems, which typically correspond to the most frequently failing ones. In such instances, classifiers were trained to distinguish only among these predominant subsystems, grouping all other subsystems in the "other" category. This approach was pursued to investigate the potential enhancement in classification outcomes when fewer label categories are required for predictions.

An overview of the most frequently failing subsystems is provided in **Table 12** together with the RDS-PP codes. If we were to consider the subsystems based on their frequency of mention in the maintenance reports, the drive train system swaps places with the control system.

Table 12. Most frequently failing subsystems within the analysed dataset

	Subsystem	Corresponding RDS-PP code
1	Rotor system (incl. pitch system)	MDA
2	Converter system	MSE
3	Control system	MDY
4	Drive train system (incl. main bearing and gearbox)	MDK
5	Power generation system (incl. generator)	MKA
6	Yaw system	MDL

Based on these findings, the text classifier in test scenario 23 is configured to predict the six subsystems highlighted in **Table 12**, reverting to the label "other" for instances where none of the specific categories are applicable. This configuration yields a total of seven distinct categories. In contrast, text classifiers from test scenarios 24 to 26 increasingly relegate more subsystems under the "other" category. For example, in test scenario 26, the classifier discerns among the three different categories: MDA, MSE or "other".

The corresponding F1 scores of these classifiers are shown in **Figure 21**. As these test scenarios employ predefined categories rather than a hierarchical system, distinctions between "soft correct" and "fully correct" evaluations become redundant. In these cases, predictions are either entirely accurate or mistaken, which means only two F1 scores per classifier are depicted in **Figure 21**. The results indicate that narrowing the focus to frequent labels, while reducing the overall number of labels, enhances classification performance. In comparison to the results of the up-to-level-2 classifier of test scenario 1, which registered a micro F1 score of 0.89 and macro F1 score of 0.63, scenarios 23 to 26 exhibit superior evaluation metrics. This convergence becomes more pronounced as the number of prediction categories diminishes, mitigating label imbalance and diminishing the effects of weighting.

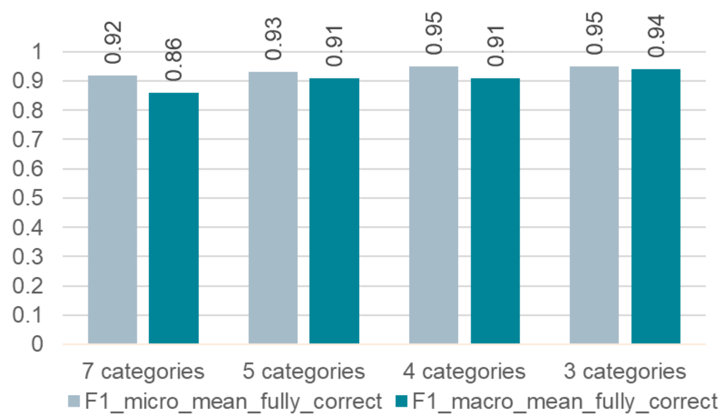

Figure 21. Comparison of F1 scores for test scenarios 23 to 26

6.4.2.5 How well do the classifiers perform for different wind farms?

Until now, all discussed results were derived from classifiers trained and tested on data from the same wind farm or collective group of wind farms. Given the encouraging outcomes for practical application, the study sought to understand how these classifiers perform if applied on datasets from different wind farms.

In scenarios 4, 5 and 6 classifiers were trained with the same level of detail as in scenario 1, 2 and 3. However, training was undertaken using data from a distinct operator. These scenarios provided a preliminary insight into the adaptability if pre-trained classifiers are transferred to alternative datasets. F1 scores from these scenarios were only about half or two-thirds of those achieved in scenarios 1 to 3, as depicted in **Figure 18**. A plausible explanation for this could be variations in terminology and report structures across different organisations. These outcomes suggest that tailoring a classifier to each operator seems necessary, albeit being more labour-intensive due to the need for manual data labelling.

Subsequently, classifiers from test scenarios 1, 2 and 23 were assessed on a comprehensive dataset from wind farms not used during classifier training. These wind farms, however, were still under the portfolio of the original operator. **Figure 22** presents the results, which clearly indicate a suboptimal performance relative to prior scenarios. Specifically, the up-to-level-2 and up-to-level-3 classifiers correctly predict only around half of the labels. Even though the up-to-level-2 classifier's performance mirrored that in test scenario 4, the up-to-level-3 classifier showed improved accuracy than in test scenario 5. This suggests possible similarities in terminology at the subsystem level, whereas greater disparities exist at the component level. When comparing these outcomes with results from test scenarios 1 and 2, it is evident that there is a significant decline in performance when classifiers are extended to different wind farms – even if they belong to the same operator. Such inconsistencies might arise from divergent documentation standards across service entities operating in varied regions. Notably, the 7-categories classifier managed to correctly predict 74% of the maintenance activities, outperforming the other two classifiers. However, in comparison with test scenario 23 in which 92% of labels were predicted correctly, it is clear that there is room for improvement.

Therefore, it has to be noted that when an existing trained classifier is to be applied to other wind farms the semantic needs to be analysed carefully. If significant differences emerge, investing in retraining the classifier can be beneficial—even if it necessitates additional manual efforts for curating a training dataset.

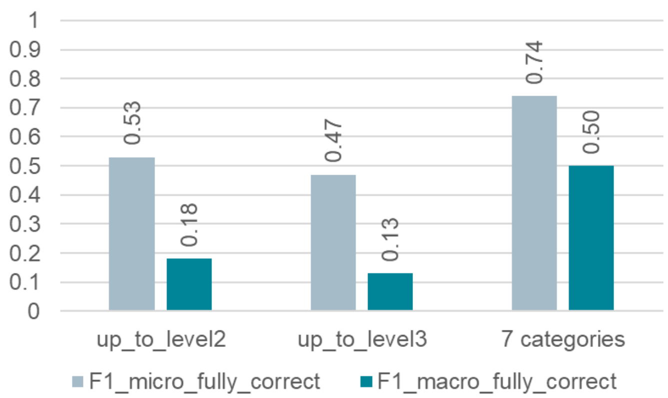

Figure 22. Classifier performance for different wind farms belonging to the portfolio of the same operator

6.4.3 Industry perspective on productive use of classifiers

To understand industry needs and pinpoint the most valuable classifiers, six structured face-to-face interviews were conducted with senior staff from different operators and service providers being active in asset management. These discussions were twofold: Firstly, to determine the requisite level of detail when labelling maintenance reports based on internal processes and secondly, to gauge preferences for the level of detail of classifiers considering the F1 scores achieved in the above test scenarios. The approach involved presenting interviewees with a series of either-or questions to discern their priorities. A visualisation of evaluated answers can be found in **Figure 23**.

From the data in this figure, it is evident that there are no clear tendencies among interviewees. Responses varied significantly across interviewees and organisations. Some showed a leaning towards a more generic classifier (up to level 2) with a higher performance, while others exhibited a bias for a more specific classifier (up to level 4) even if it came with slightly lower performance scores. Both these preferences were equally popular.

Another aspect explored, was whether the interviewees would opt for (a) a high-performing classifier, trained on a comprehensive dataset, even if it necessitates more labour-intensive and costly data preparation, or (b) a classifier with optimised performance only for regularly mentioned components within the maintenance reports as it is trained on a smaller dataset. Here, there was a discernible tilt towards the former – a preference for larger datasets, even if they required increased effort.

Lastly, interviewees were asked about their interest in the classifiers from test scenarios 1, 23, 24, 25, and 26. These represented classifiers with approximately 25, seven, five, four and three distinct component categories, respectively. Half of the respondents favoured the up-to-level-2 classifier. The other half expressed a preference for the classifier, which labels the six most frequently failing subsystems, relegating all other activities under the "other" category. Notably, none of the interviewees expressed an interest in the classifiers of test scenarios 24 to 26, which labelled fewer than six subsystems.

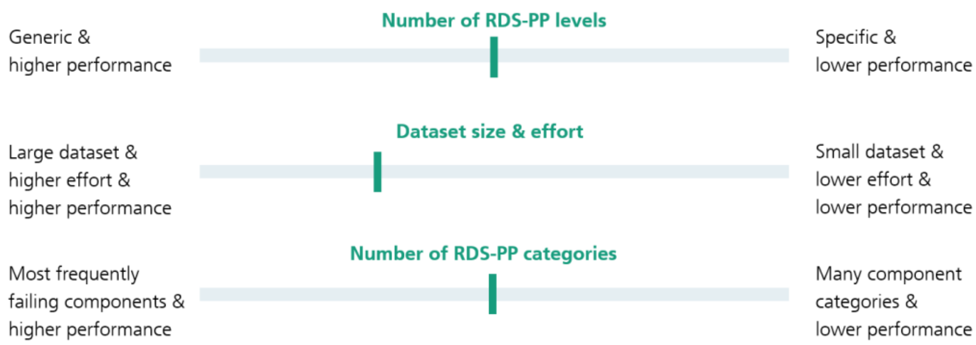

Figure 23. Summary of interviewees' preferences for classifier configurations using either-or questions

Contrary to the authors' expectations of receiving a consistent set of responses, the interviews revealed a variety of preferences for classifier configurations. This diversity is largely attributed to the distinct requirements and motivations inherent to each interviewed company. Hence, the idea of a "one fits all" solution is not deemed viable.

6.4.4 Failure rate comparison of differently preprocessed datasets

To gauge the uncertainty tied in various preprocessed datasets, failure rates of wind turbines' subsystems and components were selected as KPIs, aside from machine learning metrics like F1 scores and accuracy. Industry often relies on failure rates to understand the frequency at which components and subsystems fail. They serve as important KPIs for both benchmarking operational wind farms and planning for future wind farm projects. Consequently, the authors aimed to discern the potential variation in these KPIs based on different preprocessing approaches applied to maintenance reports. Therefore, two distinct analyses were carried out: Firstly, maintenance reports were labelled using selected classifiers. Without having access to these results, the same reports underwent manual labelling. Secondly, maintenance reports were manually classified by two different organisations both using RDS-PP as labelling guidelines. Following these processes, the failure rates derived from each differently preprocessed dataset were computed and set for comparison. The results of these investigations are detailed in the subsequent sections.

6.4.4.1 Manual labelling vs. text classifier

The dataset and text classifiers used in this study where the same as the ones employed in the analysis presented in Subsection 3.1.5, with the results visualised in **Figure 22**. For the 13 subsystems that fail most frequently, normalised failure rates were deduced from the different preprocessed datasets and are illustrated in **Figure 24**.

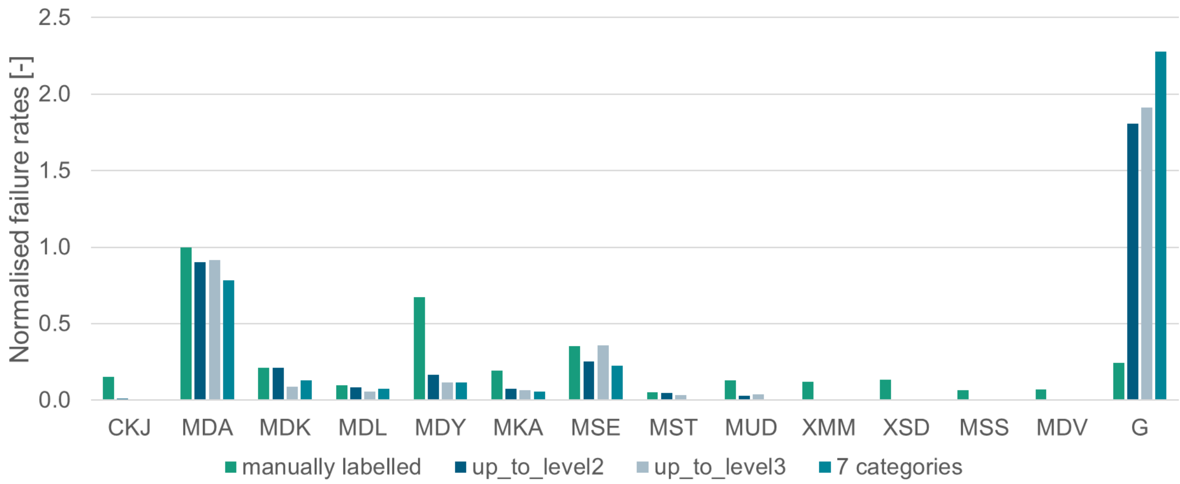

Figure 24. Comparison of normalised failure rates of differently preprocessed datasets (for translation of RDS-PP codes see Table 25 in Appendix C)

To normalise the data, the failure rate of the most frequently failing subsystem, i.e., the rotor system (MDA) as per the manually labelled dataset, was utilised. Although the F1 scores already conveyed that the classifiers' performance for varying wind farms (even from the same operator) might not be fit for productive use (cf. **Figure 22**), the normalised failure rates provide deeper insights into how these classifiers' function and perform.

Generally, the results reveal that the five most frequently failing subsystems are consistent with the ones of the dataset initially used for both training and testing the classifiers (cf. **Table 12**). Only the order switched between the converter system (MSE) and control system (MDY). For many of the top failing subsystems, the failure rates calculated from classifier-processed data somewhat align with the manually labelled data. In contrast, categories like the environmental measuring system (CKJ) or ancillary systems (XMM, XSD) barely make a mark in the statistics when processed by the classifier. Interestingly, there are two major subsystems, namely the control system (MDY) and the power generation system (MKA), which showcase substantially reduced failure rates when deduced from classifier-labelled data. Therefore, this suggests that these component categories are difficult to predict for the classifiers, despite their relatively frequent occurrences in the training dataset, in comparison to the ones of e.g., ancillary systems. Due to the classifiers' inherent struggle to predict labels with a limited representation in the training dataset, the failure rate of the category "G", which stands for the "overall system energy conversion" and is the fallback label for all text descriptions, which were not possible to be sorted into one of the more specific categories, is seven to nine times higher when derived from classifier-processed data, in contrast to the manually labelled data.

As concluded in Subsection 6.4.2.5, applying a pre-trained classifier to data from different wind farms, even if they are from the same operator, can lead to unsatisfactory categorisation results. However, the failure rate comparison demonstrates that the prediction for some subsystem categories is still reasonably accurate, e.g., when utilising the classifier predicting labels up to RDS-PP level 2, even though it had only half of its predictions correct for the entire dataset. Hence, failure rate deviations will be notably smaller for datasets labelled by a classifier specifically trained for them. Another approach would be to artificially enhance the training dataset with text examples from subsystem or component categories that are underrepresented. This could help achieve comparable performance for typically less frequent labels to those that appear regularly. Moreover, given the common assignment to the overarching category "G", it might be advisable to first use a trained classifier for labelling the dataset. Subsequently, entries labelled as category "G" can be manually relabelled. This approach would considerably reduce the manual work while maintaining the reliability of the result.

6.4.4.2 Uncertainty related to manually labelling maintenance reports

In the previous section, we analysed the differences in failure rates that arose from differently labelled datasets, benchmarking the classifiers against the expertise of wind energy professionals. However, during the data preparation phase for training the classifiers, it became evident that even expertly labelled data can vary. To gauge the significance of this variation, failure rates from the same dataset of operator 2, categorised by two distinct organisations, were calculated. The subsequent step involved determining the difference in failure rates as multiples from organisation 1. Results at the level of wind turbine subsystems are showcased in **Figure 25**. Variances span from 0.71 for category MSC (Generator Switching System) to a 3.5-fold higher failure rate for category XGM (Fire Extinguishing System). Substantial discrepancies in KPIs are also evident in subsystems like the drive train system (MDK) and lifting gears (XMM), which exhibit 1.79- and 1.77-times higher failure rates, respectively.

This can be explained by several aspects:

1) Interpretation of RDS-PP Guidelines: The guidelines of RDS-PP are crafted to categorise any wind turbine technology. As a result, component categories are not too specific. Some components could arguably fit into multiple RDS-PP categories, leaving room for interpretation. Even though both organisations engaged in regular discussions about such categorisation ambiguities, it remains challenging to unanimously decide

which components "clearly" fall into a particular RDS-PP category. Ultimately, the decision is up to the individual.

2) Uncertainties in ZEUS labelling: The State-Event-Cause-Code "ZEUS" [41] served as a guide to standardise the labelling of a component's state or the corresponding maintenance actions on the wind turbine. Within this study, a failure is defined as a fault necessitating technician intervention and spare parts usage to restore the function of the wind turbine. Therefore, only corrective maintenance measures involving component replacement is deemed a failure event. The ambiguity arises when maintenance reports lack detail, leading experts to potential judge certain replacements as corrective or planned differently. This can subsequently result in variant KPIs.
3) Human factors in manual labelling: Manually labelling maintenance reports is exhaustive and time-consuming. Factors like an individual's expertise, their mental state during labelling or their specialisation in either electrical or mechanical components can influence labelling decisions.

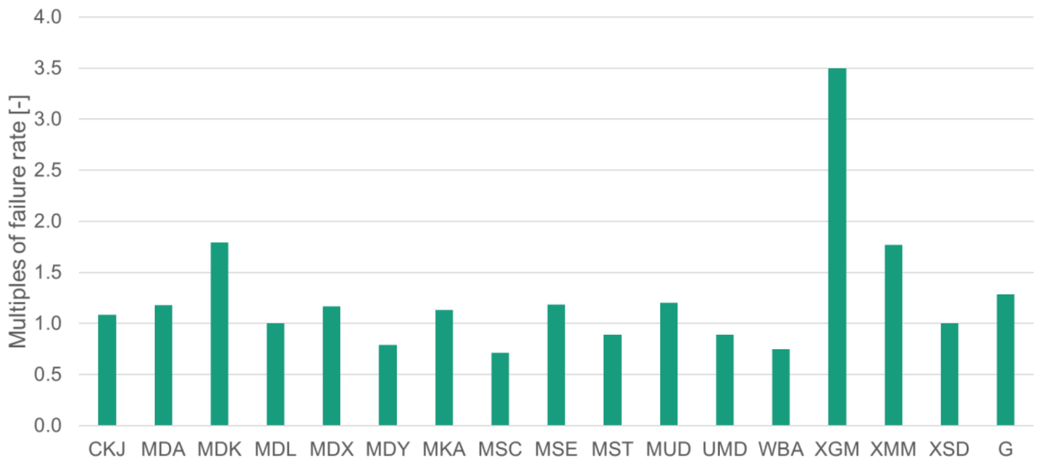

Figure 25. Multiples of failure rate for each wind turbine subsystem shown by RDS-PP categories comparing results based on preprocessed datasets by organisation 1 and organisation 2 (for translation of RDS-PP codes see Table 25 in Appendix C)

To illustrate the last point, a comparison is presented in **Figure 26,** showcasing failure rates at the component level. As an example, the subsystem frequency converter (MSE) is chosen.

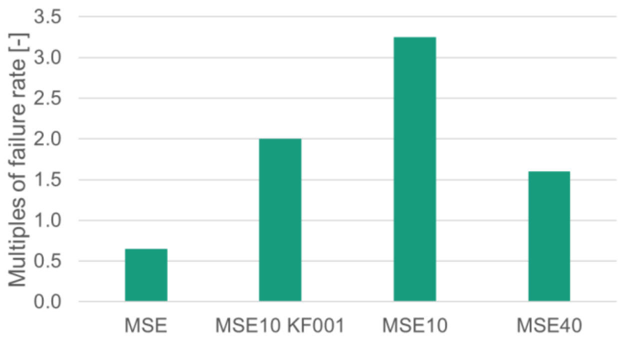

Figure 26. Multiples of failure rate exemplarily for components of the converter system shown by RDS-PP categories comparing results based on preprocessed datasets by organisation 1 and organisation 2 (for translation of RDS-PP codes see Appendix C)

Within the subsystem, differentiation is made among three component categories (MSE10 KF001, MSE10, and MSE40), as well as the general category MSE. This general category accumulates all failures that were not described with enough precision in the maintenance reports to label a specific affected component.

In this example, the failure rates for specific component categories from organisation 2 are at least 1.6 times higher than the ones of organisation 1. Meanwhile, the failure rate for the general category is 0.65 times lower in comparison. This discrepancy could indicate the presence of experts in the field of power electronics in organisation 2 who might be more confident in labelling specific component categories over a broader category, given the more detailed information available in the maintenance reports. However, this information might not be as comprehensible to individuals from different engineering backgrounds. Since RDS-PP adopts a hierarchical structure, choosing a more general category for labelling is not incorrect, though the aim should always be to label as specifically as possible. Yet, this approach can also result in variances in failure rates, particularly at the component level. This observation emphasises the challenges inherent to comparing failure rates across different publications, even when they employ the same taxonomies or failure definitions.

This critical finding underlines the necessity of further efforts in standardising the labelling process of maintenance reports also across organisations. While standards and guidelines as RDS-PP and ZEUS give recommendations how to proceed, the forementioned analyses have shown that instructions seem to fail achieving consistency. Therefore, specific examples of maintenance descriptions and how to apply these standards would be beneficial. Most helpful would be a parts list of each turbine type provided by the OEMs with respective RDS-PP translations attached. Making such information publicly available would greatly contribute to consistent data preprocessing allowing for better interpretation and comparability of KPI calculations.

6.4.4.3 Barriers to the adoption of text classifiers and potential applications in the wind energy sector

One of the primary technological barriers is the variability and inconsistency in maintenance report formats. Different operators use diverse terminologies and reporting standards, complicating the training of robust classifiers. The implementation of industry-wide standards, such as RDS-PP, can mitigate this issue by providing a uniform framework for categorising maintenance activities. Furthermore, while text classifiers can achieve high accuracy, their performance can vary significantly based on the quality and representativeness of the training data. Ensuring that classifiers are trained on comprehensive and diverse datasets is crucial to maintain reliability across different wind farms and operators. Additionally, integrating NLP models with existing maintenance management systems (MMS) and enterprise resource planning (ERP) systems poses a challenge. Seamless integration requires APIs and middleware that can handle the specific data structures and workflows of these systems.

Adoption of new technologies necessitates significant change management. Maintenance staff and engineers need to be trained to trust and effectively use these automated systems. There might be resistance due to perceived threats to job security or scepticism about the reliability of automated systems. Moreover, introducing text classifiers into established workflows can initially disrupt operations. Careful planning and phased implementation, starting with pilot projects, can help mitigate disruption and demonstrate the benefits gradually.

Developing, training, and integrating text classifiers involves upfront costs, including technology investments, data labelling efforts, and training programs for staff. For smaller operators, these costs might be prohibitive without clear demonstrations of return on investment (ROI). Furthermore, text classifiers require ongoing maintenance and updates to handle new terminologies, equipment, and failure modes. This ongoing cost needs to be factored into the economic feasibility of adopting such technology.
The integration of text classifiers can be envisioned through several steps. First, an initial data assessment is essential to evaluate the quality and standardisation of existing maintenance logs. Following this, a pilot project can be implemented in a controlled environment, such as a single wind farm or a specific subset of maintenance reports. This pilot phase allows for testing and adjustments before broader deployment. Training and onboarding sessions for maintenance staff and engineers are crucial to familiarise them with the new system and its benefits. Eventually, a full-scale implementation can be pursued, gradually expanding the use of text classifiers across all operations, ensuring continuous monitoring and feedback.

Incorporating text classifiers into current MMS and ERP systems can significantly enhance their functionality. Systems like SAP PM (Plant Maintenance) or IBM Maximo, which manage extensive maintenance records and data from diverse sources, including sensor readings, operational logs, and manual reports, can benefit significantly. These systems can automate the categorisation and standardisation of maintenance records by embedding text classifiers. This automation facilitates easier tracking and analysis of component failures and maintenance activities, thus improving data accuracy and efficiency. Additionally, platforms such as Microsoft Azure and AWS provide NLP services that can be tailored for specific industry needs, offering scalable and secure deployment options. Demonstrating real-world applicability involves showcasing the efficiency gains and accuracy improvements in maintenance data processing. By leveraging cloud-based solutions, real-time data analytics, and user-friendly interfaces, the adoption of text classifiers can be streamlined, providing tangible benefits in terms of reduced downtime and optimised maintenance schedules.

Furthermore, text classifiers can be seamlessly integrated with predictive maintenance tools that utilise data and machine learning algorithms to foresee equipment failures before they happen. Accurate and standardised maintenance records provided by text classifiers improve the precision of predictive models. This enhancement leads to more effective maintenance strategies, further preventing unexpected breakdowns and extending the life of turbine components (cf. **Figure 27**).

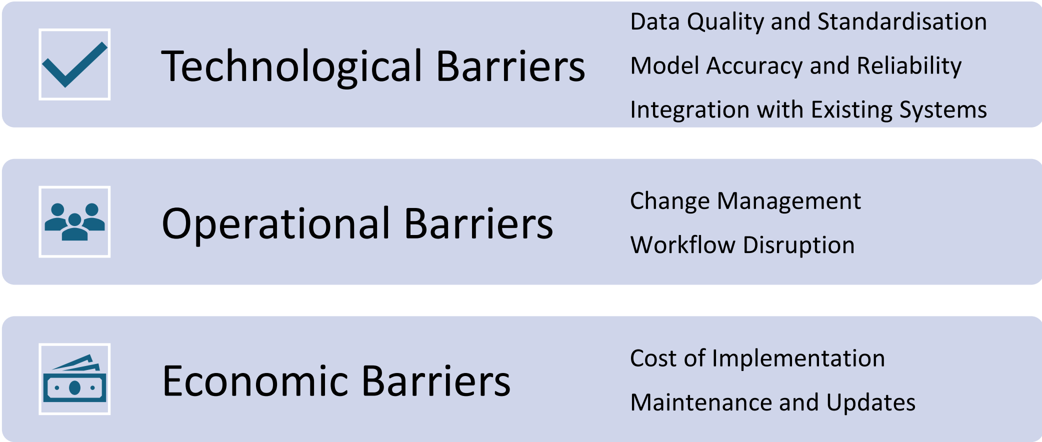

Figure 27. Barriers to the Adoption of Text Classifiers

6.5 Conclusions and outlook

This study assessed the viability of text classifiers for preprocessing wind turbine maintenance reports, highlighting their potential to reduce manual data processing efforts significantly. Main conclusions can be summarised as follows:

- Text classifiers achieved high micro F1 scores when trained on specific datasets, demonstrating their effectiveness. However, their performance decreased when applied to different wind farms, indicating the necessity for context-specific training.
- The research also underscored the importance of cost and resource efficiency, showing that smaller, well-curated training datasets can still produce competitive results. This finding emphasises the need to balance manual labelling efforts with classifier performance for practical application.
- Industry feedback revealed diverse classifier configuration preferences, suggesting that custom solutions are essential to meet varied stakeholder needs.
- While text classifiers tended to over-generalise, leading to skewed KPI calculations, they remain valuable when combined with manual verification for critical categories, enhancing overall reliability.
- A significant insight from the study is the need for standardisation in maintenance reporting. Both automated and manual methods face uncertainties due to inconsistent documentation. Standardised designation systems like RDS-PP can improve data accuracy and reliability, resulting in more meaningful KPIs.

Looking ahead, large language models (LLMs) such as GPT-3 and GPT-4 offer the potential to overcome current limitations. Fine-tuning these models with domain-specific datasets may enhance their applicability in the wind energy sector, improving classification performance. Successful applications of encoder models in other fields, like healthcare (BioBERT) and finance (FinBERT), provide blueprints ([146], [147]). Additionally, developing comprehensive datasets that capture the technical jargon and variations in maintenance reports, along with better guidelines for applying standards like RDS-PP, will be crucial. Integrating text classifiers into maintenance management and ERP systems can enhance operational efficiency and decision-making. By focusing on these future directions, this study aims to contribute to the improvement of maintenance data processing in the wind energy industry, ensuring more accurate and reliable analysis and reporting.

6.6 Acknowledgements

The present work was mostly carried out within the research project "Digitalization of Maintenance Information (DigMa)" funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), grant number 03EE2016A. The authors thank the project partners for providing field data and for sharing their requirements and experience. Further financial support was received by EPSRC through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1.

7 Reliability and O&M key performance indicators of onshore and offshore wind turbines based on field-data analysis

Based on maintenance data from over 1000 onshore and offshore wind turbines covering more than 4200 operating years, this study presents an analysis of failure rates, repair times, and maintenance resource requirements, focusing on subsystem-level reliability. Failure rates per turbine and megawatt are compared and failure behaviour over time is examined. Next to failure events, further corrective and preventive maintenance interventions are analysed. To provide more detailed insights for operations and maintenance simulations, a distinction is made between total major component replacements and those specifically requiring a jack-up vessel. Results show that onshore wind turbines have higher failure rates per megawatt than offshore wind turbines. Key subsystems including the pitch system, the control system, and power converter system are identified as critical to overall wind turbine reliability for both onshore and offshore wind turbines. For the overall wind turbine system, a failure behaviour over time following a bathtub curve is identified, with distinct trends for individual subsystems. The material of this chapter is currently under peer review for publication in [4].

7.1 Introduction

The global shift towards renewable energy has driven significant investments in wind energy, positioning it as a cornerstone of sustainable power generation. With the growing reliance on wind energy, especially in offshore environments, the reliability and performance of wind turbines have become critical factors that directly influence energy yield, operational costs, and overall asset integrity [19]. The effective management of these assets is particularly crucial as the industry aims to optimise operational efficiency and minimise downtime. However, achieving this requires a profound understanding of failure mechanisms and maintenance needs, underpinned by reliable data ([148], [149], [150]).

While there has been considerable progress in the development of wind turbine (WT) technology, the reliability assessment and optimisation of operations and maintenance (O&M) of these systems has often been hampered by a lack of comprehensive, high-quality field data. Existing studies on wind turbine reliability mostly rely on limited datasets or combine data from diverse turbine types and operating conditions without sufficient granularity ([12], [151]). Such approaches can obscure the differences in reliability performance across turbine types, manufacturers, and environmental contexts. Consequently, there is a critical need for detailed analyses based on comprehensive field data that can provide more accurate and actionable insights into failure rates and maintenance strategies [32].

This study represents a significant advancement in the field by presenting an extensive analysis of wind turbine reliability based on a large, representative sample of field data. Drawing on

[4] Julia Walgern, Nils Stratmann, Martin Horn, Nathalene Then Wei Ying, Moritz Menzel, Fraser Anderson, Athanasios Kolios, Katharina Fischer, 2025. “Reliability and O&M key performance indicators of onshore and offshore wind turbines based on field-data analysis”. Submitted to Wind Energy for publication [199]

maintenance reports spanning more than 4,200 operational years from both onshore and offshore wind turbines, this research provides one of the most comprehensive evaluations of failure rates and further O&M-related key performance indicators (KPIs) to date. The analysis includes data from nine different onshore and four offshore wind turbine original equipment manufacturers (OEMs), covering a range of turbine capacities and operational contexts. This breadth and depth of data allow for a more comprehensive understanding of reliability performance across various wind turbine systems, informing both design optimisation and O&M strategies for future wind farms.

Furthermore, the study introduces a detailed categorisation of wind turbine failures using the reference designation system RDS-PP, which is applied systematically for standardised component classification across all different turbine types and designs. By focusing on system and subsystem level and calculating average failure rates along with corresponding confidence intervals, this work studies the reliability behaviour of wind turbines with a rare and considerable level of detail, providing unnormalised KPIs and uncertainty quantifications. The findings reveal detailed insights into the different failure rates of onshore versus offshore turbines, the impact of turbine rated power on reliability, and the temporal patterns that characterise wind turbine failures.

This level of granularity in data analysis not only enhances the reliability modelling of current wind turbine fleets but also serves as a valuable resource for OEMs, operators, and policymakers looking to improve the design and operation of future wind farms. The analysis conducted in this study highlights specific reliability challenges as well as opportunities for technological improvements and maintenance optimisation, making it an essential input for risk management and decision-making in the wind energy sector [152].

Recognising the sensitivity of the data involved, we have systematically evaluated and implemented measures to ensure the confidentiality and security of the datasets used in this research. These measures were crucial for protecting proprietary information and maintaining the trust of data providers while enabling the comprehensive analysis presented herein.

In the sections that follow, we provide a thorough review of the current state of the art in wind turbine reliability research and outline the methodologies and datasets used in this study. This is followed by an in-depth presentation of the results, discussing their implications for both the operational management of wind farms and future research directions.

7.2 State of the art literature on wind turbine reliability

7.2.1 Overview of wind turbine reliability research

Understanding wind turbine reliability is crucial for optimising their performance and minimising operational costs, especially for offshore installations. Early studies primarily focused on analysing key reliability metrics such as average failure rates, mean time to repair, and availability, which are essential for developing maintenance strategies [153]. Over the years, efforts have been made to standardise the collection and analysis of reliability, availability, and maintainability (RAM) data across diverse turbine types and environments [32]. Prominent initiatives like the WInD-Pool common knowledge base in Germany and the SPARTA program in the United Kingdom have been instrumental in adopting structured methodologies to gather operational data from wind farms ([101], [154]).

Due to the strict confidentiality of maintenance data, only a limited number of reliability studies have been published. European initiatives such as WMEP [155], as well as WSD, WSDK, and LWK (e.g. [25], [9]) were among the first to analyse WT maintenance data, covering periods from the 1990s until 2004. More recent studies include those from the ReliaWind project [10], the University of Strathclyde [12], the AWESOME project [28], and the SPARTA initiative [29]. Additionally, detailed reviews of published failure rate statistics have been conducted by [31], [32], [33], and [156]. However, as most of these studies rely on datasets recorded before 2015, with SPARTA being the only initiative providing more recent reliability and performance KPIs from 2020/21 [13], there remains a need for comprehensive and high-quality field data, particularly for modern, larger turbines.

7.2.2 Key performance indicators for reliability assessment

KPIs are critical for evaluating wind turbine reliability and optimising O&M. Among the most widely used KPIs in reliability studies are failure rates, mean time to failure (MTTF), mean time to repair (MTTR), and time-based or energy-based availability [31]. An overview of commonly used KPIs is presented in **Table 13**. The failure rate, typically measured as the number of failures per turbine per year, is a fundamental metric that provides important insights into turbine reliability and is typically utilised as input for O&M modelling [157].

Comparative studies reveal significant differences not only in the aforementioned failure rates but also in further O&M-related KPIs between onshore and offshore wind turbines and associate these with varying environmental conditions, maintenance access, and design complexities [11]. Subsystems such as the pitch system, hydraulic systems, rotor, power converter system, generator, and gearbox are often identified as having the highest failure rates (cf. [12], [29]), especially in offshore installations where repairs are more challenging. [158] highlighted that corrective maintenance for these critical components often results in substantial downtime, underscoring the need for robust designs and advanced monitoring systems.

Identifying general trends in reliability and maintainability helps operators to pinpoint where reliability improvements could lower the levelized cost of energy (LCOE). However, challenges such as inconsistent data collection practices complicate the comparison of reliability metrics across different studies. Efforts like those by the International Energy Agency (IEA) Wind Task 33 aim to address these challenges by providing standardised frameworks for data collection and analysis [20].

The variability in methodologies and reliability indicators points to the need for more standardised approaches to provide actionable insights. Enhancing RAM databases with detailed failure and operational data is crucial for advancing wind turbine design and maintenance strategies [159].

Table 13. Summary table of key performance indicators (KPIs) for wind turbine reliability

KPI	Definition	Importance	Common Calculation Methods
Failure Rate	Average number of failures per unit (e.g., per turbine) per year for a specific component or subsystem	Indicates the reliability of wind turbine components and subsystems; high failure rates can lead to increased maintenance costs and downtime.	Empirical analysis using maintenance data
Corrective Maintenance Rate	Frequency of corrective maintenance interventions	Indicates the reliability of wind turbine components and subsystems	Empirical analysis of maintenance records
Unscheduled Maintenance Rate	Frequency of maintenance activities related to unexpected failures	High rates suggest frequent unexpected failures and may affect downtime and operational planning.	Empirical data analysis from maintenance reports
Preventive Maintenance Rate	Frequency of preventive maintenance interventions	Helps in understanding the maintenance strategy. High rates suggest that preventive maintenance interventions and associated costs are accepted to prevent failures	Empirical data analysis from maintenance reports
Mean Time to Failure (MTTF)	Average time between failures for a non-repairable specific component or subsystem	Helps in understanding the expected lifetime of components; a higher MTTF indicates better reliability.	Statistical modelling; Survival analysis; Weibull distribution
Mean Time to Repair (MTTR)	Average time required to repair a failed component or subsystem and restore it to operational condition	Critical for planning maintenance resources and minimising downtime; a lower MTTR indicates more efficient maintenance processes.	Empirical analysis based on maintenance records
Mean Time Between Failures (MTBF)	Average time between successive failures of a repairable system or component	Indicates the reliability of wind turbine components and subsystems; a higher MTBF indicates better reliability.	Calculated as the inverse of the failure rate
Availability (time-based)	The proportion of time a wind turbine is operational and capable of generating power	Reflects overall performance and reliability of wind turbines; high availability is key to maximising energy production and minimising losses.	Time-based calculations using operational and downtime data, i.e. typically SCADA data; Markov models
Downtime	Total time during which a wind turbine is not operational due to failures or maintenance	Directly impacts energy yield and economic returns; high downtime leads to significant losses in revenue.	Derivation from SCADA data

7.2.3 Common causes of failures and reliability challenges

Understanding the prevailing causes of failures in wind turbines is crucial for enhancing their reliability and maintenance strategies. In the literature, the following failure modes and causes are reported: The gearbox frequently fails due to bearing and gear fatigue, misalignment, and lubrication issues, leading to significant downtime ([12], [28]). Additionally, tribological failures such as pitting and scuffing affect gearboxes due to inadequate lubrication. The generator faces electrical and mechanical failures such as stator faults and insulation degradation due to electrical surges and thermal stresses [160]. Power converter failures are dominated by failures of the power semiconductor modules, their driver boards, the converter control system as well as the cooling system ([36], [104]). The pitch system is vulnerable to mechanical wear from continuous blade angle adjustments in varying wind conditions [161]. Meanwhile, blades are prone to erosion, fatigue, and lightning strikes, affecting turbine performance [162].

Common failure mechanisms include fatigue, particularly in moving parts like blades and gear teeth due to cyclic loading. For many years, fatigue due to power and thermal cycling was postulated to be the main failure mechanism also in power converters, until comprehensive field-data and damage analyses revealed that climatic influences, which drive corrosion and affect insulation integrity in the converter, play a more important role in the wind-power application ([163], [36]). Corrosion is a relevant failure mechanism also for support structures, especially in offshore environments where saltwater accelerates degradation ([164], [165]).

It is important to keep in mind that the detailed identification of failure root causes and the underlying mechanisms can be a complex and laborious task, often requiring comprehensive data evaluation and analyses of damaged components. As the above example of power converters shows, there is a certain risk that hypotheses or postulates about prevailing failure mechanisms propagate through the literature and divert attention from the reality observed in the field.

7.2.4 Impact of turbine design, manufacturer, and age on reliability

The reliability of wind turbines is significantly influenced by their design and the manufacturer. Studies have shown that design choices, such as drivetrain configurations (e.g., geared vs. direct drive) and control systems, affect failure rates and maintenance needs ([166], [167]). For instance, direct-drive turbines eliminate the gearbox, reducing failures associated with gears and bearings, but they may have higher rates of electrical component failures due to the larger size and complexity of the generator and converter systems. Additionally, differences in manufacturing quality and component selection between manufacturers can lead to variability in reliability performance [168]. Standardisation, stringent quality control during the design and manufacturing phases as well as test-based reliability validation are essential to reduce such variability, ensuring consistent reliability across different turbine models and brands.

The operating age of wind turbines also significantly impacts their reliability. As turbines age, wear and tear from continuous operation, exposure to harsh environmental conditions, and fatigue loading can lead to increased failure rates [153]. Studies indicate that older turbines often experience failures in components such as blades, gearboxes, and electrical systems, which degrade due to prolonged exposure to mechanical stresses and environmental factors like temperature and humidity variations [169]. Other subsystems, such as the power converter, exhibit pronounced early failures [47]. In general, failure patterns of technical systems typically follow a "bathtub curve”, where failure rates are decreasing during the early-failure phase, remain relatively constant during a "useful life" phase, and increase again as components degrade in the deterioration phase [170]. Understanding these patterns is crucial for optimising maintenance strategies and extending the operational life of wind turbines.

7.2.5 Data-driven approaches and advanced analytical methods

The use of big data and machine learning (ML) has transformed the field of wind turbine reliability analysis, enabling more accurate early fault detection and enhanced maintenance strategies. Recent advancements leverage data-driven approaches using large datasets from SCADA (supervisory control and data acquisition) systems, which provide high-frequency data on turbine operations and performance ([171], [172]). Machine learning techniques such as neural networks, random forests, and support vector machines have been employed to detect patterns in operational data, predict failures, and optimise maintenance schedules, thereby reducing downtime and maintenance costs ([173], [174], [175]). AI-based predictive maintenance approaches also incorporate data fusion techniques that combine SCADA data with environmental and maintenance records, offering a more comprehensive view of turbine health and enabling proactive interventions [176].

Recent meta-analyses and systematic reviews have consolidated findings across multiple studies to provide higher-level insights into wind turbine reliability management. For example, a meta-analysis by [177] aggregated reliability data from diverse sources, revealing trends in failure rates and highlighting critical components that require attention. These reviews often use statistical methods to compare data from different regions, turbine types, and operating conditions, offering a benchmark for reliability performance. By synthesising data from various studies, systematic reviews inform best practices for condition monitoring, component design, and maintenance planning, addressing gaps in existing literature and guiding future research. Such efforts help standardise reliability metrics and improve the robustness of reliability models, ensuring more effective asset management strategies for both onshore and offshore wind farms.

7.2.6 Knowledge gaps and future research directions

Despite significant advancements in wind turbine reliability research, several gaps remain. A summary of those is shown in **Table 14**. Many studies rely on limited sample sizes and data from specific regions, which may not accurately represent broader operational contexts [101]. There is also a lack of comprehensive field data that captures the full spectrum of failure modes and environmental influences, especially for offshore turbines [32]. This is often related to strict data confidentiality. Additionally, existing research often focuses only on a few subsystems (e.g. [29]), leading to gaps in reliability modelling. For example, [178] highlight that main bearings are frequently overlooked in reliability analyses. In reliability analyses, ensuring the recentness of data and coverage of modern WT technology remains a key challenge. As a result, many studies frequently reference literature based on older datasets that primarily reflect outdated turbine technology.

Table 14. Summary table of research gaps

Research Gap	**Description**
Limited sample sizes	Many studies use data from small, specific samples, limiting the generalisability of the findings.
Lack of diversity in field data	Inadequate data coverage on different environments and conditions, especially for offshore sites
Insufficient coverage of certain subsystems	Underrepresentation of specific subsystem failure types or insufficient reliability data for certain subsystems
Lack of recent field data	Most studies are based on old datasets, not covering modern WT technology.
Need for standardisation and harmonisation	Lack of standard methodologies and definitions across studies complicates comparative analysis.

This study aims to address these gaps by using a more representative sample size and conducting a comprehensive analysis of both onshore and offshore wind turbine maintenance data. By integrating diverse datasets including modern turbine technology and systematically evaluating failure modes across various subsystems, this research offers a more holistic view of turbine reliability which is applicable for future wind farm design and operation.

7.3 Methodology and datasets

7.3.1 Methodology

7.3.1.1 Field-data collection and preprocessing

Maintenance reports, which are available for each visit of a wind turbine, of more than 1000 wind turbines were collected making an effort to include a variety of turbine types of both onshore and offshore turbines. Attention was paid to incorporate recently commissioned turbines as well as having datasets of turbines which have a certain track record already. This leads to a unique field-data collection with respect to its size, diversity and recentness.

Maintenance records include information about what maintenance intervention was carried out on which turbine on which date. Those reports can have different lengths and levels of detail. Typically, at least spare parts and / or work descriptions are recorded, which allow one to understand what kind of work technicians have performed on the turbines. In order to conduct different reliability analyses, the data needs to be machine-readable and comparable even though the reports stem from different organisations and sites. Within this study standards and guidelines like the reference designation system RDS-PP for wind turbines [40] and the State-Event-Cause-Code “ZEUS” [41] are utilised to support the preprocessing. RDS-PP is used to classify maintenance interventions according to the components and subsystems that were maintained. Using ZEUS, activities performed by technicians are labelled as corrective and preventive and further differentiated according to the specific maintenance action undertaken. The preprocessing results in a comprehensive field-data base covering:

- Wind turbine ID and respective wind farm
- Wind turbine manufacturer and type
- Commissioning date of the turbine
- Rated power of the turbine
- Technical information about the different subsystems
- Coordinates of the turbine
- Data provider
- Time stamps of start and end date of each maintenance activity
- Number of technicians involved
- Components and subsystems affected (standardised codes of RDS-PP)
- Type of maintenance activity (standardised codes of ZEUS)

7.3.1.2 Reliability analyses

In order to assess O&M activities and WT reliability performance, different reliability analyses are performed and KPIs computed. Respective KPIs can be utilised for benchmarking of different assets, understanding failure patterns as a basis for developing countermeasures, or as input for development and O&M simulation of future wind farms.

KPIs are assessed for corrective and preventive maintenance interventions. Particular attention is paid to failures of components and subsystems as those are afflicted with costly downtimes requiring maintenance and the use of spare parts. Within this study, a failure is defined as an event necessitating corrective maintenance (ZEUS code "02-08-01") and which is not resettable but requires a component to be replaced (ZEUS code "02-09-09-01"). In order to compare reliability KPIs of different components, subsystems and overall turbines, the following average rates are calculated:

$$corrective\ maintenance\ rate\ c = \frac{\sum_{i=1}^{I} C_i}{\sum_{i=1}^{I} X_i T_i} = \frac{C}{T} \tag{7.1}$$

$$preventive\ maintenance\ rate\ p = \frac{\sum_{i=1}^{I} P_i}{\sum_{i=1}^{I} X_i T_i} = \frac{P}{T} \tag{7.2}$$

$$failure\ rate\ f = \frac{\sum_{i=1}^{I} N_i}{\sum_{i=1}^{I} X_i T_i} = \frac{N}{T} \tag{7.3}$$

Herein, C_i is the number of corrective maintenance visits, P_i is the number of preventive maintenance visits, and N_i is the number of failures of the analysed component or subsystem in the time interval i. X_i is the number of WTs analysed within this time interval of duration T_i. Consequently, the average rates are equal to the quotient of the sum of all corrective, preventive or failure events, C, P and N, respectively, and the total amount of considered WT operational years T.

As WTs of different power classes are included in the analyses, next to average rates per WT and year, average rates per rated capacity in MW and year are also calculated.

Moreover, confidence intervals for the failure rates are computed to quantify the uncertainty stemming from the size of the datasets. According to [179], the confidence intervals for failure rates based on time-censored data are estimated using:

$$\left[\frac{\chi^2(\frac{\alpha}{2},2N)}{2T},\frac{\chi^2(1-\frac{\alpha}{2},2N+2)}{2T}\right] \tag{7.4}$$

Herein, $\chi^2(\alpha/2{,}2N)$ is the $(\alpha/2)$-quantile of the χ^2 distribution with $2N$ degrees of freedom. In this study, $\alpha = 0.1$ is utilised to provide confidence intervals with a confidence level of 90%. As explained in more detail in [36], these confidence intervals based on sample data are to be interpreted in terms of frequency: if a large number of samples (in this case failure or maintenance datasets covering a part of a WT population) was evaluated, the confidence intervals determined according to Equation (7.4) would cover the true value of the failure rate in 90% of the cases.

7.3.2 Datasets

The datasets underlying this analysis are based on maintenance reports of onshore and offshore wind turbines. In total, more than 4200 operational years are covered. A detailed overview of the datasets is provided in **Table 15**. While the offshore data stem from turbines of four different OEMs with turbine capacities ranging up to 9 MW, the onshore data comprise turbines of nine different manufacturers. In total, 1089 WTs located in seven different European countries are considered in the present study.

Table 15. Information about the datasets which have been considered in the analysis

	Offshore	Onshore
WT operational years considered	1755	2489
Number of WT OEMs covered	4	9
Rated capacity considered	Up to 9 MW	
Available data period	2006-2024	

The dataset analysed in this study encompasses the following technical concepts:

- Pitch system: hydraulic, electrical
- Drive train concepts: geared, direct drive, hybrid drive
- Generator types: doubly-fed induction generator (DFIG), electrically excited synchronous generator (EESG), permanent magnet synchronous generator (PMSG), squirrel-cage induction generator (SCIG); including low voltage (LV) and medium voltage (MV) generators
- Converter technology: air-cooled, liquid-cooled; including LV and MV converters

While it is important to include data of both, WTs, which have been operated already for some time to analyse failure behaviour over time, and WTs, which have just recently been commissioned to incorporate newest technologies, this leads to a diverse dataset of different turbine generations. The data period analysed in this study is nearly identical for both onshore and offshore WTs resulting in comparable age distributions across the two categories. Note that 12.5% of the WTs have a capacity smaller than 2 MW. Most WTs covered within this study can be considered as recent turbine technology.

7.4 Results and discussion

7.4.1 Comparison of failure rates for onshore and offshore wind turbines

Figure 28 and **Figure 29** illustrate a comparative analysis of failure rates for onshore and offshore WTs, calculated per WT and year, as well as per MW of turbine capacity and year, respectively. In addition to presenting the average failure rate of the entire WT, **Table 17** provides the average failure rates for all 29 subsystems defined by RDS-PP, along with a corresponding translation of RDS-PP codes. For better clarity in the presentation of results, the analysis in this section is limited to the eleven most critical subsystems, selected based on failure frequency. Components that could not be unequivocally assigned to a specific subsystem are categorised under "G", representing "other components". It is important to note that the sum of all subsystem failure rates exceeds the overall WT failure rate, as certain failure events involve the replacement of components across multiple subsystems.

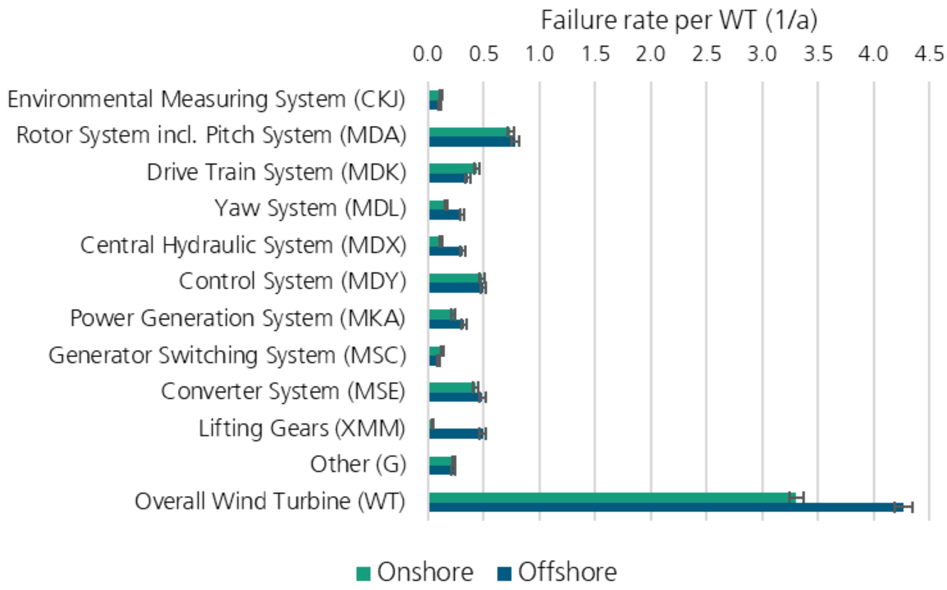

Figure 28. Failure-rate comparison per WT and year of onshore and offshore WTs including the eleven most critical subsystems

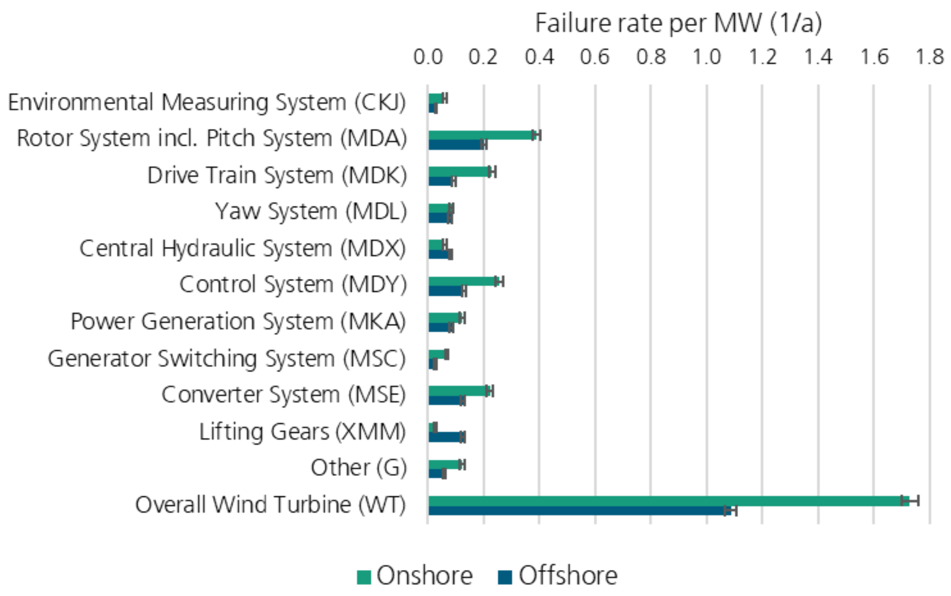

Figure 29. Failure-rate comparison per MW of turbine capacity and year for onshore and offshore WTs including the eleven most critical subsystems

The comparison of average failure rates per WT and year indicates a higher reliability of onshore WTs (3.3 vs. 4.3 failures per offshore WT and year), consistent with findings frequently reported in the literature [32]. However, when normalised per MW and year, the data reveal that onshore WTs exhibit a higher failure frequency per WT capacity, with an average failure rate of 1.729 failures per MW per year, compared to 1.088 failures per MW per year for offshore WTs. Given the strong dependence of average failure rates on WT size – shown e.g. in [9], [180], [105], [47], and also found in our analyses – further analysis and interpretation are based exclusively on failure rates normalised per MW and year. While for onshore WTs the subsystems rotor system (MDA) including the pitch system, the control system (MDY), the drive train system (MDK), and the converter system (MSE) are identified as most critical, for offshore WTs the highest failure rates are recognised for the subsystems rotor system, control system, lifting gears (XMM), and converter system.

In previous publications by Fraunhofer IWES, which focused exclusively on the power converter, the converter subsystem also encompassed failures related to main circuit breakers and contactors (cf. [36], [163], [104], [47]). In contrast, this study categorises these failures separately within the "Generator Switching System" (MSC) subsystem in order to follow the RDS-PP classification. Additionally, while some of our earlier studies normalised failure rates based on the rated power of the converter, it is important to note that in the present analysis all failure rates, including that of the converter system, are normalised by the rated power of the turbine.

Note that the drive train system covers the subassemblies rotor bearing, speed conversion, drive train brake, high speed shaft, drive train auxiliary systems, main and offline gear oil systems, oil lubrication system, rotor lock, rotor slewing unit, and drive train cooling system. Therefore, the subsystem is evaluated across both WTs with and without gearboxes. A more detailed examination of the MDA system category reveals that for onshore WTs the pitch system accounts for approximately 80.8% of MDA system failures, whereas for offshore WTs it constitutes nearly 82.5% of failures within this category (cf. **Table 17**). Provided KPIs in **Table 17** can be utilised for estimating failures and maintenance interventions. However, it is important to note that the failure behaviour is not solely characterised by turbine size making more sophisticated reliability models necessary to support such analysis.

7.4.2 Failure-rate comparison across WT OEMs

Although it is common practice to report average failure rates derived from mixed fleets comprising different WT types, as presented in Section 4.1, this approach carries inherent risks. Reporting only a group-averaged failure rate without further differentiation might obscure major reliability differences, which can serve as key indicators for root-cause analysis and design optimisation. To address these limitations, an OEM-specific analysis is performed. **Figure 30** and **Figure 31** present the average failure rates of offshore WTs from four different OEMs and onshore WTs from six different OEMs. Where a manufacturer is included in both **Figure 30** and **Figure 31**, they do not share the same label for confidentiality reasons. This means that OEM1 in **Figure 30** is not the same manufacturer as OEM1 in **Figure 31**.

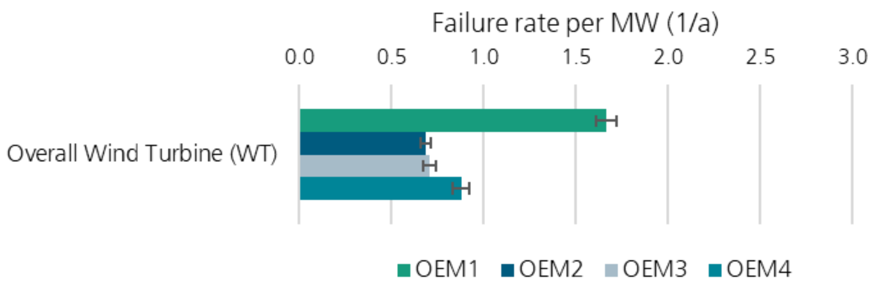

Figure 30. Failure-rate comparison per MW and year across WT OEMs of offshore assets

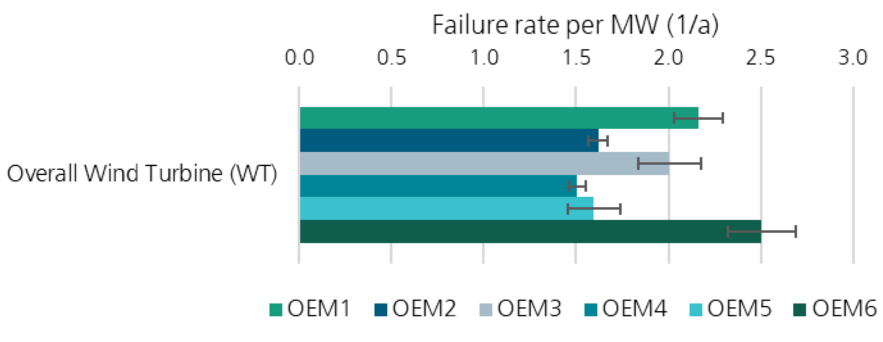

Figure 31. Failure-rate comparison per MW and year across WT OEMs of onshore assets

To ensure that the comparison reflects only technological differences, failure rates are again normalised per MW and year. Analysis results reveal significant disparities in failure rates between WTs from different manufacturers. For offshore WTs, the average failure rate for OEM1 is 1.6 to 2.4 times higher than that of the other three OEMs, with a distinct failure rate of 1.7 failures per MW per year. In the case of onshore WTs, failure rates range between 1.5 and 2.5 failures per MW per year. The variability in confidence intervals reflects the uncertainty associated with the sizes of the underlying data subsets. While datasets for all offshore OEMs and onshore OEMs 2 and 4 include at least 1100 MW-years, analysis for onshore OEMs 1, 3, 5, and 6 are based on smaller datasets ranging from 200 to 330 MW-years. Onshore OEMs 7, 8, and 9 are excluded from this analysis due to insufficient sample sizes. Overall, onshore OEM failure rates generally exceed those of offshore OEMs, with the exception of offshore OEM1, which exhibits a failure rate comparable to the three best-performing onshore OEMs.

7.4.3 Failure-rate behaviour through time

An essential aspect of reliability analysis is the evolution of failure behaviour over time. This is assessed by calculating failure rates across different operating years. To isolate the effect of WT aging, the analysis is conducted for specific WT types, avoiding the confounding influence of mixed turbine designs. As an example, **Figure 32** presents a comparison of normalised failure rates across different operating years, grouped into five periods of WT operating age, for a single WT type including eight representative subsystems.

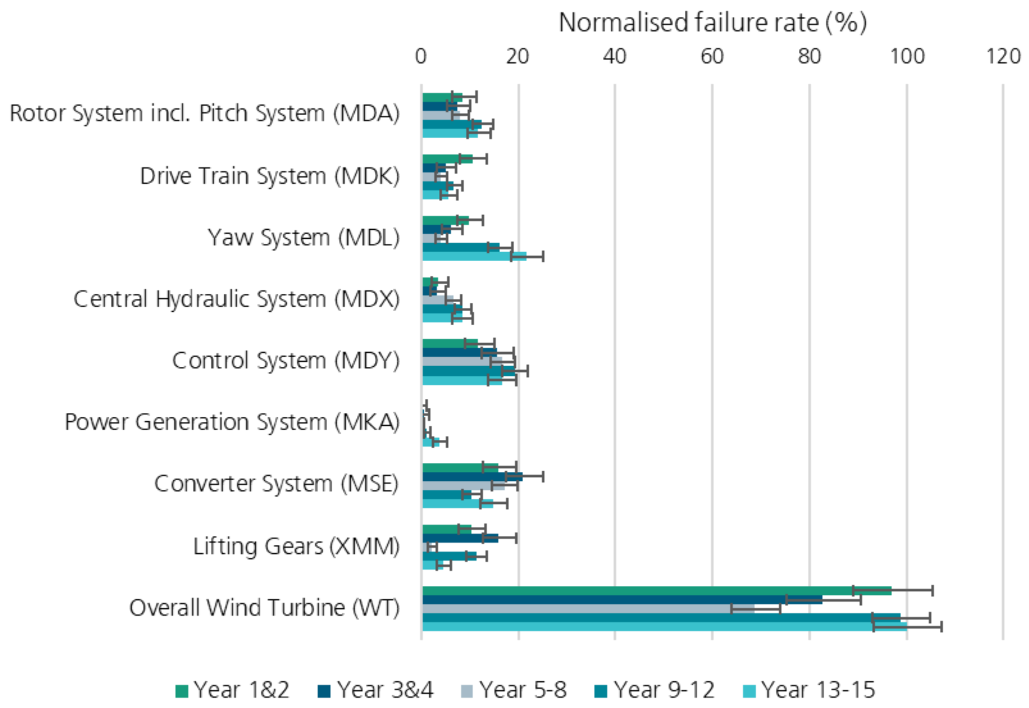

Figure 32. Comparison of normalised failure rates across different operating years for a specific WT type including eight exemplary subsystems

The failure rate trajectory for the entire WT system follows the characteristic shape of a bathtub curve [37]: During the initial years of operation, elevated failure rates are observed, corresponding to early failures. Over time, failure rates decline, reaching a lower and more stable level through operating years 5 to 8. From year 9 onward, failure rates increase again, indicative of degradation-related failures. Although confidence intervals show a slight overlap between some groups, the overall trend is clearly visible and observable across different WT types, both onshore and offshore.

The failure behaviour of individual subsystems varies significantly depending on the specific subsystem under analysis. While certain subsystems, such as the drive train system (MDK), yaw system (MDL), and converter system (MSE), exhibit a failure trend similar to that of the overall WT system, others, such as the central hydraulic system (MDX) and the power generation system (MKA), show a steadily increasing trend suggesting that these are primarily suffering from degradation-related failures. Additionally, some subsystems do not display a distinct trend due to overlapping confidence intervals, either because no distinct trend exists, or the dataset is too limited to detect one. These findings emphasise that the well-established bathtub curve in reliability modelling results from the superposition of different failure mechanisms and trends.

7.4.4 Other O&M relevant KPIs

When utilising reliability data for O&M simulations or OPEX calculations, additional O&M KPIs beyond failure rates are required as input. To address this, further analyses based on the offshore data subset are presented in the following. These include a comparison of corrective and preventive maintenance interventions, an analysis of major component replacements (MCR), and an evaluation of average repair times and the average number of maintenance technicians required per failure event and subsystem. Due to limited access to cost data and the impact of inflation, cost figures for spare parts are not provided, as comparisons across different datasets and years would be challenging. Reference values can be found in [12], [50], and [5].

7.4.4.1 Comparison of corrective and preventive maintenance interventions

Within this study the failure definition is based on the consumption of spare parts, while other corrective maintenance activities not requiring spare parts are classified under the category “Corrective Maintenance other”. In addition to addressing failure events and conducting troubleshooting and repairs – both classified as corrective maintenance interventions – technicians are also responsible for preventive maintenance interventions, such as scheduled maintenance. Furthermore, statutory inspections, functional tests, condition monitoring related activities – such as oil sampling – and routine tasks like topping up coolants or lubricants are categorised as preventive maintenance interventions. **Figure 33** displays the corresponding maintenance rates per MW and year.

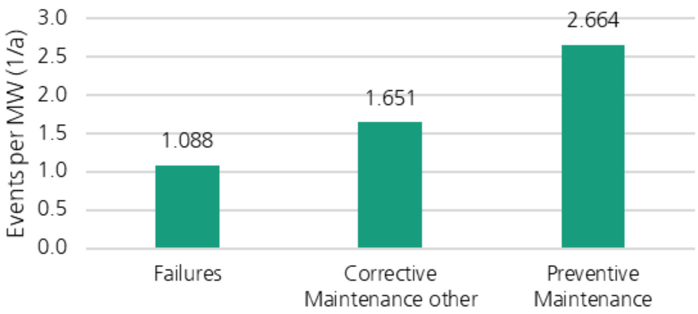

Figure 33. Comparison of corrective and preventive maintenance interventions for offshore wind assets differentiating corrective interventions into failures and other corrective maintenance

As detailed in Section 5.3, offshore WTs experience an average of 1.088 failures per MW per year. For example, this is equivalent to 5.4 failures per year for a 5 MW turbine and 10.9 failures per year for a 10 MW turbine. Additionally, the category “Corrective Maintenance other” accounts for 1.651 interventions per MW and year, while preventive maintenance actions total 2.664 interventions per MW and year. In total, this results in 5.403 maintenance interventions per MW per year. This translates to approximately 27 maintenance interventions annually for a 5 MW WT and to around 54 maintenance interventions for 10 MW WT. Similar intervention frequencies are observed across offshore wind farms with different WT power classes included in the datasets used for this analysis.

7.4.4.2 Major component replacements

The average failure rates per subsystem presented above are based on all corrective maintenance interventions involving the use of spare parts, regardless of the size or cost of the replaced component. To provide further details relevant for O&M simulations, a distinction is made between total major component replacements (MCR) and those that specifically require a jack-up vessel (JUV), as outlined in [181]. MCR encompasses replacements across six subsystems: the rotor system (MDA), the drive train system (MDK), the power generation system (MKA), the generator transformer system (MST), the nacelle (MUD), and the tower system (UMD). The components considered for each subsystem are listed in **Table 16**. Average offshore MCR rates as well as rates of interventions requiring a JUV are presented in **Table 17**.

Table 16. Considered components for major component replacements (MCR) and MCR requiring a jack-up vessel (JUV)

Subsystem	MCR requiring no JUV	MCR requiring a JUV
Rotor system (MDA)	-	blade, hub, blade bearing
Drive train system (MDK)	damaged high and low speed shaft	main bearing, gearbox, rotor shaft assembly
Power generation system (MKA)	generator bearings	generator
Generator transformer system (MST)	-	transformer
Nacelle (MUD)	-	nacelle
Tower system (UMD)	-	tower, transition piece, foundation

With an average of 0.0209 MCR per MW and year, the power generation system MKA accounts for the highest MCR rate, followed by the drive train system MDK at 0.0149 MCR per MW and year. Of these, 0.0097 MCR per MW and year require a JUV, making the drive train system the primary contributor to MCR events necessitating a JUV. For the rotor system MDA only blade and blade bearing replacements were observed, while no MCR events were recorded for the nacelle, tower, transition piece, or foundation. Across the entire WT, the total MCR rate is 0.0366 per MW and year, with 0.0117 MCR per MW per year requiring a JUV. For a wind farm comprising 50 WTs, each with a rated capacity of 10 MW, this corresponds to approximately 37% of WTs undergoing a MCR annually, with 12% requiring a JUV – equivalent to roughly six WTs.

7.4.4.3 Average repair time

The average repair time per subsystem is displayed in **Table 17**. It represents the total duration from the technicians' arrival to their departure from the turbine, regardless of the number of personnel involved in the maintenance intervention. Unlike downtime or time to repair, it does not account for travel time, lead time of spare parts, delays due to inaccessibility, or other external factors [12]. It is important to note that the average repair time is calculated across all failure events without distinguishing between failure severity. On average, component replacements for the overall WT system require 2.7 hours. Other corrective maintenance activities take approximately 1.5 hours, while preventive maintenance tasks involve an average technician presence of 3.8 hours.

The longest repair times are observed for the drive train system, rotor system, generator transformer system, and converter system. While extended repair durations are expected for subsystems containing major components, their overall impact on turbine availability remains limited due to relatively low failure rates in most cases. In contrast, the power converter system has a substantial effect on availability, as it exhibits both a high failure rate and prolonged average repair time.

7.4.4.4 Average number of technicians required

Similarly to the average repair time, the average number of technicians required per maintenance intervention for each subsystem is shown in **Table 17**. This value represents the mean number of technicians who recorded working hours on the WT or were listed in maintenance records. However, this information was available for only half of the WTs in the offshore dataset, resulting in a reduced sample size for analysis. Consequently, the dataset is insufficient to provide specific figures for MCR beyond the overall averages for all failure events. As a result, the variation in technician requirements across subsystems is relatively small, ranging from 1.9 technicians for the common cooling system to 3.5 technicians for the generator transformer system and generator switching system. On average, 2.5 technicians are required for both other corrective maintenance activities and preventive maintenance interventions.

7.4.5 Comparison with results from literature

Although a direct comparison with existing literature is not feasible due to variations in turbine sizes, technologies, and generations considered in different studies, this section aims to contextualise the findings of this paper within the existing body of reliability and O&M research. For offshore WTs, studies by the University of Strathclyde [12] and SPARTA ([29], [13]) are referenced, while for onshore turbines, comparisons are drawn with findings from WMEP [11], ReliaWind [10] and AWESOME [28]. However, direct comparisons remain challenging due to differences in categorisation systems and variations in KPI definitions. For example, Carroll et al. report annual failure rates, whereas SPARTA provides monthly repair rates. This shows that the definition of failure itself varies across studies. [30] emphasise that such differences in failure definitions in field-data-based studies significantly impact the reported KPI values. Despite these challenges, a general comparison remains valuable to place our results in the context of other research work.

For onshore WTs, an overall average failure rate of 1.729 failures per MW and year has been determined in the present study, with the pitch system, control system, drive train system, and converter system identified as the most critical subsystems. Similar findings were reported by [11], who calculated an annual failure rate of 2.4 failures per WT – consistent with the smaller rated capacities of the turbines in their study – while also highlighting the electrical and control systems as particularly critical. Although [10] reported only normalised failure rates, their findings similarly identified the power module (including power converter, generator, transformer and switchgears), rotor module (including pitch system, blades and hub), control system, and drive train system among the five most frequently failing subsystems. In contrast, [28] highlight the gearbox, the blades, the blade brake, generator, and controller as most critical, while reporting lower normalised failure rates for the pitch system and the frequency converter.

For offshore WTs, an annual average failure rate of 1.088 per MW have been determined in this study. [12] reported approximately 8.3 failures per turbine per year, including major component replacements, as well as major and minor repairs, for turbines with rated capacities between 2 and 4 MW. Transforming the findings of our study to a 3 MW turbine results in an estimated 3.3 failures per turbine and year, which appears significantly lower. However, considering discrepancies in failure definitions and incorporating the additional 1.651 interventions per MW per year associated with corrective maintenance interventions beyond component failures, the estimated corrective maintenance rate reaches approximately 8.2 for a 3 MW turbine – closely aligning with the figures reported by [12]. This highlights the substantial impact that failure definitions and the inclusion criteria for corrective maintenance activities have on reported failure rates.

Regarding the most failure-prone subsystems of offshore WTs, the pitch system, control system, and converter system have emerged as critical in our study, consistent with the top four failing subsystems identified in the [29] report. Similarly, [12] highlighted the pitch system as a major contributor to failure events. Furthermore, significant differences in annual failure rates were observed across different OEMs, a finding also noted by [13] when comparing forced outages per turbine between two OEMs for selected subsystems.

The analysis has also revealed variations in failure behaviour over time, with the overall WT system following the characteristic bathtub curve. At the same time, different subsystems exhibit different failure trends. [11] reported a similar trend for overall onshore WTs. [13] assessed temporal patterns for repairs for specific components and subsystems not directly comparable with failure events and their trends evaluated within this study. The increasing repair rate observed for the generator in the SPARTA evaluation aligns with the trends found for the power generation system (MKA) in the present study, whereas other subsystems are not directly comparable due to differences in component classification.

Regarding major component replacements, both this study and [12] identified the power generation system and drive train system as the primary contributors to JUV interventions. Finally, reported average repair times and the number of technicians required for replacements were compared with findings from [12]. Repair times in this study were generally lower than those reported by Carroll et al., even when compared with Carroll's "minor repairs" category, which primarily includes small spare parts driving overall failure rates. On average, 2.8 technicians were required per replacement according to our results, which is in the same range as the figures reported in [12].

Table 17. Input parameters for O&M simulation, including average failure rates for onshore and offshore wind turbines with 90% confidence interval bounds, average offshore major component replacement (MCR) rates, rates of MCR requiring a jack-up vessel (JUV), average number of technicians required, average repair times, corrective maintenance rate (excluding failures) and preventive maintenance rate for the overall wind turbine

Subsystem		RDS-PP Code	Average Failure Rate Onshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Failure Rate Offshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Rate MCR Offshore	Average Rate MCR JUV required	Average technicians required	Average repair time	Corrective Maintenance Rate (excl. Failures)	Preventive Maintenance Rate
		Unit	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ (MW*a)	1/ WT visit	h/ WT visit	1/ (MW*a)	1/ (MW*a)
		Onshore / offshore	onshore	onshore	onshore	offshore	offshore	offshore	offshore	offshore	offshore	offshore	offshore	offshore
Environmental Measuring System		**CKJ**	0.0590	0.0534	0.0652	0.0272	0.0240	0.0307			2.8	2.6		
Rotor System incl. Pitch System	**Rotor System**	**MDA**	0.0773	0.0708	0.0843	0.0379	0.0341	0.0420	0.0006	0.0006	3.2	4.8		
	Pitch System		0.3138	0.3006	0.3275	0.1638	0.1559	0.1721			3.2	3.3		
Drive Train System		**MDK**	0.2281	0.2168	0.2398	0.0921	0.0861	0.0983	0.0149	0.0097	2.9	7.6		
Yaw System		**MDL**	0.0846	0.0778	0.0919	0.0784	0.0730	0.0842			3.0	2.9		
Central Hydraulic System		**MDX**	0.0599	0.0542	0.0660	0.0806	0.0751	0.0865			3.1	2.7		
Control System		**MDY**	0.2548	0.2429	0.2671	0.1268	0.1198	0.1341			3.2	2.7		
Power Generation System		**MKA**	0.1206	0.1124	0.1292	0.0835	0.0779	0.0895	0.0209	0.0012	2.5	2.9		
Generator Switching System		**MSC**	0.0670	0.0610	0.0735	0.0250	0.0219	0.0283			3.5	2.1		

Subsystem	RDS-PP Code	Average Failure Rate Onshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Failure Rate Offshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Rate MCR Offshore	Average Rate MCR JUV required	Average technicians required	Average repair time	Corrective Maintenance Rate (excl. Failures)	Preventive Maintenance Rate
Converter System	**MSE**	0.2226	0.2115	0.2342	0.1243	0.1174	0.1315			3.4	4.2		
Generator Transformer System	**MST**	0.0252	0.0215	0.0293	0.0083	0.0066	0.0103	0.0003	0.0003	3.5	4.3		
Nacelle	**MUD**	0.0267	0.0229	0.0309	0.0202	0.0175	0.0232	0.0000	0.0000	3.3	1.9		
Remote Monitoring System	**MYA**	0.0071	0.0053	0.0095	0.0004	0.0001	0.0011			2.0	2.2		
Tower System	**UMD**	0.0204	0.0171	0.0241	0.0151	0.0128	0.0178	0.0000	0.0000	2.7	2.5		
Personnel Rescue Systems	**WBA**	0.0059	0.0042	0.0081	0.0019	0.0011	0.0030			2.2	1.4		
Fire Extinguishing System	**XGM**	0.0078	0.0058	0.0102	0.0041	0.0029	0.0056			2.9	2.3		
Lifting Gears	**XMM**	0.0246	0.0210	0.0287	0.1255	0.1185	0.1327			2.5	2.8		
Obstacle Warning System	**XSD**	0.0328	0.0286	0.0374	0.0222	0.0194	0.0254			2.6	1.9		
Low Voltage Electrical Main Supply System	**BFA**	0.0006	0.0002	0.0016	0.0243	0.0213	0.0276			2.6	1.2		
Fire Alarm System	**CKA**	0.0044	0.0030	0.0064	0.0155	0.0132	0.0183			3.4	3.0		

Subsystem	RDS-PP Code	Average Failure Rate Onshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Failure Rate Offshore	Lower bound of 90% confidence interval	Upper bound of 90% confidence interval	Average Rate MCR Offshore	Average Rate MCR JUV required	Average technicians required	Average repair time	Corrective Maintenance Rate (excl. Failures)	Preventive Maintenance Rate
Transformer Station	**UAB**	0.0025	0.0015	0.0041									
Equipotential Bonding / Earthing System	**XFB**	0.0029	0.0018	0.0046	0.0049	0.0036	0.0066			2.2	2.6		
Lightning Protection System	**XFC**	0.0025	0.0015	0.0041	0.0096	0.0077	0.0118			2.4	1.3		
Ventilation Systems	**XAM**				0.0029	0.0019	0.0042			2.2	1.8		
Central Lubrication System	**MDV**	0.0218	0.0184	0.0257	0.0015	0.0008	0.0025			2.5	1.9		
Compensation System	**MSS**	0.0084	0.0063	0.0109	0.0004	0.0001	0.0011						
Common Cooling System	**MUR**	0.0078	0.0058	0.0102	0.0026	0.0017	0.0039			1.9	2.2		
Telephone System	**Y**	0.0204	0.0171	0.0241	0.0064	0.0049	0.0082			2.0	0.9		
General / Other	**G**	0.1210	0.1128	0.1296	0.0579	0.0533	0.0629			3.2	2.6		
Wind Turbine (WT) overall	**WT**	1.7286	1.6973	1.7602	1.0875	1.0669	1.1084	0.0366	0.0117	2.8	2.7	1.6508	2.6637

7.5 Conclusions and outlook

This study provides a comprehensive analysis of failure rates for offshore and onshore wind turbines (WTs), as well as repair times and maintenance resource requirements for offshore assets, with a particular focus on subsystem-level reliability. Based on real-world maintenance data from over 1000 onshore and offshore WTs covering more than 4200 operational years, this dataset offers unique diversity, size and recentness when compared to those used in previous reliability studies. The results highlight that while onshore WTs exhibit lower failure rates per turbine and year, their failure rates per megawatt and year are higher compared to offshore WTs. Given the strong dependence of failure rates on the turbines' rated power, further analyses have been conducted based on failure rates per MW and year to ensure comparability. Onshore WTs exhibit an average failure rate of 1.729 failures per MW per year, whereas offshore WTs demonstrate a lower annual average failure rate of 1.088 failures per MW.

The analysis of subsystem-level failure rates has revealed that certain components, such as the pitch system (0.314 vs. 0.164 failures per MW and year), the control system (0.255 vs. 0.127 failures per MW and year), and the converter system (0.223 vs. 0.124 failures per MW and year), contribute disproportionately to overall WT unreliability for both onshore and offshore turbines. While the drive train system exhibited notably high failure rates for onshore WTs, offshore WTs experienced elevated failure rates in the lifting gear system. Particularly the power converter system has been identified as a critical subsystem due to its combination of a high average failure rate and extended repair duration, making it a major factor affecting overall WT availability next to long-lasting replacement campaigns of major components. Additionally, major component replacements (MCR) have been analysed, distinguishing between those requiring a jack-up vessel (JUV) and those that do not. The power generation system and drive train system accounted for the majority of MCRs, with the latter also being responsible for the highest share of JUV-requiring replacements.

The study has also examined failure behaviour through time, demonstrating that the overall WT failure pattern follows the well-established bathtub curve, with high early failure rates, a period of stability, and increasing failure rates due to degradation in later years of turbine operation. However, subsystem-specific trends vary, with some following the same pattern as the overall WT and others dominated by degradation failures or displaying no clear trend.

In addition to failure rates, i.e. the frequency of corrective measures including spare-part consumption, corrective maintenance interventions without spare-part use and preventive maintenance tasks have also been analysed. On average, 2.7 hours are required for component replacements, while other corrective maintenance and preventive maintenance activities take 1.5 hours and 3.8 hours, respectively. The number of technicians required per maintenance intervention varies by subsystem, ranging from 1.8 to 3.5 technicians, with an overall average of 2.5 technicians per other corrective and preventive maintenance task. While major component failures have significant repair times, their relatively low failure rates limit their impact on availability. In contrast, frequently failing subsystems such as the power converter system have a substantial influence on turbine performance and should be prioritised in reliability-driven design improvements.

Our findings emphasise the importance of detailed, subsystem-level reliability analyses to enhance the accuracy of O&M simulations and operational expenditure (OPEX) calculations.

Aggregated failure rates derived from mixed turbine fleets may obscure critical differences in reliability between turbine types, underscoring the necessity of subgroup-specific analyses. At the same time, the coverage of a variety of WT types and manufacturers is an important prerequisite for providing representative results.

Ultimately, this study underscores the complexity of WT reliability and maintenance planning, highlighting the need for continued field-data based analysis to optimise O&M strategies and improve the long-term sustainability of wind energy operations. Future research will extend beyond basic failure rate calculations to develop advanced reliability models that capture temporal trends in failure behaviour and quantify the effect of various factors on reliability, including design aspects and operating conditions.

7.6 Acknowledgements

The present work was mostly carried out within the research project "Reduction of uncertainties for continued operation of offshore wind farms combining reliability and yield analysis (RUN25+)" funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant number 03EE3106. The provision of comprehensive field data by project partners is gratefully acknowledged. Further financial support was received by EPSRC through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1.

8 Reliability of electrical and hydraulic pitch systems in wind turbines based on field-data analysis

The pitch system is notably one of the critical subsystems of a wind turbine, supporting its effective control towards maximising wind capture and at the same time protecting its integrity in cases of excessive loads. A pitching mechanism is also responsible for operational downtime, hence its reliability performance needs to be carefully evaluated so as to ensure operational availability. This chapter aims to derive failure rates of two configurations of pitch systems, namely the electrical and hydraulic, based on statistical analysis of a large population of onshore assets, followed by a classification of findings by turbine rating, effect of seasonality, and reliability performance of different manufacturers. The datasets underlying the present analysis are based on maintenance reports and comprise 1847 operational years of wind turbines with electrical and 848 operational years of turbines with hydraulic pitch system. Results of this study show high failure rates in pitch systems of both types, with hydraulic systems performing slightly lower than electrical (0.54 than 0.56 failures per turbine per year), a significant variation between turbines of different manufacturers, and a tendency for higher failure rates for larger turbines. The material of this chapter has been peer reviewed and published in [5].

8.1 Introduction

With increasing deployment of wind energy and especially with the rapid development of offshore wind farms, it is crucial to reduce operations and maintenance (O&M) costs of wind turbines. Since O&M costs sum up to 25% to 40% of levelized cost of energy (LCoE), reliability is one of the main levers for further LCoE reduction [18] [19].

Several reliability studies have been conducted in the past. A comprehensive overview of available reliability data is given in [32] and [31]. The pitch system has been identified as one of the most critical sub-systems of a wind turbine (WT) in regards to failure rate and downtime, see e.g. [10], [12] and [29]. The RELIAWIND project analysed a dataset covering 373 WTs with 1115 operational years and found the pitch system to be the main contributor to the overall failure rate of the WTs with 22% [10]. In addition, within the project a failure modes effects and criticality analysis (FMECA) was performed in order to determine the most important failure modes of the critical sub-systems. Carroll et al. published a reliability study analysing around 350 offshore WTs with over 1768 WT years of operation and found the sub-system “pitch / hydraulics” to stand out with most failures per WT per year [12]. The System Performance, Availability and Reliability Trend Analysis (SPARTA) initiative identified the blade adjustment system with the second highest monthly repair rate analysing 1045 offshore WTs located in UK waters [29]. In comparison, a study from Moog and DNV GL conducted a specific pitch system failure analysis including electrical and hydraulic pitch systems and

[5] Julia Walgern, Katharina Fischer, Paul Hentschel, Athanasios Kolios, 2023. „Reliability of electrical and hydraulic pitch systems in wind turbines based on field-data analysis”. Energy Reports, 9, 3273-3281, doi: 10.1016/j.egyr.2023.02.007 [105]

failure rates for different subsets of data were determined based on a data base of 1330 WTs from North America, Europe and China [35]. However, most of those studies used data which had been recorded before 2010. Moreover, neither these system-level studies nor the pitch-system specific study presented by [35] differentiate between hydraulic and electrical pitch systems or analyse underlying failure patterns and related failure rates of the pitch systems' components. At the same time, understanding which failure modes drive the failure rate is key to develop countermeasures. Therefore, this work presents a deepened reliability analysis of both electrical and hydraulic pitch systems including failure rates of the respective components. Additionally, temporal patterns are investigated to gain further insights into the failure behaviour. Outcomes of this work will be of value to further researchers and practitioners who aim to evaluate and optimise design and operational management of wind turbines, as well as for supporting further technological improvements of next generation pitch systems. Obtained failure rates can be utilised for O&M simulation tools such as the Operation and Maintenance Cost Estimator (OMCE) of the Energy Research Center of the Netherlands (ECN) [14], the Norwegian Offshore Wind cost and benefit (NOWIcob) tool presented by [15], the openO&M tool [182], or OffshoreTimes, a simulation tool developed by the Fraunhofer Institute for Wind Energy Systems IWES [17].

Modern WTs use pitch regulation to control operations. Pitch systems allow changing the blade pitch angle dependent on incoming wind speed. From cut-in wind speed to rated wind speed, the pitch angle is adjusted actively so that optimal power output is achieved. From rated wind speed onwards, power production is limited by rotating the rotor blades out of the wind. Therefore, the pitch system is not only responsible for maximising power output but also functions as an aerodynamic break. Due to safety requirements, there is a pitch system for each blade axis and the systems are entirely independent. There are electrical and hydraulic pitch systems: Electrical pitch systems can be divided into AC or DC systems which drive the pitch motor. In case of interruption of voltage supply, batteries feed the system to guarantee that the WT can be stopped by pitching the blades out of the wind. In comparison, hydraulic pitch systems are driven by hydraulic cylinders. Additional components ensuring its operation are hydraulic valves, accumulator units and oil tanks. A further description of both systems can be found in [183].

The chapter is outlined as follows: First, an introduction of the used methods is given and the analysed dataset is described (Section 8.2). Afterwards, the paper presents findings of a deepened statistical analysis for WTs with electrical and hydraulic pitch systems and compares those with previously published results of field-data analysis summarised above. Next to failure rates of the pitch systems' components for different subsets, seasonal patterns are evaluated (Section 8.3). Last, a summary of main conclusions as well as an outlook to future work are given (Section 8.4).

8.2 Methodology and datasets

8.2.1 Methodology

Within this study, a failure is defined as a fault that leads to downtime of the wind turbine and is not resettable remotely but requires maintenance and the use of spare parts. In case repeated maintenance activities are needed to resolve the same technical problem, the activities are assigned to one failure event. The failed components are classified using the reference designation system RDS-PP for wind turbines [40]. From all maintenance measures recorded for a wind turbine, only maintenance interventions which are related to the pitch system are analysed in this study.

In order to compare the reliability of different components, their average failure rates are calculated as follows:

$$f = \frac{\sum_{i=1}^{I} N_i}{\sum_{i=1}^{I} X_i T_i} = \frac{N}{T} \tag{8.1}$$

Herein, N_i is the number of failures of the analysed component in the time interval i, X_i is the number of WTs considered in this time interval and T_i is the duration of the time interval. Therefore, the average failure rate is equal to the quotient of the sum of all failures N and the total amount of analysed WT operational years T.

Additionally, the corresponding confidence intervals are determined to quantify the uncertainty of the calculated failure rates resulting from the size of the datasets [179] [36]:

$$\left[\frac{\chi^2(\frac{\alpha}{2}, 2N)}{2T}, \frac{\chi^2(1-\frac{\alpha}{2}, 2N+2)}{2T}\right] \tag{8.2}$$

Herein, $\chi^2(\alpha/2, 2N)$ is the $(\alpha/2)$-quantile of the χ^2 distribution with $2N$ degrees of freedom. In this chapter $\alpha = 0.1$ is used so that the 90% confidence intervals are provided. These can be interpreted as follows: If a large number of samples (in this case failure datasets of WTs) would be analysed, in 90% of the cases the given confidence intervals would cover the real value of the failure rate.

8.2.2 Datasets

The datasets underlying the present analysis are based on maintenance reports and comprise 1847 operational years of WTs with electrical and 848 operational years of WTs with hydraulic pitch system. All WTs are located onshore. Detailed information about the datasets is presented in **Table 18**. While for the electrical pitch system data from turbines of six different original equipment manufacturers (OEMs) with turbine capacities ranging from 500 to 6000 kW are considered, for the hydraulic pitch system data from turbines of three different manufacturers with capacities from 600 to 3000 kW are evaluated. A total number of 2695 WT operational years stemming from 1022 WTs is underlying the present study.

Table 18. Information about the datasets which have been considered in the analysis

	Electrical pitch system	**Hydraulic pitch system**
WT operational years considered	1847	848
Number of WT OEMs covered	6	3
Rated capacity considered	500-6000 kW	600-3000 kW
Available failure data period	2006-2015	2013-2017

8.3 Results and discussion

8.3.1 Comparison of failure rates for hydraulic and electrical pitch systems

8.3.1.1 Electrical and hydraulic pitch system comparison

Figure 34 and **Figure 35** present the resulting component failure rates along with the overall pitch-system failure rates for the electrical and the hydraulic pitch systems, respectively. For the presentation of results, component categories are chosen based on frequency of failure and level of detail of the available maintenance reports. All pitch-system components that do not fail often and are of no specific interest for the analysis are summarised in "Other Components". Note that the sum of the component failure rates is higher than the overall failure rate of the system, as there are failure events involving the exchange of components from several categories.

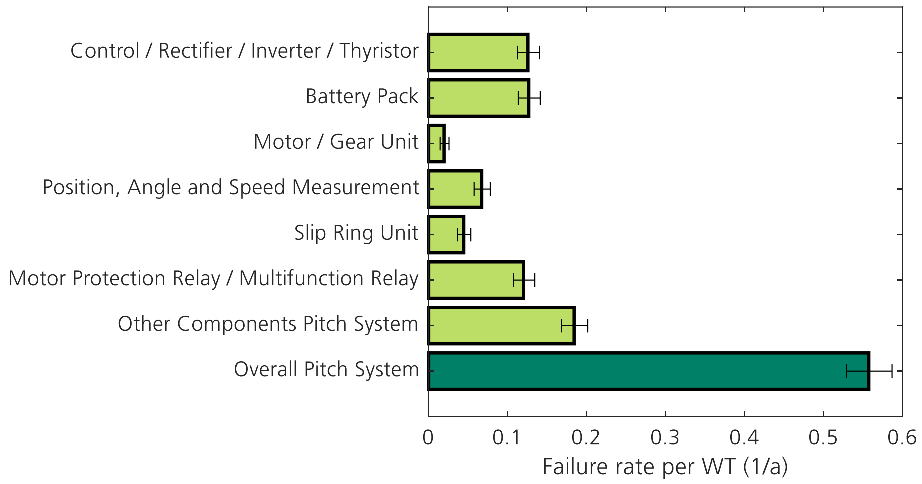

Figure 34. Average failure rates of the electrical pitch system

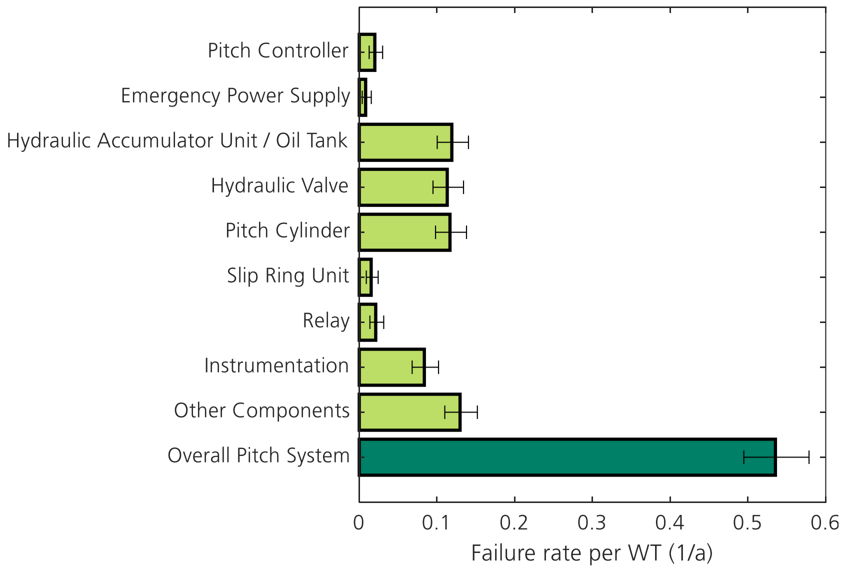

Figure 35. Average failure rates of the hydraulic pitch system

The results confirm the occurrence of high failure rates in pitch systems of both types. With 0.54 failures per WT and year, the overall failure rate of the hydraulic pitch system is slightly lower than that of the electrical pitch system with 0.56 failures per WT per year. However, due to overlapping confidence intervals, there is not sufficient evidence to conclude that hydraulic pitch systems are more reliable than electrical pitch systems. While for the electrical pitch system the component categories “Battery Pack”, “Control / Rectifier / Inverter / Thyristor” and “Motor Protection Relay / Multifunction Relay” are identified as most critical, the hydraulic pitch system shows the highest failure rates in the component categories “Hydraulic Accumulator Unit / Oil Tank”, “Pitch Cylinder” and “Hydraulic Valve”. This highlights that main concerns of the hydraulic pitch system are related to the hydraulic system itself. An interesting finding in the context of electrical pitch systems is that, in contrast to the main power converters of WTs where power electronics are subject to frequent failure (cf. [184], [36]), they are only a minor contributor to failure of pitch systems. A possible explanation for that could be the significantly lower rated power of the pitch drives. A more sophisticated design better withstanding the harsh environmental conditions could be another reason.

8.3.1.2 Comparison with literature

When comparing those results with the reliability studies mentioned in the introduction, similarities can be identified. While the RELIAWIND project provides only normalized failure rates, the study by [35] found an average failure rate of 0.7 per WT per year for a combined dataset of 545 WTs with electrical and 785 WTs with hydraulic pitch system all being installed onshore. This number is slightly higher in comparison to the average failure rates presented above even when considering the confidence intervals. Also [12] have identified a higher failure rate of 1.076 failures per WT per year. However, the comparison can only be made with caution since Carroll et al. used a sub-system category which combines the pitch system with all other hydraulic components within a turbine since only hydraulic pitch systems were analysed. In the RELIAWIND project, a FMECA was conducted identifying the top five failure modes of the critical sub-systems of which the pitch system has been one. For the electrical pitch system, “battery failure”, “pitch motor failure” and “pitch motor converter failure” were mentioned as most important failure modes, whereas for the hydraulic pitch system different kinds of leakages were described as top three failure modes [10]. [12] described oil issues, valve issues and accumulator problems as the most common failure modes in the component category “pitch / hydraulic”. Furthermore, based on the quantitative study of [12], [185] performed a case study for fluid power pitch systems in which a fault tree analysis (FTA) and FMECA revealed valves and accumulators as most critical components. Those findings are in line with the components’ average failure rates shown in **Figure 34** and **Figure 35**. Even though criticality of the faults cannot be derived directly from the presented results, all component failures need to be considered as critical since the pitch system is part of the emergency shut down system of turbines. Additionally, instead of giving just a ranking for top failure modes, in this study the determined average failure rates of the components can be used to quantify to which extent a certain component drives the overall failure rate of the pitch system analysed.

8.3.2 Failure-rate comparison across WT OEMs

While it is common practice to provide average failure rates calculated from mixed fleets containing different types of turbines as in Section 3.1, this practice is afflicted with risks: One is that certain WT types with a particularly low or high reliability level might bias the result. Another one is that providing only a group-averaged failure rate without further elaboration masks such reliability differences that are important indicators to trigger root-cause analysis and design improvements. Therefore, more detailed analyses based on subgroups of turbines are presented in the following.

8.3.2.1 Electrical pitch systems

Further analyses have shown that there are significant differences in pitch system failure rates when comparing failure rates across different WT manufacturers or when clustering WTs according to their rated power. **Figure 36** shows the results for average failure rates for the electrical pitch system and its components for three different OEM categories. The dataset of the category OEM1 is characterised by an average rated power of 1511 kW and 1011 operational years analysed in which 850 failure events have been recorded, thus having an average failure rate of 0.84 per WT per year. The analysis for OEM2 is based on 700 operational years with an average rated capacity of 1686 kW and results with 149 logged failures in a lower failure rate of only 0.21 per WT per year for the overall pitch system. The last analysis combines the failure events of four different OEMs as the datasets available for each OEM separately would have been too small for sufficient interpretation of the results. Therefore, the last category OEM3-6 includes 137 operational years of WTs from four different OEMs with 29 failures recorded. The WTs within this dataset have an average rated capacity of 1996 kW. Even though the confidence interval is slightly larger due to the smaller dataset evaluated, also for this category there is an average failure rate of 0.21 per WT per year. Moreover, it can be noted that the distribution of most contributing components varies slightly dependent on the OEM. While for OEM1 the component category "Motor Protection Relays / Multifunction Relays" plays a significant role, this is not the case for OEM2 and OEM3-6.

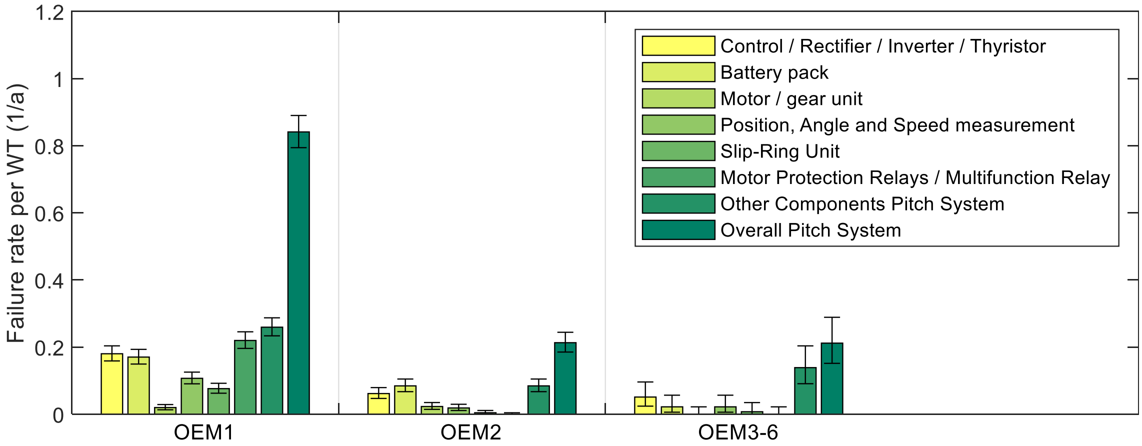

Figure 36. Failure-rate comparison across WT OEMs for electrical pitch systems

8.3.2.2 Hydraulic pitch systems

The same evaluation has been conducted for the hydraulic pitch system. In this case three OEMs are compared. Results are presented in **Figure 37**. The data-subset of OEM1 comprises 118 operational years. With 87 failures counted, an average failure rate of 0.74 per WT per year for the overall pitch system is calculated. The WTs within this subset can be characterised by an average rated capacity of 1420 kW. The category OEM2 contains 552 operational years and WTs within this subset have an average rated capacity of 2004 kW. With 312 failures noted in this period, a lower average failure rate than for OEM1 of 0.57 per WT per year is determined. The analysis for OEM3 is based on 178 operational years and the subset has an average rated capacity of 1226 kW. For this subset 55 failure events have been recorded resulting in an average failure rate of 0.31 per WT per year. Comparing the failure rates of the components it can be noted that the component categories "Hydraulic Valve" and "Pitch Cylinder" have higher failure rates for WTs of OEM2 whereas the overall pitch system failure rate of OEM1 is driven by the component category "Instrumentation". In comparison to OEM1 and OEM2, the component category "Hydraulic Valve" is the only component category that plays a significant role for OEM3.

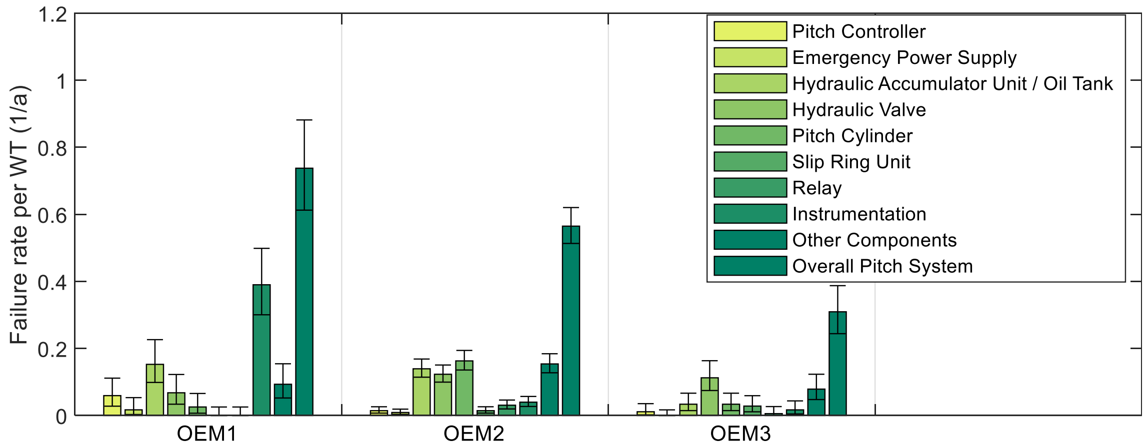

Figure 37. Failure-rate comparison across WT OEMs for hydraulic pitch systems

8.3.3 Failure-rate comparison: Role of WT rated power

8.3.3.1 Electrical and hydraulic pitch system

In a next step, the pitch-system failure rates of WTs with different ranges of rated capacity are compared. In order to ensure comparability, a data-subset is chosen in which only one OEM is considered, and which allows for splitting the available failure data in different capacity classes.

For the electrical pitch system this is only the case for OEM2. **Figure 38** shows the failure-rate comparison for this case. Because of using a subset for this evaluation, the number of operational years considered is reduced to 164 which leads to larger confidence intervals.

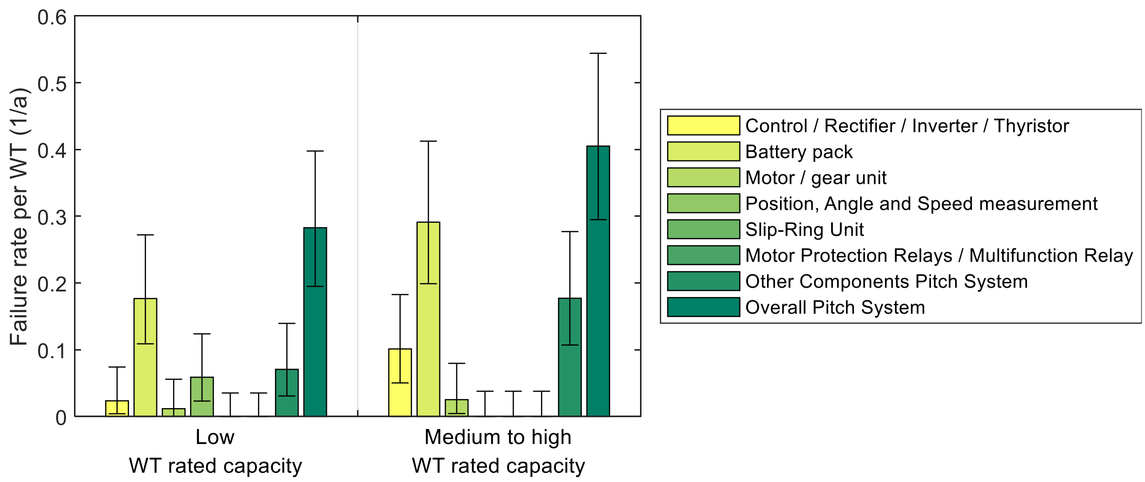

Figure 38. Failure-rate comparison for WTs with different categories of rated power for electric pitch systems

The category "low WT rated capacity" comprises WTs with rated capacities below 1500 kW, whereas the category "medium to high WT rated capacity" contains WTs ranging from 1500 kW to 6000 kW.

A similar analysis is performed for two subsets with WTs with a hydraulic pitch system. Results can be seen in **Figure 39** and **Figure 40**. For the first case the comparison is made for a data-subset comprising 112 operational years since only failure events of OEM1 are considered for comparability reasons. While the category “low WT rated capacity” contains WTs with rated capacity below 1500 kW as for the electrical pitch system, the category “medium WT rated capacity” consists of WTs ranging from 1500 kW to 2500 kW. The second case analyses a data-subset containing WTs of OEM3 which considers 178 operational years. The category “low WT rated capacity” describes WTs with rated capacity below 1500 kW, whereas the category “medium WT rated capacity” comprises WTs with rated capacity ranging from 1500 kW to 2500 kW as for OEM1.

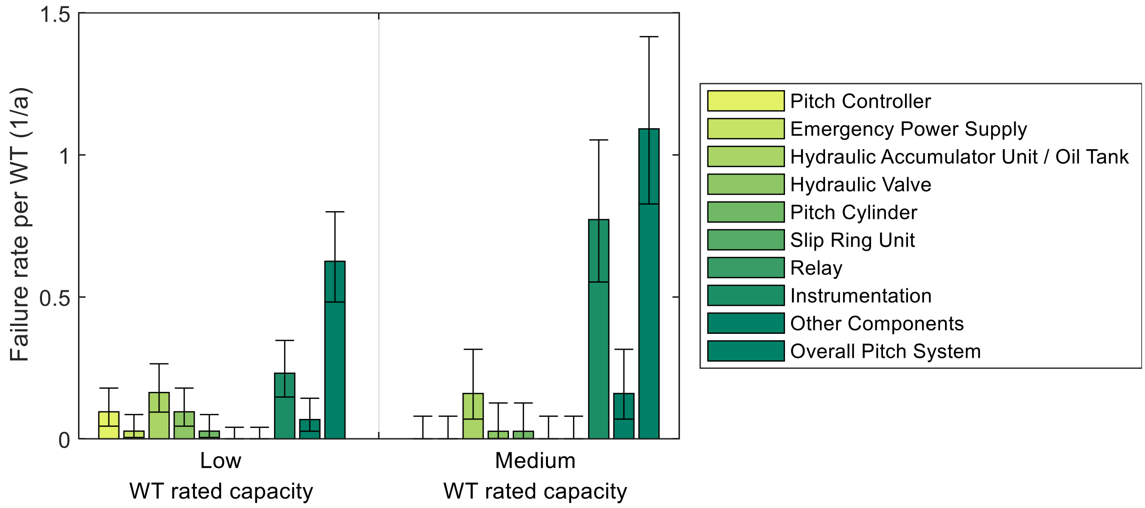

Figure 39. Failure-rate comparison for WTs of OEM1 with different categories of rated power for hydraulic pitch systems

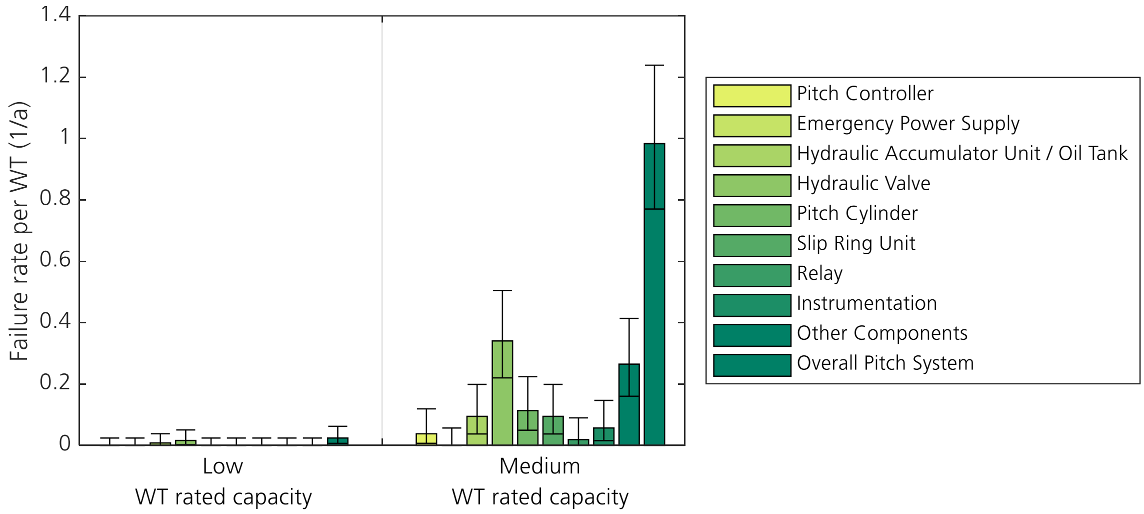

Figure 40. Failure-rate comparison for WTs of OEM3 with different categories of rated power for hydraulic pitch systems

While there is not in all cases clear evidence due to the overlapping confidence intervals, a trend of failure rates increasing with the WT rated power can be observed both for the hydraulic and the electrical pitch systems. Besides the component category “Position, Angle and Speed measurement”, there is a trend for all other components of the electrical pitch system failing more often in larger turbines as well. For the components of the hydraulic pitch system of OEM1 no clear trend can be observed. Comparing the two categories of OEM3 for the hydraulic pitch system, a distinct tendency of higher failure rates for WTs with higher rated power can be seen. However, it has to be noted that only three failure events for the category with low WT rated capacity have been recorded within 125 operational years resulting in the low average failure rate. Considering this, it becomes clear that the lower average failure rate of OEM3 in comparison to OEM1 and OEM2 in **Figure 37** is mainly driven by the majority of small WTs being represented in the data-subset of OEM3.

8.3.3.2 Comparison with literature

The findings above can be compared with the study of [35] which also differentiated into two categories of turbine sizes. One turbine class was defined with rated power ranging from 1.5 MW to 2.5 MW, and the other turbine class included WTs with a rated capacity between 2.5 MW and 3 MW. The same trend was identified: The larger the turbine, the greater the failure rate of the pitch system. However, the failure rate of 1.6 failures per WT per year obtained for the larger turbine class differs from the ones in this study. While for the electrical pitch system significantly lower failure rates are found (compare **Figure 38**), the upper boundary of the confidence interval of the hydraulic pitch system for medium WT rated capacity differs only slightly (compare **Figure 39**). Since no confidence intervals are presented in the study of [35], it is difficult to judge how the smaller dataset affects the calculated failure rate.

8.3.4 Seasonal patterns in the failure behaviour

Next to comparing failure rates under consideration of design factors (OEM, size of turbine, type of pitch system), it is evaluated if any seasonal patterns can be identified in the failure behaviour. For this purpose, component failure rates are calculated for each month. In order to allow a comparison of the failure behaviour with the environmental conditions the WTs have been exposed to, monthly averaged wind, temperature and humidity conditions derived from ERA5 reanalysis data are included for each evaluated wind farm. (ERA5 provides hourly estimates of a variety of atmospheric and oceanographic variables based on global modal data. The dataset covers the earth on a grid of approximately 30 km x 30 km. For detailed information on the ERA5 reanalysis data, please refer to [59] [186], for information on how this data is processed to [187]).

Figure 41 shows the component failure rates through the year and respective ERA5 data from the same wind farms and time periods failure data has been available for. Each line indicates the environmental conditions at one analysed wind farm location. Results for the electrical pitch system are shown on the left and for the hydraulic pitch on the right side, respectively. It can be observed that the WTs operate within similar climatic conditions whereas the wind characteristics for each site can differ significantly. Looking at the overall pitch-system failure rates, no pronounced seasonal patterns can be identified.

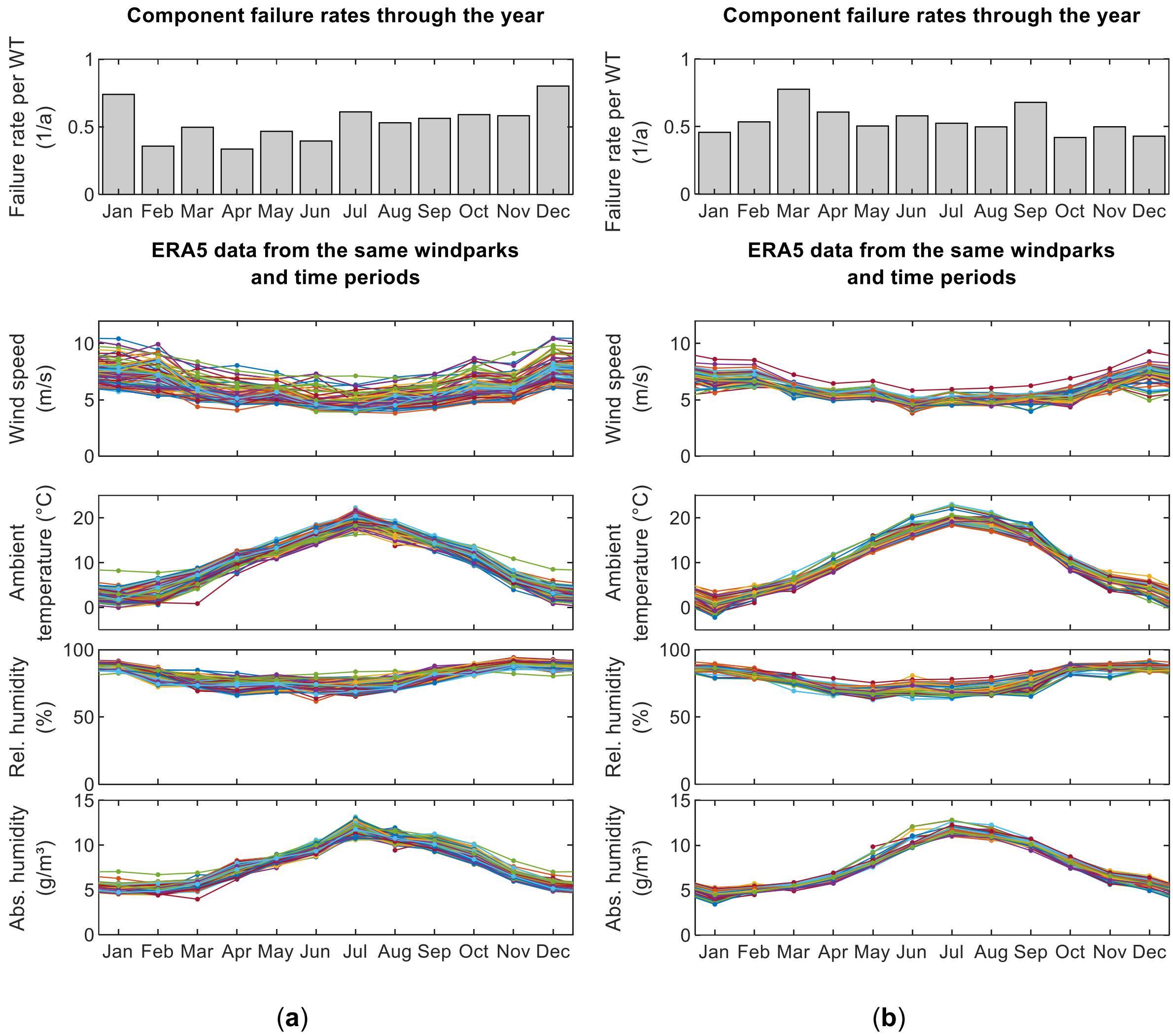

Figure 41. Average failure rates through the year and respective ERA5 data from the same wind farms and time periods. (a) Electrical pitch system. (b) Hydraulic pitch system.

When the same analysis is repeated for specific components of the electrical pitch system, the situation looks different. **Figure 42** presents component failure rates through the year for selected components for which seasonal patterns can be identified:

The battery packs have higher failure rates from September to January in comparison to the summer months. This could partially be related to low ambient temperatures (compare **Figure 41**). The colder it is, the lower is the battery voltage and the higher the probability that required minimum voltage values are not met anymore. Consequently, the battery pack needs to be replaced.

Motor protection relays and multifunction relays show two different trends. On the one hand, failure rates are higher from July to October, which could be explained with a correlation with higher temperature and absolute humidity. On the other hand, there are peaks in the winter months of December and January, which likely have a different cause.

Slip ring units are found to have the highest failure rates in December and January. Those are the months with highest average wind speed but also low temperatures (compare **Figure 41**). A correlation with high wind speeds could be explained with more pitch activity and possibly increased friction related to the higher main-shaft speed during operation at or close to rated power. Consequently, the slip ring unit faces increased wear and needs to be replaced more often.

In comparison to the three components mentioned above, electronic and power-electronic components of the electrical pitch system, namely control, rectifier, inverter and thyristor, do not exhibit any seasonal clusters. This is an interesting finding since for power electronics in main power converters of WTs pronounced seasonal patterns have been reported in [163], which could be related to their climatic operating conditions in [187].

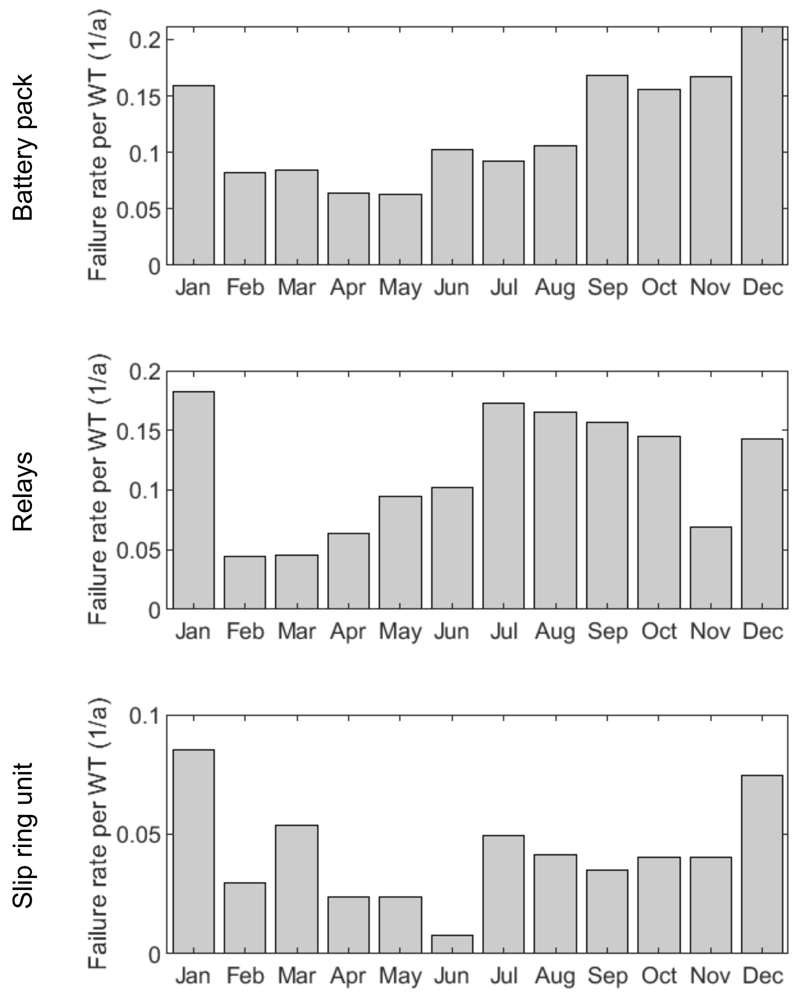

Figure 42. Component failure rates through the year for different components of the electrical pitch system

The same evaluation is conducted for the hydraulic pitch system and its components. **Figure 43** shows component failure rates through the year for selected components. No pronounced seasonal patterns are found for components of the hydraulic pitch system. This can partially be related to the fact that the number of operational years covered by this data-subset is smaller in comparison to the one of the electrical pitch system. Especially for components with small average failure rates (compare **Figure 35**) only a few failure events have been recorded. Therefore, patterns are more difficult to identify. On the contrary, also components with higher average failure rates (see **Figure 43**) do not show seasonal clusters. Solely a peak in September can be observed for the component categories “Hydraulic Valve” and “Pitch Cylinder”. However, there is no evident reason found.

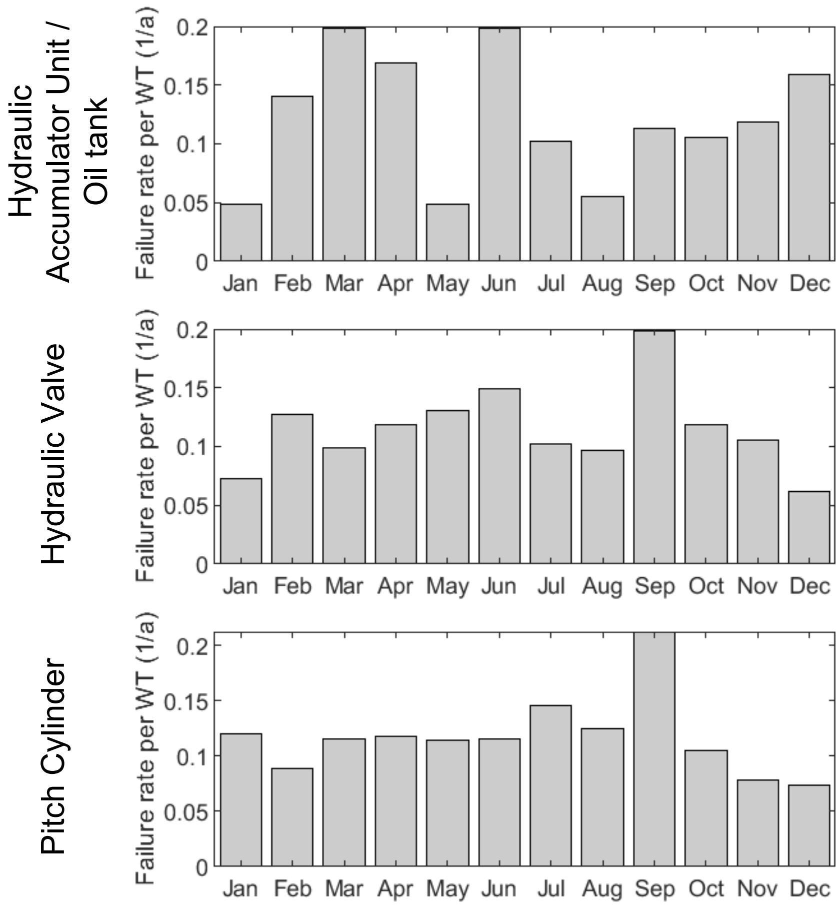

Figure 43. Component failure rates through the year for different components of the hydraulic pitch system

8.4 Conclusions and outlook

This chapter has investigated the reliability performance of electrical and hydraulic pitch systems based on a large population of wind turbines with the objective to derive representative failure rates for the overall populations and evaluate the impact of certain parameters to the failure rate values. This can be utilised for more representative availability assessments, optimisation of operational strategies or prioritisation of design improvements. As the study has been performed on a larger dataset than any previous study, its results can be considered more representative. Findings of the study can be summarised as follows:

- Failure rates are high in pitch systems of both types, with hydraulic systems performing slightly lower than electrical (0.54 than 0.56 failures per WT per year). However, due to overlapping confidence intervals, there is no sufficient evidence to conclude that hydraulic pitch systems are more reliable than electrical ones.
- Among the different OEMs comprised by the dataset, the failure rates have been found to differ significantly depending on OEM, and hence technology.
- The classification of rating to low and medium-high capacity has indicated that the failure rates of the overall pitch system tend to increase with the WT rated power.

- While for the electrical pitch system the component categories “Battery Pack”, “Control / Rectifier / Inverter / Thyristor” and “Motor Protection Relay / Multifunction Relay” have been identified as most critical, the hydraulic pitch system has shown the highest failure rates in the component categories related to the hydraulic system itself, namely “Hydraulic Accumulator Unit / Oil Tank”, “Pitch Cylinder” and “Hydraulic Valve”.
- Seasonal patterns in the failure behaviour have been found for components of the electrical pitch system but could not be identified for hydraulic pitch system’s components based on the evaluated dataset. As temporal failure patterns typically become more evident with higher numbers of evaluated failures, further investigations with an extended data base are recommendable especially for components with low average failure rates in the future to reveal potential further conclusive correlations with environmental conditions.

8.5 Acknowledgments

The present work was partly carried out within the “Fraunhofer-Innovationscluster Leistungselektronik für regenerative Energieversorgung”. The support by the Federal State of Lower Saxony with funds from “Niedersächsisches Vorab” (grant number VWZN2989) and by Fraunhofer-Gesellschaft as well as the provision of comprehensive field data by the project partners are gratefully acknowledged. ERA5 reanalysis data were obtained from the Copernicus Climate Change and Atmosphere Monitoring Services. Further financial support was received by EPSRC through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1.

9 Medium-voltage versus low-voltage converter reliability in wind turbines: a field-data based study

High failure rates of typical IGBT-based low-voltage converters remain a challenge for wind-turbine reliability. A field-data based reliability study for IGCT-based medium-voltage converters in offshore wind turbines is presented within this chapter. Compared with low-voltage converters, these are found to exhibit lower failure rates per MW of converter capacity and fewer components with susceptibility to climatic influences. The failure behaviour through time of both the medium- and low-voltage converters is mostly characterised by early-failure behaviour in the first years of turbine operation and subsequent reliability deterioration. In case of the medium-voltage converters, these trends are less pronounced so that the reliability behaviour is more stable. Overall, the findings support the potential for wider adoption of medium-voltage converter technology in wind turbines. The material of this chapter is currently under peer review for publication in [6].

9.1 Introduction

Medium-voltage (MV) power converters have been judged an attractive option for application in multi-MW wind turbines (WTs) in the literature for more than a decade (see e.g. [188], [189], [190]). However, the long-awaited large-scale turnover to MV converter technology is still not visible in the field, despite its advantages of reduced current levels and lower number of components. Although MV converters for the wind application and turbine models with MV converters have been available on the market for many years, the vast majority of WTs in operation are equipped with insulated-gate bipolar transistor (IGBT)-based low-voltage (LV) converters. In addition, LV converter technology is still applied in the latest turbine generations of most major manufacturers. With regard to possible causes, a known obstacle of switching to MV converters is the associated need for specially trained personnel. Besides this, the limited track record of MV converter application in WTs might have presented another obstacle. With the further increase of WT rated capacity into the double-digit MW range, the arguments in favour of MV converter technology become more and more important. Since recently, new wind farms using MV converters are being installed and commissioned in both European and US waters.

Using a field-data based approach, comprehensive research at Fraunhofer IWES has been dedicated to analysing the reliability and failure causes of power converters in the wind application in recent years. So far, the research was focused on LV IGBT-based voltage source converters as the prevailing technology applied in WTs (see e.g. [191], [163], [46], [104], [47]). In this contribution, a field-data based reliability study of MV converters in WTs is presented, which we expect to become increasingly applied in future wind farms, and a comparison with LV converters is drawn.

[6] Katharina Fischer, Fraser Anderson, Julia Walgern, 2025. “Medium-Voltage versus Low-Voltage Converter Reliability in Wind Turbines: A Field-Data Based Study”. Submitted to PCIM Europe 2025 for publication [213]

9.2 Evaluated datasets and wind turbine fleets

9.2.1 Medium-voltage converter data

In case of the MV converters, the study is based on converter failure data from offshore wind farms in European waters. The dataset is derived from maintenance records, ranges from 2015 to 2023 and covers in total 1249 WT operating years. All WTs are from the same manufacturer, use a MV-PMSG (permanent-magnet synchronous generator) and have a rated power of 5 MW. Their commissioning dates range from 2015 to 2017. The MV converters are liquid-cooled 3-level (3L) voltage source converters of neutral point clamped (NPC) topology with press-pack integrated gate-commutated thyristors (IGCTs) as power switches, suitable for AC voltage levels up to 4.3 kV, see **Figure 44**.

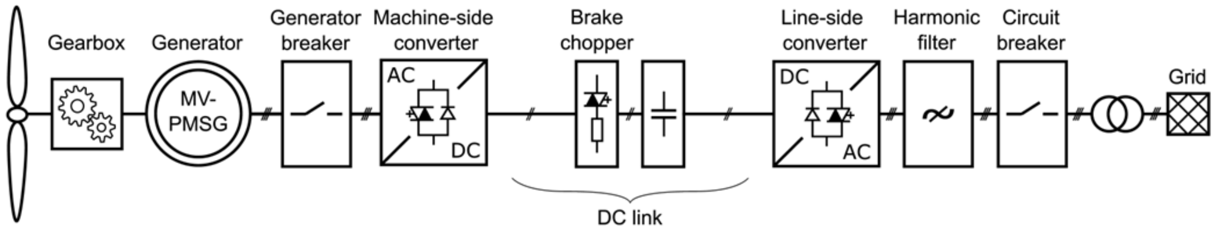

Figure 44. Scheme of the investigated wind turbines with medium-voltage permanent-magnet synchronous generator (MV-PMSG) and fully rated medium-voltage converter

9.2.2 Low-voltage converter data

For better comparability, the results for LV converters presented in this study are also based solely on liquid-cooled LV converters of offshore wind turbines in European waters. The analysis uses data from years 2007-2024 covering in total 1904 years of turbine operation. The evaluated WT fleet consists of turbines of different manufacturers and generator converter concepts, namely doubly fed induction generators (DFIG) with partially rated LV converters, squirrel-cage induction generators (SCIG) with fully rated LV converters, or permanent-magnet synchronous generators (PMSG) with fully rated LV converters (see e.g. [163] for further information about these concepts). The rated power of the evaluated WTs falls into the range from 2 to 9 MW. The WTs have been commissioned from 2007 to 2020. All evaluated LV converters are two-level (2L) IGBT-based voltage source converters, with AC voltage levels below 1000 V (mostly 690 V) and DC-link voltages up to 1250 V.

9.2.3 Site-specific environmental data

In addition to the failure data of MV and LV converters described above, we use site-specific environmental data from all evaluated wind farms, obtained from the publicly available ERA5 reanalysis data ([59], [192]). Of the various meteorological quantities provided on an approx. 30 km x 30 km grid in this dataset, we use the wind speed, ambient temperature and ambient humidity.

9.2.4 Data preparation and processing

To facilitate a script-based analysis, all failure data are preprocessed into a uniform format, using the reference designation system RDS-PP [40] for component classification. Only faults requiring onsite repair and the use of spare parts are counted as failures. Routine maintenance such as refilling of coolant or the exchange of deionisation cartridges in the cooling system are not considered, neither are repairs by means of sealing, retightening, reconnecting or resetting components.

Note that our analysis is based on failure events; i.e., an incident requiring the replacement of e.g. IGCTs (which, in the MV converters investigated here, always involves also the replacement of the corresponding driver boards) is counted as one failure event, regardless of how many IGCTs had to be exchanged and if the issue was remedied during a single maintenance intervention or if several turbine visits were necessary to achieve this. The dates of the failure or the maintenance intervention are used as timestamp of the failure events where this information is available, the booking date of the corresponding maintenance report otherwise.

The failed converter components are grouped into the same categories previously used for analysis of LV converters: 'Phase module' (including IGBT modules and their driver boards in LV converters; IGCTs, their driver boards, and diodes in MV converters; DC-link capacitors and busbars), 'Converter control system', 'Heating & cooling system', 'Main circuit breakers & contactors', and 'Other converter components'. Both datasets underlying the present work include left- and right-censored data.

9.3 Analysis methods and results

Based on the data preprocessed in this way, we derive and present average failure rates for the different component categories of the converter systems, investigate seasonal failure patterns and compare these with the monthly averaged wind speed and environmental conditions at the wind farms. In addition, we analyse the converter failure behaviour through time.

9.3.1 Average failure rates of MV and LV converters

Average failure rates are the most straightforward measure to describe the reliability of a system or component. The most frequently presented type of failure rates indicates the average number of failure events per WT and year. As previous investigations (e.g. [47], [166], [193]) have shown that the frequency of failures typically scales with the size of the converters, it makes sense to calculate average failure rates per MW of rated converter capacity and year whenever reliability comparisons between converter systems of different power classes are intended. Failure rates per WT and year and their corresponding confidence intervals are calculated as described in [191]. They can be transformed to failure rates per rated converter capacity by means of multiplication with a weighted mean of the converter capacity

$$P_{rated,wmean} = \frac{\sum_j^J T_j \cdot P_{rated,j}}{\sum_j^J T_j} \tag{9.1}$$

with T_j denoting the evaluation period and $P_{rated,j}$ the rated converter capacity of wind turbine j. The equation for direct calculation of average failure rates per converter capacity is:

$$f_{MW} = \frac{\sum_j^J N_j}{\sum_j^J T_j \cdot P_{rated,j}} \tag{9.2}$$

Herein, N_j denotes the number of failure events on turbine j during its evaluation period. Note that the rated capacity of the converters is assumed to be equal to the WT rated power in case of PMSG and SCIG turbines with fully rated converters. In case of WTs with DFIG, the rated capacity of the partially rated converter is approximated with 1/3 of the WT rated power.

Figure 45 (a) presents the average failure rates per MW of converter capacity and year for the different component categories of the MV converters. The light bars represent the component failure rates whereas the dark bar indicates the failure rate of the overall converter system. Note that the component failure rates typically sum up to a value larger than the converter-system failure rate since there are failure events affecting components of more than one category.

The overall MV converter system failure rate of 0.075 /MW/year corresponds to 0.375 failures per 5 MW WT and year, i.e. on average one converter failure per year on 37.5% of the WTs. Within the MV converter system, the category 'phase module' stands out with the highest average failure rate of 0.029 /MW/year. Replacements of diodes and IGCTs with their corresponding driver boards account for almost all failure events in this category, whereas DC-link capacitors and busbars play a negligible role. Besides the 'phase module' category, also 'Converter control', 'Heating & cooling system' and the category 'Other converter components' contribute with relevant portions to the converter-system failure rate. Within the heating & cooling system, pressure and temperature sensors, expansion vessels, connectors and pipes / hoses and coolant pumps contribute most to the failures. In the category 'Other converter components', power supply units, DC-link charging units and AC filter components stand out.

Figure 45 (b) shows the corresponding average failure rates in the evaluated WT fleet with LV converters. Differences from previously published average converter failure rates of LV converters as e.g. in [47] can be attributed to the limitation of the present analysis to offshore WTs commissioned from year 2007 onwards. Note that slightly different amounts of field data are underlying the phase-module failure rate and those of the other component categories, as part of the datasets covers only phase-module failures. In summary, the phase-module analysis is based on 1904 WT operating years whereas the failure rates of all other component categories as well as of the overall converter system are derived from a total of 1527 WT operating years. The average rated converter capacity is in the range of 2.5 MW for the evaluated fleet with LV converters.

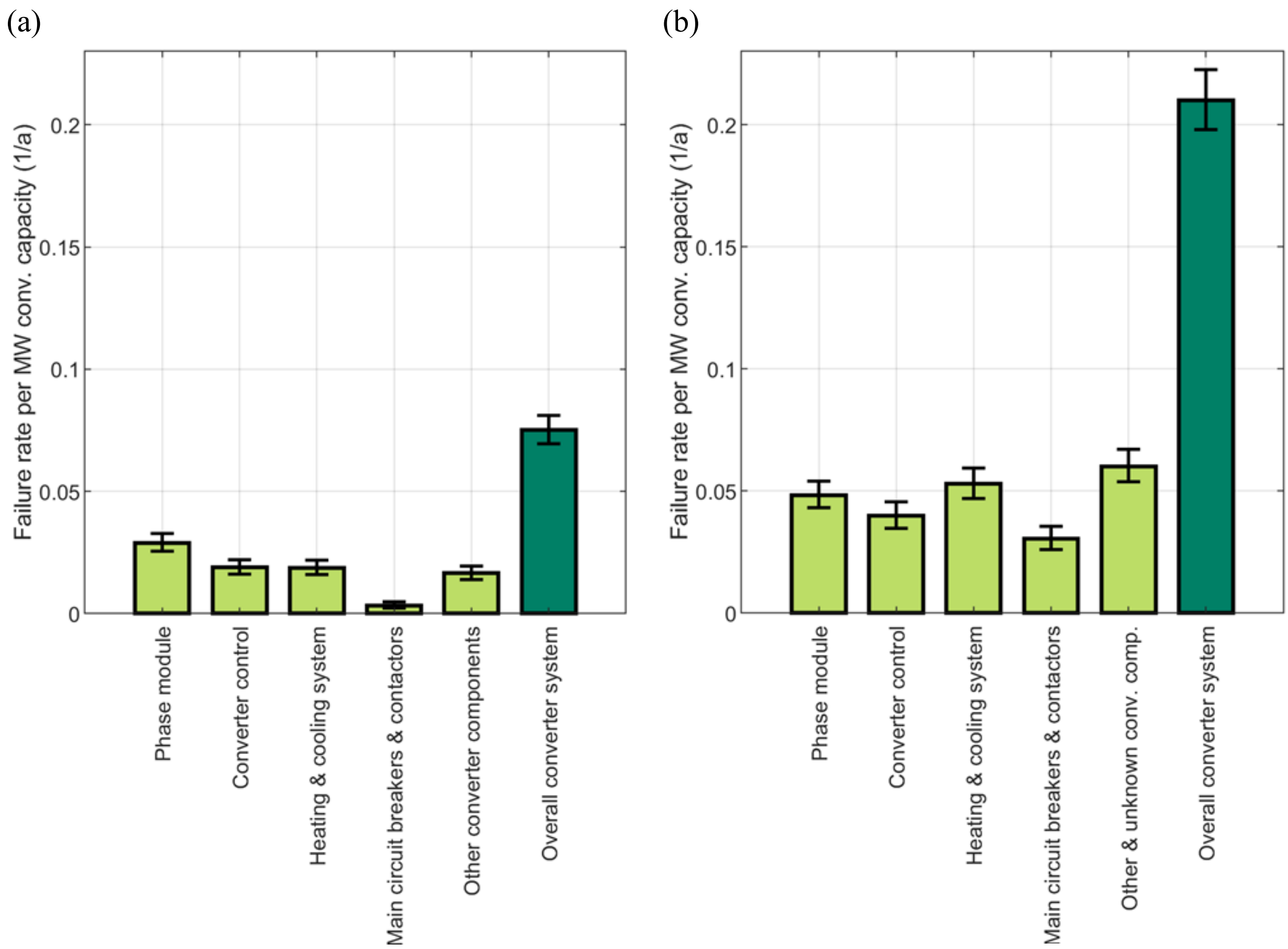

Figure 45. Average failure rates per converter capacity of the overall converter system (dark green) and its components (light green) for medium-voltage converters (a) and for the evaluated low-voltage converters (b)

The comparison between the average failure rates of the investigated MV and LV converters in **Figure 45** reveals a considerable difference in reliability: The failure rates per MW of converter capacity are in all component categories much higher for the LV than for the MV converters. In case of the particularly costly phase-module failures, the failure rates differ by a factor of approx. 1.6. With on average 0.21 /MW/year, the overall converter-system failure rate of LV converters is found to be almost three times higher than that of the MV converters.

It is interesting to note the even lower reliability level of LV converters reported in a previous study of the authors, which was based on a much larger WT fleet also including onshore turbines and turbines commissioned before 2007: With an average of 0.48 failures per MW converter capacity and year published in [104] and [47] for liquid-cooled LV converters, the difference to the MV converters systems comes even close to a factor of 6.5.

9.3.2 Distribution of power hardware failures over the generator-side and grid-side converter

In case of power hardware failures (i.e. IGCTs & driver boards, diodes; category 'phase module'), an interesting question is if the damage is predominantly located in either the machine-side (MSC) or the line-side converter (LSC). In the datasets underlying the present

MV converter analysis, this information was available for approximately half of the phase-module failure events. **Figure 46** shows how these cases distribute over MSC and LSC. It makes clear that the machine-side and the line-side part of the MV converters are rather equally frequently affected by failures. In only 2% of the failure events, semiconductors in both MSC and LSC had to be replaced.

A corresponding analysis for LV converters in turbines with DFIG was presented in [36]. It revealed different results for WTs of different OEMs, ranging from prevalence of MSC failures in case of one OEM over equal shares of MSC and LSC failures in a second case to a dominance of LSC failures in a third one.

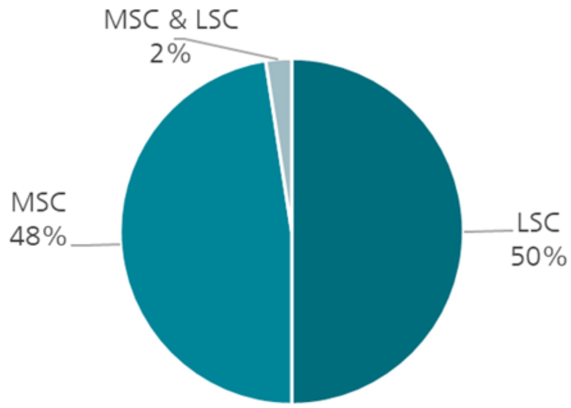

Figure 46. Distribution of power hardware failures over machine side (MSC) and line side (LSC) of the medium-voltage converter system

9.3.3 Seasonal patterns in the failure behaviour

Seasonal patterns in the failure behaviour can provide indications about load or environmental conditions promoting the emergence of failures. In order to identify such patterns, **Figure 47** (a) and (b) present the variation of failure rates (in this case per WT and year) through the course of the year, along with the averaged wind speed and ambient climatic conditions of the evaluated wind farms with MV converters and LV converters, respectively.

Figure 47 (a) shows that the highest phase-module failure rates in the MV converters occurred in the months with the highest average wind speeds. The low failure rates observed during the warm and humid summer months indicate that – in contrast to the phase-module components of LV converters in previous analyses (cf. [36], [163]) and also to the LV converters in the present study – the power hardware of MV converters is much less susceptible to climatic influences. The MV converter control system shows the opposite pattern, with particularly few failures in the strong-wind season and more failures during spring and summer, suggesting climatic influences have a relevant effect on this category. While no obvious seasonal pattern is observed in **Figure 47** (b) for the control system hardware of LV converters, previous analyses of the authors based on a larger LV fleet have revealed a significant effect of mean ambient absolute humidity on converter control system reliability [47].

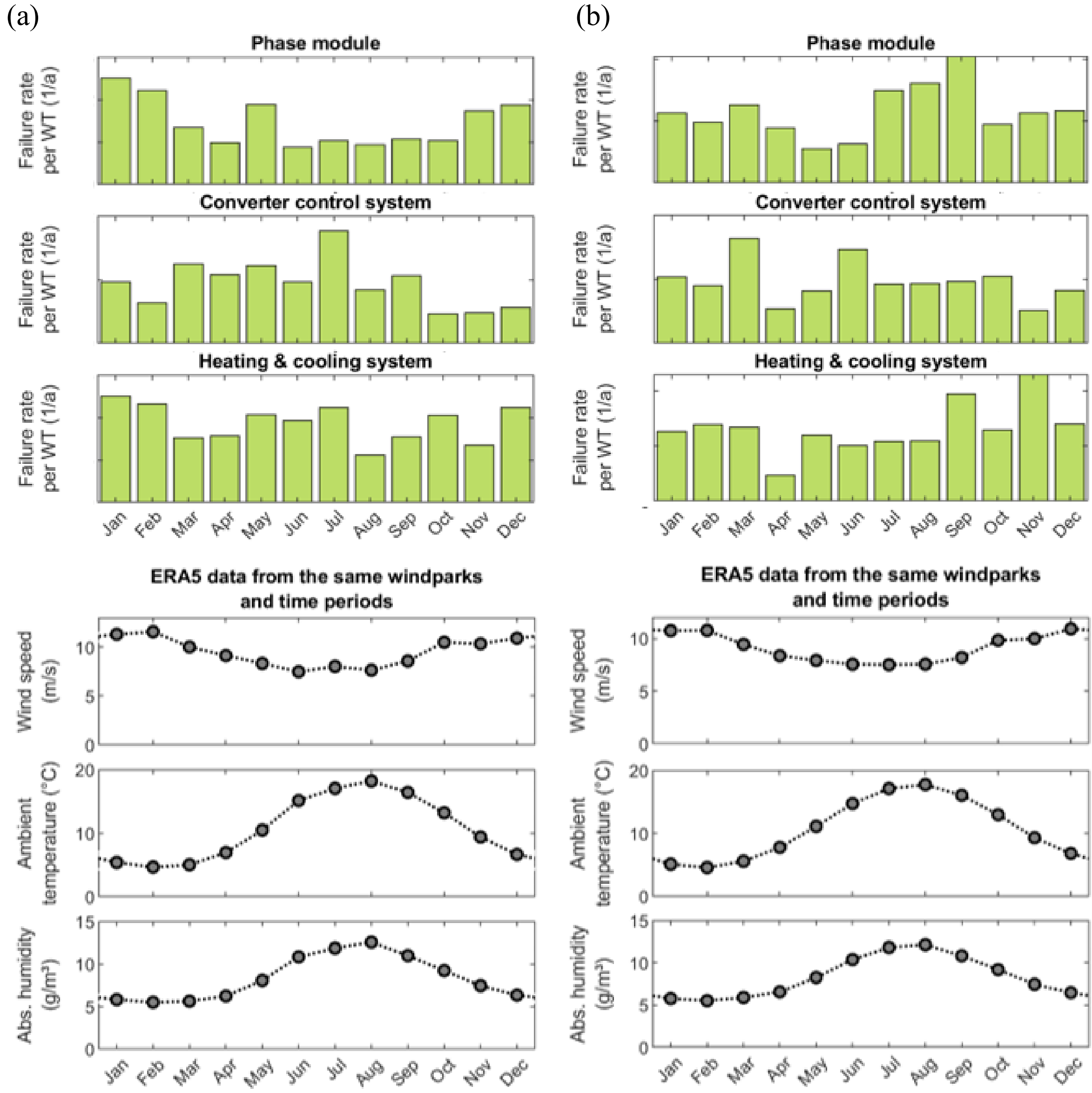

Figure 47. Seasonal variation of failure rates in (a) MV converters and (b) LV converters with corresponding monthly average values of wind speed, ambient temperature and ambient absolute humidity derived from ERA5 data

Patterns in the other converter component categories, of which only the heating & cooling system is included in **Figure 47** (a) and (b), are less clear. For a better judgement of these results, it is important to note that two adverse effects add uncertainty to the failure-rate plots in this section. The first is the limited size of the underlying dataset, leading to an enhanced scatter in the monthly failure rates, which makes a visual identification of seasonal patterns more difficult. The second lies in the nature of the timestamps in the evaluated datasets: For a major part of these, the exact timestamps of the failure incidents were not available. Instead, the datasets provide the timestamps of the resulting maintenance intervention or the corresponding dates on which these were included in the maintenance reports. Where an analysis of the temporal deviation between the former and the latter timestamps was possible, it remained below 12 days in the majority of cases, whereas large deviations of more than a month were encountered in single cases. This introduces a certain but limited uncertainty to the present and the subsequent analysis that are both based on the timestamps of failure events.

9.3.4 Failure behaviour through time

Nelson Aalen plots can reveal valuable information about a system's failure behaviour through time. Each cross-shaped marker in these plots represents a failure event of the respective component category. The slope δ of the resulting graphs indicates if the failure behaviour is dominated by:

- early failures or reliability improvement (with decreasing failure intensity: $\delta < 1$),
- intrinsic failures (constant intensity, $\delta \approx 1$) or
- deteriorating reliability (with increasing failure intensity through time, $\delta > 1$),

i.e. which part of the 'bathtub curve' shown in **Figure 48** characterises the failure behaviour. This is relevant for reliability modelling, but also for the clarification of failure causes.

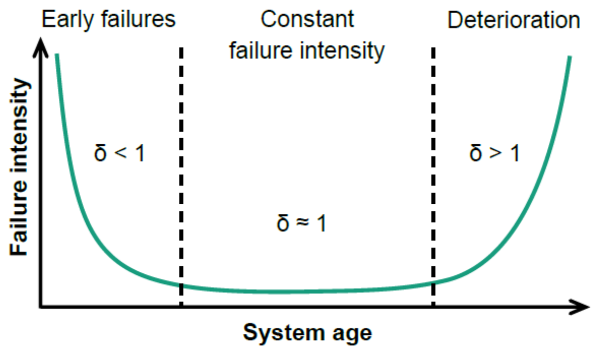

Figure 48. 'Bathtub curve' of repairable technical systems with shape parameter δ as an indicator of early failures, intrinsic failures and deterioration

When analysing Nelson Aalen plots as presented in **Figure 49** (a) for MV and in **Figure 49** (b) for LV converters, it is important to be aware of the logarithmic scale of the axes. On the horizontal axis representing the natural logarithm of the system age (here: operating age of the WT), this leads to a visual stretching of early operating years and a strong compression of later operating years. To further illustrate this effect, the first five operating years take up large parts of the plot, ending at a value of $\ln(5) \approx 1.6$ on the horizontal axis. In contrast to that, the operating years 11 to 15 are represented by the visually much shorter section from $\ln(10) \approx 2.3$ to $\ln(15) \approx 2.7$.

Due to this stretching effect in the left half of the plot and the squeezing effect towards its right side, the assessment of slopes will mainly focus on the right half of the plots, which represents the WT operating age from approx. 4 months to around 15 years.

Comparing the Nelson Aalen plots for the phase-module components of MV and LV converters reveals interesting similarities. Up to an operating age of approx. 3 years, the graphs indicate early-failure behaviour. With a slope of $\delta \approx 0.8$ and the corresponding confidence interval including the 1, this is less pronounced in case of the MV converters than for the LV converters with a slope of only $\delta \approx 0.5$. Another similarity is that the right-most part of the graph indicates deterioration behaviour, with δ ≈ 1.2 (from an age of approx. 6 years onwards) for the MV and δ ≈ 1.6 (from an age of approx. 10 years onwards) for the LV phase modules.

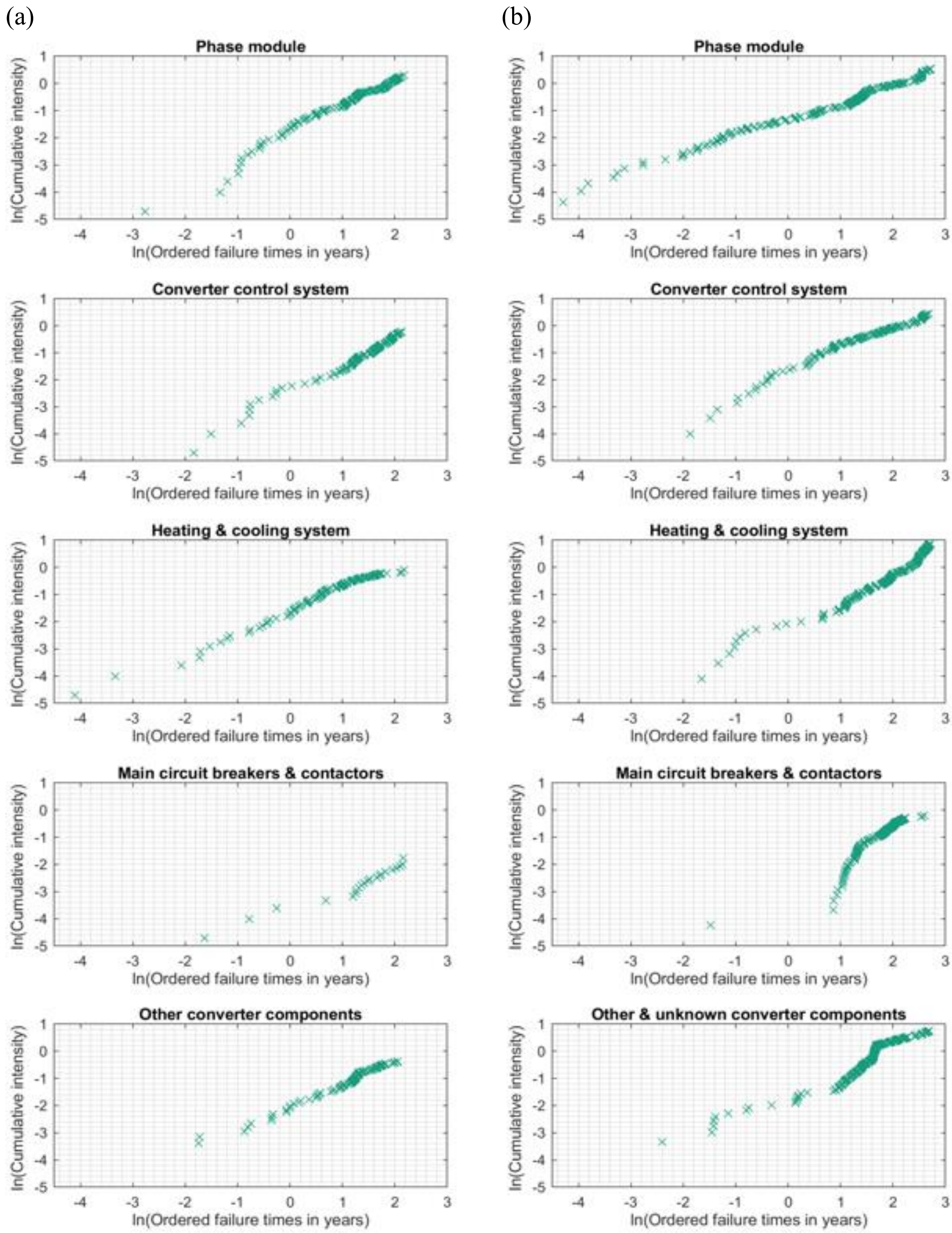

Figure 49. Nelson Aalen plots of (a) medium-voltage converter and (b) low-voltage converter component failures characterising their failure behaviour through time

Again, in this case, the MV phase-module failure behaviour is closer to the desirable behaviour of intrinsic failures characterised by constant failure intensity with δ ≈ 1. On the other hand, the onset of deteriorating reliability of the MV phase modules is observed already at a lower WT operating age. In between these two clearly distinguishable phases, there is in both MV and LV phase modules a period with mixed behaviour, in which a time with increasing failure rates around a WT age of 3 to 5 years is followed by another period with early-failure characteristics.

The converter control systems of both MV and LV converters exhibit a transition from early failures directly to deteriorating failure behaviour, without any notable period of constant failure intensity in between. In case of the MV converters, this is indicated by a change in slope from $\delta \approx 0.7$ to $\delta \approx 1.2$ (with the corresponding confidence intervals touching or even slightly crossing the value of 1). In case of the LV converters, the slope changes from $\delta \approx 0.7$ to $\delta \approx 1.7$ (with confidence intervals not including 1 in this case). A major difference between the control system failure behaviour of MV and LV converters is the turning point in which the transition is observed, which is already around the operating age of 3 years for the MV and as late as around 12 years for the LV converters.

The heating & cooling systems of the MV and the LV converters show considerably different failure behaviour through time: While both are close to failure behaviour with constant intensity in the earlier years of operation (with $\delta \approx 1.1$ and confidence intervals crossing 1 in both cases), an interesting flattening out with improving reliability ($\delta \approx 0.4$) can be observed for the MV converters, starting already during the third year of operation. This is most likely attributable to a learning curve, related to increasing spare-part reliability, improving maintenance practices or a combination of both. In contrast to this, the heating & cooling systems of the evaluated fleet with LV converters shows a relatively stable failure behaviour until an age of approx. 12 years and clear signs of deterioration afterwards ($\delta \approx 2.3$).

In case of the component category “Main circuit breakers & contactors”, we refrain from evaluating the failure behaviour for the MV converters, due to the low number of failure events. In case of the LV converters, this component category shows a very unusual failure pattern, with almost no failures in the first 2.5 years, followed by a phase with strongly increasing failure rates ($\delta \approx 3.4$), most likely related to component degradation. Subsequently, a flattening towards failure behaviour with constant intensity is observed, also in this case possibly related to improved components and / or maintenance practices.

The category “Other (and unknown) converter components” covers a wide variety of mostly smaller components. The corresponding Nelson Aalen plot is therefore a blend of different failure behaviours of a broad range of components. An interesting pattern that the graphs of the corresponding plots for MV and LV converters have in common is the occurrence of a relatively short intermediate phase with a steep increase in failure intensity after approx. 3 and 5 years, respectively, with a sudden transition to subsequent early-failure behaviour. In case of the MV converters, a likely explanation is that among the components grouped in this category, there is at least one suffering from strong aging, so that the resulting failures dominate the pattern until many of these degraded components have been replaced with new ones. In the LV case, however, a deeper investigation has made clear that the short steep section is related to a single event in one of the wind farms and should therefore not be regarded as typical behaviour.

Finally, decreasing failure intensity is observed from year 4 and 6 onwards for the category “Other (and unknown) converter components” of MV and LV converters, respectively. Also in this case, a learning curve with respect to spare-part quality, retrofits or maintenance practices might have contributed to the reliability improvement.

Comparing the levels of cumulative failure intensity between MV and LV converters in **Figure 49** (a) and (b) might leave the impression that there is a reliability advantage of the former only in the categories "Main circuit breakers & contactors" and "Other converter components" whereas levels appear similar for the other three categories. However, it is important to note that such a comparison neglects the different converter capacities in the MV and LV fleet. Taking into consideration that the MV converters have approx. twice the rated capacity of the LV converters evaluated in this study, it becomes clear that also the Nelson Aalen plots confirm the superior reliability of the MV converters.

9.4 Conclusions

This chapter presented a field-data based reliability study for IGCT-based medium-voltage (MV) converters in offshore wind turbines. It is based on failure data from European offshore wind farms covering in total 1249 operating years. For comparison with IGBT-based low-voltage (LV) converters as the by far prevailing technology applied in wind turbines, failure data from European offshore sites from in total 1904 turbine operating years have been additionally evaluated. Besides the converter failure data, site-specific environmental data from the ERA5 reanalysis have been used for all evaluated wind farms in this study.

The analysis of failure rates indicates that the investigated MV converters have considerably lower failure rates per megawatt of converter capacity and year compared to the LV converters: The frequency of phase-module failures, which are usually afflicted with particularly high repair costs, differs by a factor of approximately 1.6, with MV converters showing a failure rate of 0.029 /MW/year compared to 0.048 /MW/year for LV converters. With 0.075 /MW/year for the MV systems and 0.210 /MW/year in the LV case, the overall converter-system failure rate is almost three times higher for the evaluated LV converters than for the MV converters. The difference is even larger when also LV converters in onshore WTs and in wind turbines commissioned before 2007 are included in the failure rate analysis as in [47].

Seasonal patterns in failure behaviour reveal that the highest phase-module failure rates of MV converters occurred during months with the highest average wind speeds. This indicates that, in contrast to LV converters, the power hardware of MV converters is not or much less susceptible to climatic influences, whereas the results suggest a climatic impact on control system failure of the MV converters to be likely, as it is known to be the case also for LV converters (cf. [47]).

An analysis of the converter failure behaviour through time by means of Nelson-Aalen plots has revealed that not only LV, but also most MV converter components show early-failure behaviour with decreasing failure rates in the first years of wind-turbine operation, followed by a transition to deteriorating reliability, which is typically related to component aging or degradation, respectively. Comparing the failure behaviour of the evaluated MV and LV converters in offshore wind turbines, the MV system exhibit a more stable overall failure behaviour with early-failure and deterioration phases being less pronounced, thus coming closer to the desirable pattern of constant failure intensity through time.

In summary, the investigated MV converters show a certain reliability advantage over LV converters in wind turbines, with lower failure rates per converter capacity, reduced susceptibility to climatic influences, and a slightly more mature failure behaviour. As a result, these findings indicate that MV converters are an attractive option for multi-MW wind turbines not only under technical aspects but also from a reliability perspective.

9.5 Acknowledgement

This work is based on data collection, processing and analysis conducted across the two research projects “ReCoWind2” and “RUN25plus”, funded by the German Federal Ministry for Economic Affairs and Energy under grant number 03EI4066A and 03EE3106. In addition to the funding, we kindly acknowledge the provision of field data by our partners in the above and previous projects. ERA5 reanalysis data used in this work were obtained from the Copernicus Climate Change and Atmosphere Monitoring Services [192].

10 Field-data based wind turbine reliability modelling: quantifying effects of operating age, design and technological development

Based on maintenance data from over 1,000 onshore and offshore wind turbines covering more than 4,200 operating years, this study presents an analysis of wind turbine failure behaviour over time and identifies key factors influencing reliability. Failure trends are assessed using Nelson-Aalen plots while non-homogenous Poisson process regression models are developed to quantify the effect of design and technological development, incorporating a range of covariates. Results reveal that while some subsystems exhibit failure intensities following a classical bathtub curve, others transition directly from early failures to deterioration or are monotonically increasing throughout time. The regression modelling results indicate that reliability improves with later commissioning years, highlighting the effectiveness of technological advancements. Rated power negatively affects reliability, with larger turbines experiencing higher failure intensities. Additionally, offshore turbines are generally found to be more reliable than onshore ones, except for the yaw subsystem, which exhibited higher failure rates in offshore environments. Subsystem-specific findings further underscore the influence of design choices: hydraulic pitch systems outperform electrical ones in reliability, and direct-drive turbines demonstrate lower failure intensities in both the drive train and power generation subsystems compared to geared alternatives. The material of this chapter is currently under peer review for publication in [7]

10.1 Introduction

In 2023, global wind energy capacity surpassed 1000 GW due to new onshore and offshore installations. While onshore wind accounts for 92.6% of the total installed capacity, offshore wind is gaining increasing significance [1]. Notably, for offshore wind assets, operations and maintenance (O&M) expenses contribute up to one-third of the Levelized Cost of Energy [50]. Leveraging operational insights to enhance reliability and optimise O&M strategies is therefore essential for cost and risk reduction. However, significant uncertainties remain in the O&M phase due to the limited availability of reliability data for wind turbines and their components. Existing research on wind turbine reliability predominantly relies on outdated and limited datasets, often focusing on annual average failure rates (e.g., [155], [25], [9], [10], [12], [28], [29]).

This lack of field-based, technology-specific input, particularly for newer turbine generations, directly impacts the accuracy of O&M process modelling. Consequently, detailed analyses based on comprehensive field data have significant potential to improve understanding of failure behaviour and maintenance strategies, ultimately supporting more effective decision-making in the wind energy sector [32].

[194] and [13] addressed this gap by presenting failure rates as a function of operating age and analysing the proportion of major repairs relative to asset age. Other studies have examined

[7] Julia Walgern, Fraser Anderson, Athanasios Kolios, Katharina Fischer, 2025. "Field-data based wind turbine reliability modelling: Quantifying effects of operating age, design and technological development". Submitted to Wind Energy for publication [207]

various factors influencing reliability, including the works of [166], [38], [39], [195], [196], [197], and [105], which explore different methodological approaches and environmental or design influences on wind turbine failure patterns.

This study advances the field of wind turbine reliability by providing a comprehensive analysis based on a large and representative set of field data. Utilising maintenance reports covering over 4,200 operational years from both onshore and offshore wind turbines, the research presents deep insights into the reliability of modern turbine technologies. A regression-based reliability modelling approach similar to that developed by Fraunhofer IWES for the power converter subsystem (e.g., [46], [47]) is employed in this study. This research extends beyond previous work to the application of reliability models to critical subsystems other than the converter subsystem. For the analysed subsystems, this allows us to examine differences in failure behaviour between onshore and offshore wind turbines, assess the impact of turbine rated capacity on reliability, and analyse how failure intensity evolves over time. These findings support design optimisation and improving O&M strategies for future wind energy projects.

The following sections detail the methodologies and dataset utilised in this study, followed by a presentation of the results. The findings are analysed in the context of their implications for wind farm operational management and future research directions are evaluated.

10.2 Methodology and evaluated datasets

10.2.1 Methodology

10.2.1.1 Field-data collection and preprocessing

Maintenance reports documenting each turbine visit for over 1,000 wind turbines were collected, ensuring a diverse representation of turbine types across both onshore and offshore installations. The dataset was curated to include both recently commissioned turbines and those with an established operational history, resulting in a comprehensive field-data collection distinguished by its scale, diversity, and recency. To systematically categorise components across different turbine technologies and standardise recorded maintenance interventions, the Reference Designation System for Power Plants (RDS-PP) [40] and the State-Event-Cause-Code "ZEUS" [41] were employed for data preprocessing and classification. Detailed information about the preprocessing approach is provided in [198] and [199]. In this study, a failure is defined as an event requiring corrective maintenance that cannot be resolved through a simple reset and necessitates component replacement.

10.2.1.2 Reliability modelling

The most common method for modelling wind turbine reliability involves calculating the failure rates of components, subsystems, and the overall turbine. To date, studies have primarily reported average failure rates per year and per wind turbine (e.g. [10], [12], [28], [30]). Studies such as [9], [180], [105], and [47] have demonstrated a strong dependence of average failure rates on WT size. Therefore, we proposed in [199] that failure rates should always be expressed per rated capacity in MW and per year, as presented in [13] and [104]. It is important to highlight, however, that failure behaviour is typically not constant through time and that it is not exclusively determined by turbine size, underscoring the need for more advanced reliability models.

This study presents such a detailed reliability analysis that explores

- temporal trends in the failure behaviour of critical subsystems and
- the effect of influential factors, referred to as "covariates", on reliability of critical subsystems.

To address both aspects, we use a methodology based on the well-established reliability theory of repairable systems. Specifically, we apply the Nelson-Aalen estimator to identify trends in failure behaviour over time, and the non-homogeneous Poisson process (NHPP) to quantify the effects of relevant covariates.

The Nelson-Aalen estimator is employed to calculate the non-parametric cumulative intensity for a given component or subsystem, based on the corresponding field data. When plotted on a double-logarithmic scale, these intensity plots can reveal different phases of failure behaviour over a turbine's operating age, which appear as contiguous straight lines with varying gradients (cf. **Figure 50**). The gradient of a straight line in such a double-logarithmic Nelson-Aalen plot corresponds to the shape parameter δ of a power-law process, which we use for modelling the baseline failure intensity λ_0 [42]:

$$\lambda_0(\mathrm{t}) = \left(\frac{\delta}{\upsilon}\right)\left(\frac{t}{\upsilon}\right)^{\delta-1} \tag{10.1}$$

Besides the shape parameter δ>0, the equation contains υ>0 representing the scale parameter of the power-law process.

Consequently, the double-logarithmic Nelson-Aalen plots can thus be used to identify the distinct phases of reliability trends, which form the characteristic shape of the bathtub curve [37], [170]:

- Early failures, which are characterised by a decreasing failure rate (δ<1)
- Constant failures, which are described by a constant failure rate (δ=1)
- Deterioration failures, which are defined by an increasing failure rate (δ>1)

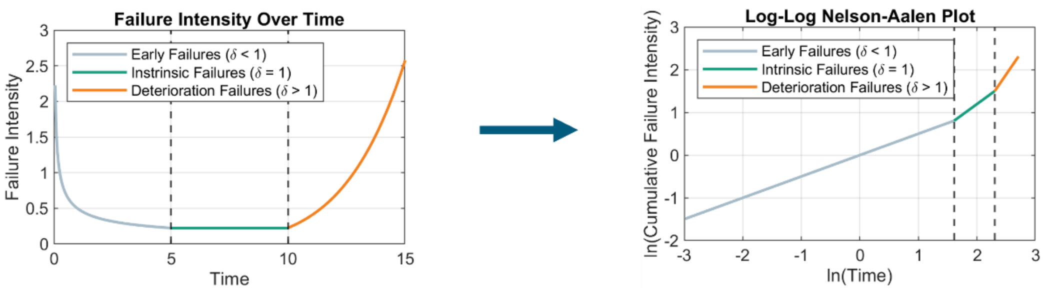

Figure 50. Derivation of the shape parameter δ using Nelson-Aalen plots [44]

We present Nelson-Aalen plots together with the estimated values of δ and the corresponding confidence intervals. To address uncertainty in classifying line segments into early, intrinsic, or deterioration stages based on their gradient, we utilise a bootstrap method. For each bootstrap sample, we conduct a linear regression to create a distribution of possible gradients for each line segment. This distribution is then used to calculate 95% confidence intervals for δ. If a line segment's confidence interval includes δ=1, we cannot confidently assign it to either the early or deterioration failure stage. In such cases, it is categorised as an intrinsic failure stage.

The NHPP is a type of counting process used to model the failure intensity of a repairable system over time. In this analysis, it is formulated such that a set of covariates x_n multiplicatively alters a baseline intensity function $\lambda_0(t)$, resulting in an observed intensity function:

$$\lambda(t) = z\,\lambda_0(t)\exp\,(\beta_1 x_1 + \cdots + \beta_n x_n) \tag{10.2}$$

Next to the baseline failure intensity $\lambda_0(t)$, the covariates x_i and their corresponding coefficients β_i, z accounts for heterogeneity that cannot be explained by the set of observable covariates.

By fitting an NHPP to the dataset of a given subsystem using maximum likelihood estimation (MLE), we can estimate the magnitude and direction of the effects of these covariates through their β coefficients. To enhance the analysis, we integrate an MLE-fitted NHPP with

- principal component analysis, that allows for the simultaneous inclusion of even highly correlated numerical covariates
- a covariate selection procedure based on [200], which enables the identification of covariates that have a significant effect on reliability, and
- a subsampling routine, that addresses uncertainty in the covariate selection procedure by relying on 100 subsamples of the dataset. Each subsample consists of 90% of the original turbine fleet. This approach produces "inclusion rate" plots, which indicate how likely a given covariate is to have a significant effect on the reliability of a particular subsystem.

A detailed presentation of the methodology, using the power converter subsystem as an example, can be found in [47]. Results were obtained using Matlab version R2023b and R version 4.3.2. The R package frailtypack [201] was used to fit the NHPP regression models.

10.2.2 Evaluated datasets

The dataset used in this analysis is derived from maintenance reports of both onshore and offshore wind turbines, encompassing 1089 WTs with a total of over 4,200 operational years. It includes WTs with rated capacities of up to 9 MW and covers data from 2006 to 2024. The analysis period is similar for both onshore and offshore turbines. The dataset encompasses operational data from turbine commissioning up to a maximum operating age of approximately 18 years. It includes both left- and right-censored data.

The dataset contains detailed information about maintenance of different components, classified into the following subsystems according to [40]: environmental measuring system (CKJ), rotor system (MDA_rotor), pitch system (MDA_pitch), drive train system (MDK), yaw system (MDL), central hydraulic system (MDX), control system (MDY), power generation system (MKA), generator switching system (MSC), converter system (MSE), generator transformer system (MST), nacelle (MUD), remote monitoring system (MYA), tower system (UMD), personnel rescue systems (WBA), fire extinguishing system (XGM), fire alarm system (CKA), lifting gears (XMM), obstacle warning system (XSD), low voltage electrical main supply system (BFA), transformer station (UAB, in case of onshore turbines), equipotential bonding / earthing system (XFB), lightning protection system (XFC), ventilations systems (XAM), central lubrication system (MDV), compensation system (MSS), common cooling system (MUR), telephone system (Y), and other (G).

The dataset analysed in this study includes various technical concepts, covering hydraulic and electrical pitch systems, as well as geared, direct-drive, and hybrid drive trains. It further encompasses different generator types (DFIG, EESG, PMSG, SCIG) across low and medium voltage levels, along with air- and liquid-cooled converter technologies. A comprehensive description of the dataset can be found in [199].

10.3 Results and discussion

In the reliability analysis presented in [199], which is based on the same dataset as this study, the pitch system, control system, converter system, and drive train system were identified as the most critical in terms of failure rates. In addition to these, we choose the rotor system, power generation system, and yaw system for deeper analysis in the present paper as these subsystems are key for the power conversion process.

The objective of the following analysis is to characterise failure patterns over time by means of Nelson-Aalen plots (Section 3.1) and identify factors that significantly affect subsystem and overall WT reliability using NHPP regression models in combination with a covariate selection procedure (Section 3.2).

10.3.1 Failure behaviour through time

The failure behaviour of a technical system over time is commonly expected to follow a bathtub curve, comprising three distinct phases: early failures with decreasing failure intensity in the initial years of operation, a constant failure intensity related to intrinsic failures, and a final phase of increasing failure intensity due to deterioration. However, prior research on the power converter subsystem and its components has demonstrated that not all subsystems necessarily exhibit all three phases of the bathtub curve (cf. [47], [43]). As explained in [202], early failures are typically related to e.g. material or manufacturing defects, insufficient testing or inadequate mounting. Intrinsic failures are e.g. caused by human errors during maintenance or other external causes like lightning strikes or excessive voltage peaks in the power grid. In the deterioration phase, failures are dominated by aging or wear-out, i.e. by degradation accumulated and progressing as the system is used. A mature and desirable reliability behaviour would consist of a long phase with a low and constant failure intensity, followed by a late transition to deterioration as the system reaches the end of its intended service life. In the context of system- or subsystem-level analysis as in the present work, it is important to note that the mix of different components with their variety of failure modes and failure mechanisms can potentially bias the identified reliability trends towards intrinsic failure behaviour.

Figure 51 presents the Nelson-Aalen plots derived from our dataset, illustrating cumulative failure intensities over time for (a) the entire WT system and (b)-(h) individual subsystems. As explained above, these plots use double-logarithmic scaling. For better readability, we provide two horizontal axes: the natural logarithm of the WT operating age is displayed on the upper horizontal axis and the actual operating age on the lower horizontal axis for reference. Among the analysed subsystems, only the converter and drive train subsystem exhibit the classical bathtub curve pattern. In contrast, the overall WT system and the control subsystem transition directly from early failures to deterioration failures, without a notable phase of constant failure intensity dominated by intrinsic failures. The yaw system displays a transition from a constant to an increasing failure intensity. The failure behaviour of the pitch and rotor subsystems is predominantly driven by deterioration. The trend of improving reliability observed in later operation years for the rotor subsystem is likely related to a learning curve, possibly due to improved maintenance practices. Also, the power generation subsystem, which includes the generator, demonstrates an atypical pattern, transitioning from a constant failure intensity to deterioration and then reverting to a stable failure intensity. This could be associated with a learning curve as in case of the rotor subsystem. A potential alternative explanation is that during the phase of deterioration, a large number of components of that subsystem was replaced with new ones, biasing the reliability trend of the last phase towards that of the early years of operation.

Comparisons with previous WT reliability studies that addressed trends in failure behaviour such as [194] and [13] remain challenging due to differences in component classification, failure definition or methodological approaches. Still some similarities can be observed. For instance, with increasing repair rates for blades, [13] reports a similar trend as we found for rotor subsystem failure intensity in this study. However, the comparability of gearbox and generator repairs with the present study is limited, as SPARTA focuses on individual major components rather than entire subsystems and considers repairs rather than failures according to our definition above. Additionally, SPARTA's "electrical" category encompasses a broader range of components than the subsystems defined in this study.

[194] analysed failure rates per turbine per year of operation for different subassemblies, facilitating better comparability with the subsystems defined in this study. However, their dataset covers only the first eight years of WT operation, allowing a comparison only with the first part of the long-term failure behaviour shown in **Figure 51**. Nonetheless, for the first years of operation, similar trends are observed for comparable subsystems, including the pitch, control, converter, yaw, and rotor subsystems. Notably, the rotor subsystem in this study aligns with Carroll's "blades" and "hub" subassemblies. However, as with the SPARTA study, direct comparisons for the generator and gearbox subassemblies remain challenging due to different component classification.

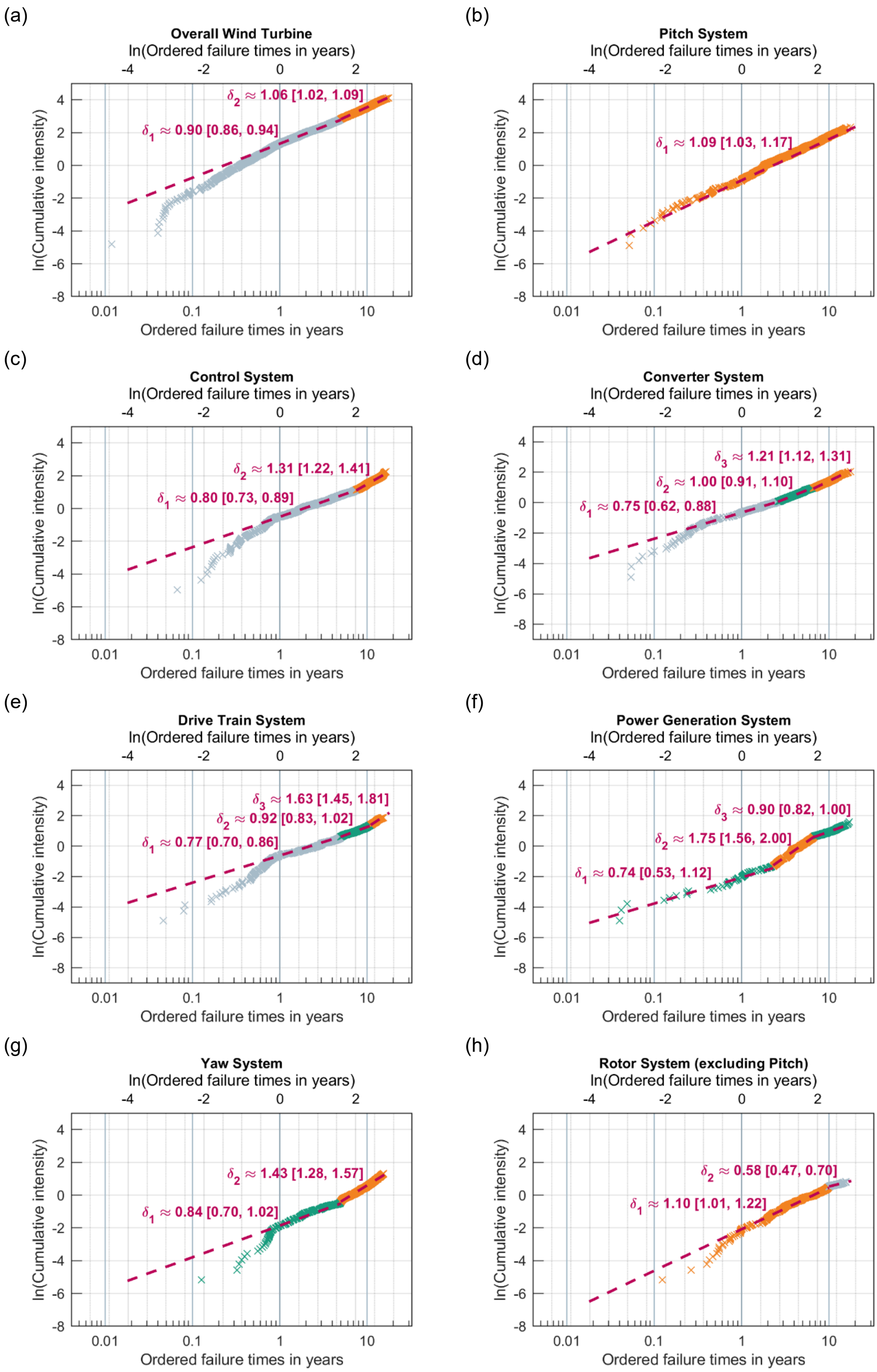

Figure 51. Cumulative failure intensity plots for the entire wind turbine system (a) and individual subsystems (b)-(h). Crosses represent observed failures from the field dataset, categorised as "early" (grey), "intrinsic" (green), and "deterioration" (orange). Red dashed lines indicate the best-fit models, with corresponding δ parameters displayed. Values in brackets represent 95% confidence intervals for the estimated δ parameters.

10.3.2 Factors influencing reliability

In order to identify the factors that have a significant effect on subsystem and overall WT reliability, NHPP regression models are utilised. For most models three covariates are considered: two numerical covariates – WT rated capacity and WT commissioning year – and one categorical covariate, distinguishing between onshore and offshore locations. As an initial step, the correlation between these covariates is assessed. While highly correlated numerical covariates can be addressed through principal component analysis (PCA), categorical covariates, which are highly correlated with numerical ones, could cause instability in covariate estimates. As an initial guideline, we use the "rule of thumb" suggested by [203], that a pair of covariates should not be included in the same regression model if the magnitude of their pairwise correlation (|r|) exceeds ~0.7.

The correlation results for the overall WT dataset, summarised in **Table 19**, indicate that the correlation coefficients of the categorical variable "onshore/offshore" remain below this threshold, allowing its inclusion in the model in the first instance. However, the influence of this correlation will be critically evaluated based on the stability of covariate estimates by means of confidence intervals and the inter-subsample variance of beta value estimates. The strong correlation between WT commissioning year and rated capacity will be managed using PCA, following the methodology described in [47].

Table 19. Linear correlation coefficients for the covariates considered in this study

	Onshore/Offshore	**WT Commissioning Year**	**WT Rated Capacity**
Onshore/Offshore	1	-0.495	-0.547
WT Commissioning Year	-0.495	1	0.788
WT Rated Capacity	-0.547	0.788	1

Additionally, we compare the effects of incorporating rated power directly versus in a logarithmised form to determine, which approach yields a better model quality, indicated by a maximised log-likelihood. Note that a direct inclusion of rated power implies an exponential effect on failure intensity, whereas a logarithmised inclusion corresponds to a root-function (for $0<\beta<1$), linear (for $\beta=1$) or a power-function effect (for $\beta>1$), as can be derived from Equation (10.2).

For specific subsystems, additional categorical covariates related to their design characteristics are incorporated into the analysis:

- For the pitch subsystem model, a covariate distinguishing between hydraulic and electrical pitch systems is included.
- For the converter subsystem model, a covariate differentiating between fully rated and partially rated converters is evaluated, as the converters in turbines with doubly fed induction generator (DFIG) have a rated capacity of only approximately one third of the WT capacity and as, in contrast to former studies conducted by Fraunhofer IWES ([104] and [47]), the rated capacity of the WT instead of the converter is utilised for the covariate "rated capacity" in the present work.

- For the drive train subsystem model, the impact of different drive train configurations is examined by differentiating between geared and direct drive turbines. Hybrid drive turbines are included in the "geared" category.
- For the power generation subsystem model analyses, a covariate distinguishing between low-speed generators and a combined category of medium- and high-speed generators is included. Medium- and high-speed generators are grouped together due to their common application in hybrid drive and fully geared turbine configurations, whereas low-speed generators are used in direct drive turbines. Consequently, the covariate "drive train configuration" serves as a proxy for the underlying generator-speed category.

For each of these additional covariates, correlation coefficients were assessed to ensure compliance with the threshold established by [203], confirming their suitability for inclusion in the analysis. Note that this limit remains a rough guide. It is also necessary to evaluate the stability of the fitted models, which is facilitated in this analysis by the subsampling procedure and uncertainty quantification in the fitted models.

Figure 52 presents inclusion rate plots illustrating the outcomes of the covariate selection procedure for both the overall wind turbine system (a) and the seven individual subsystem (b)-(h) analysed in the preceding section. Note that the covariates "Onshore/Offshore", "Commissioning Year", and "Rated Power" were included in the selection process for all subsystems. The covariate "Electric/Hydraulic" was considered exclusively in the analysis of the pitch subsystem, while the covariate "Drive Train Concept" was assessed solely for the drive train and power generation subsystems. Similarly, the covariate "Fully/Partially Rated" was included only in the analysis of the converter subsystem. Consequently, only covariates that were part of the selection procedure are displayed on the x-axes of **Figure 52**.

Covariates consistently eliminated during the selection process exhibit an inclusion rate of 0%. This applies to "Commissioning Year" for the pitch, control, and power generation subsystems, as well as "Rated Power" for the yaw subsystem. Additionally, for the control subsystem, "Rated Power" was identified as significant in only a small fraction of subsamples and was therefore excluded from the final reliability model. Similarly, the covariate "Onshore/Offshore" was deemed relevant in only approximately 20% of subsamples for the converter subsystem and was subsequently omitted from the final reliability model of that subsystem.

All remaining covariates demonstrated a significant effect on reliability and were retained in the final NHPP regression models. In this study, which focuses on reliability modelling, covariates with inclusion rates exceeding 50% are considered to have a significant effect. In contrast, previous studies on the converter subsystem by Fraunhofer IWES applied an 80% threshold, as those investigations were centred on root-cause analysis (see e.g. [47]). Based on our experience, the exclusion of a covariate during the selection process can also be attributed to limitations in dataset size rather than an actual absence of effect. Therefore, exclusion does not necessarily imply a lack of influence. As a result, only covariates identified as having a significant effect should be interpreted. **Table 20** summarises the results of the final reliability models, presenting the estimated β coefficients and their respective confidence intervals for each covariate across the overall wind turbine system and individual subsystems.

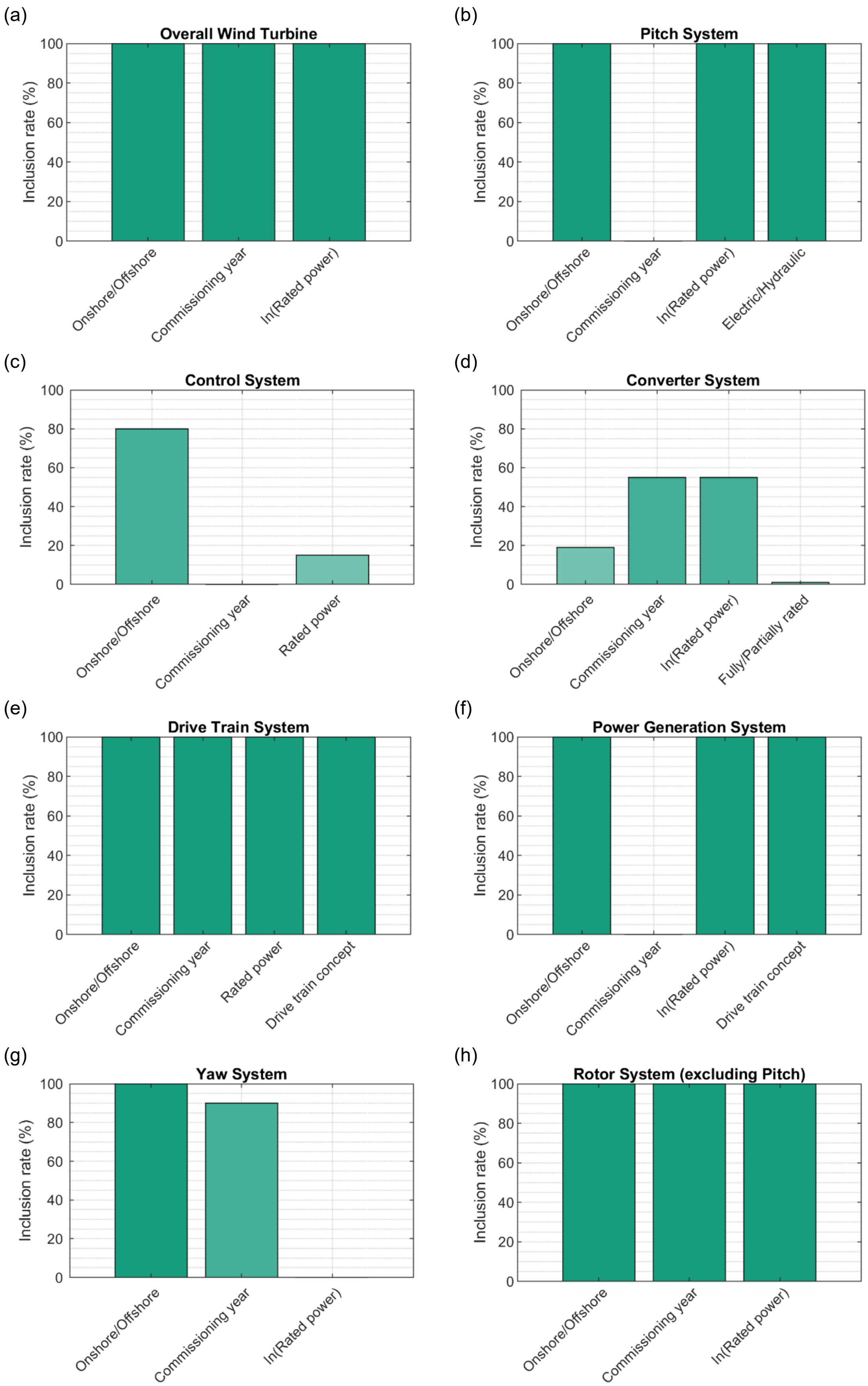

Figure 52. Inclusion rate plots for the entire wind turbine system (a) and individual subsystems (b)-(h)

For the overall wind turbine system and most subsystems, a logarithmised inclusion of the covariate “Rated Power” was found to enhance the quality of the reliability model. However, for the drive train subsystem, a direct inclusion yielded a superior fit.

Additionally, the estimated β coefficients associated with the selected covariates (see **Figure 52** and **Table 20**) provide insights into both the direction and magnitude of their effect on overall wind turbine and subsystem reliability, enabling a quantitative assessment of their influence.

For both the overall wind turbine system and the subsystems where the covariate “Onshore/Offshore” was found to have a significant effect on reliability, offshore turbines exhibited higher reliability compared to their onshore counterparts. This finding aligns with the results of [199], who reported lower annual average failure rates for offshore turbines per MW of turbine capacity. However, an exception is observed for the yaw subsystem, where an opposite effect is indicated by the identified β coefficient. While [199] normalised failure rates by rated power assuming a general scaling effect, the present analysis provides deeper insights: for the yaw system, the results indicate that rated power does not significantly influence its reliability. Consequently, in the specific case of this subsystem, a comparison based on annual average failure rates per turbine is more appropriate, demonstrating that yaw subsystems in onshore turbines are more reliable than those in offshore applications. Although it is reasonable to assume that greater efforts are made to minimise failures in offshore WTs, there is no clear explanation for the inferior reliability of yaw subsystems in offshore environments.

Analysing the influence of turbine commissioning year on reliability reveals a distinct trend across most subsystems. For the converter, drive train, and yaw subsystems, as well as for the overall wind turbine system, reliability has improved with later commissioning years. This finding highlights the effectiveness of technological advancements and reliability-improving measures implemented over time. However, an inverse trend is observed for the rotor subsystem, where reliability declines in more recent turbine generations. This may be attributed to the industry’s shift toward slimmer and more flexible blade designs, which are optimised for larger turbines but may introduce new reliability challenges affecting operational expenditure.

While reliability improvements are observed for later commissioning years, the effect of rated power on reliability is negative: larger turbines exhibit lower reliability for the overall wind turbine system, as well as for the pitch, converter, drive train, power generation, and rotor subsystems. Although this may initially appear to contradict the positive trend associated with commissioning year, the NHPP methodology enables the quantification and distinction of these independent effects on reliability. The observed decrease in reliability with increasing rated power aligns with previous findings for the power converter (e.g. [47]), the pitch subsystem (cf. [35], [105]), the drive train subsystem [180], and the overall wind turbine system (e.g. [9], [13]).

While [105] concluded, based on annual average failure rates per WT, that electrical and hydraulic pitch systems exhibit similar reliability, the present study – utilising a larger and more representative dataset along with a more advanced analysis methodology – demonstrates that turbines equipped with hydraulic pitch systems exhibit higher reliability than those with electrical pitch systems.

For the drive train subsystem, the drive train concept has been identified as a significant factor influencing reliability. It is important to note that this subsystem encompasses multiple subassemblies, including the rotor bearing, speed conversion, drive train brake, high-speed shaft, drive train auxiliary systems, main and offline gear oil systems, oil lubrication system, rotor lock, rotor slewing unit, and drive train cooling system. As a result, the analysis includes both wind turbines with and without gearboxes. Given that direct drive turbines inherently have fewer components within this subsystem category defined by RDS-PP [40], it is unsurprising that they exhibit lower failure intensity compared to geared turbines. Nevertheless, this finding aligns with [13], which reported that direct drive turbines experience fewer average monthly forced outages per MW and lower associated production losses.

Table 20. Results of the final NHPP regression models showing the β coefficients and their respective confidence intervals for various covariates across the overall wind turbine system and individual subsystems

	Reference Level	Factor Level	β	exp(β)	95% Confidence Interval
Wind Turbine Overall					
Onshore/Offshore	Offshore	Onshore	0.153	1.166	(0.065, 0.242)
Commissioning year	-	-	-0.014	0.986	(-0.024, -0.005)
ln(Rated power)	-	-	0.592	-	(0.492, 0.692)
Pitch System					
Onshore/ Offshore	Offshore	Onshore	0.444	1.559	(0.262, 0.626)
ln(Rated power)	-	-	0.876	-	(0.716, 1.037)
Electric/Hydraulic	Electric	Hydraulic	-0.437	0.646	(-0.573, -0.302)
Control System					
Onshore/ Offshore	Offshore	Onshore	0.196	1.216	(0.184, 0.207)
Converter System					
Commissioning year	-	-	-0.046	0.955	(-0.069, -0.024)
ln(Rated power)	-	-	0.426	-	(0.224, 0.628)
Drive Train System					
Onshore/ Offshore	Offshore	Onshore	0.934	2.546	(0.737, 1.132)
Commissioning year	-	-	-0.052	0.949	(-0.075, -0.029)
Rated power	-	-	0.303	1.353	(0.240, 0.365)
Drive train concept	Geared	Direct Drive	-1.137	0.321	(-1.368, -0.906)
Power Generation System					
Onshore/Offshore	Offshore	Onshore	0.651	1.917	(0.406, 0.896)
ln(Rated power)	-	-	1.421	-	(1.202, 1.642)
Drive train concept	Geared	Direct Drive	-0.473	0.623	(-0.723, -0.224)
Yaw System					
Onshore/ Offshore	Offshore	Onshore	-0.732	0.481	(-0.936, -0.528)
Commissioning year	-	-	-0.042	0.959	(-0.066, -0.018)
Rotor System (excl. Pitch)					
Onshore/ Offshore	Offshore	Onshore	1.078	2.938	(0.769, 1.387)
Commissioning year	-	-	0.092	1.097	(0.060, 0.125)
ln(Rated power)	-	-	0.953	-	(0.623, 0.283)

For the power generation system, direct-drive turbines have been found to exhibit higher reliability. While this may initially seem counterintuitive given that direct-drive turbines require large low-speed generators, the finding is plausible as these turbines are typically equipped with permanent magnet synchronous generators (PMSG) or, onshore, with electrically excited synchronous generators (EESG). Compared to doubly-fed induction generator (DFIG) configurations, the synchronous generator systems have fewer potential failure modes, such as the absence of a slipring unit, which is subject to wear-out and as such a typical driver for maintenance interventions. Our result aligns with the findings of [166], who reported that PMSG-based systems, including their auxiliary components such as cooling and lubrication systems, failed less frequently than DFIG-based configurations in turbines of identical capacity.

10.3.3 Discussion and comparison of different reliability modelling approaches and their impact on O&M simulations

Reliability modelling plays a fundamental role in optimising O&M strategies for WTs. While O&M simulations commonly rely on average failure rates per turbine (e.g. [204], [157], [205]), we suggest in [199] using failure rates per MW to account for the observed increase in average failure rates with higher WT rated capacities.

This study advances reliability modelling by employing NHPP regression models, which provide a more refined assessment of failure behaviour. Two key aspects should be considered:

First, the Nelson-Aalen plots in **Figure 51** demonstrate that failure behaviour varies over time, making the assumption of constant annual failure rates an oversimplification that can lead to inaccuracies in maintenance planning. Second, **Table 20** highlights the influence of multiple covariates on reliability, emphasising the importance of differentiating between turbine design concepts to enhance the accuracy of reliability assessment. One of the most severe implications for O&M simulations is the effect of turbine capacity. Published failure rates are often derived from datasets dominated by smaller turbines, while O&M simulations are typically conducted for currently installed or future wind farms with larger rated capacities of the WTs. This study shows that using average failure rates per turbine in such cases leads to an underestimation of maintenance requirements, as failure intensity generally scales with turbine size (cf. **Table 20**).

Figure 53 illustrates the isolated effect of WT rated capacity on subsystem reliability for those subsystems where rated capacity has a significant effect. Failure intensity scales exponentially with WT rated capacity for the drive train subsystem. In contrast, the effect is best described by a root function in case of the pitch and the converter subsystem, is close to linear for the rotor subsystem, and is represented by a power function for the power generation subsystem. The black dashed line in **Figure 53** represents the outcome of generic scaling per MW. Although previous results suggest that this approach is more accurate than assuming constant failure rates per turbine, the results of this study indicate that different subsystems require distinct scaling factors. For example, doubling the rated capacity increases the failure intensity of the overall wind turbine by a factor of $2^{0.592} \approx 1.51$ (cf. β values in **Table 20**), with subsystem-specific variations: the failure intensity of the converter subsystem scales by a factor of $2^{0.426} \approx 1.34$ when the WT rated capacity doubles. That of the rotor subsystem is multiplied by $2^{0.953} \approx 1.94$ – closely aligning with the generic MW-based scaling – and that of the power generation subsystem by $2^{1.421} \approx 2.68$, reflecting the highest sensitivity to rated capacity. These examples underline the substantial differences in reliability across turbine sizes.

It is important to note that **Figure 53** solely depicts the effect of rated capacity, while other factors, such as reliability improvements in later commissioning years (cf. Section 3.2), must also be considered. Therefore, incorporating advanced reliability models into O&M assessments is recommended over using simple average failure rates. However, if a more straightforward approach is required, scaling failure rates per MW remains preferrable to assuming constant failure rates across different turbine sizes, even though rated capacity does not significantly affect all subsystems.

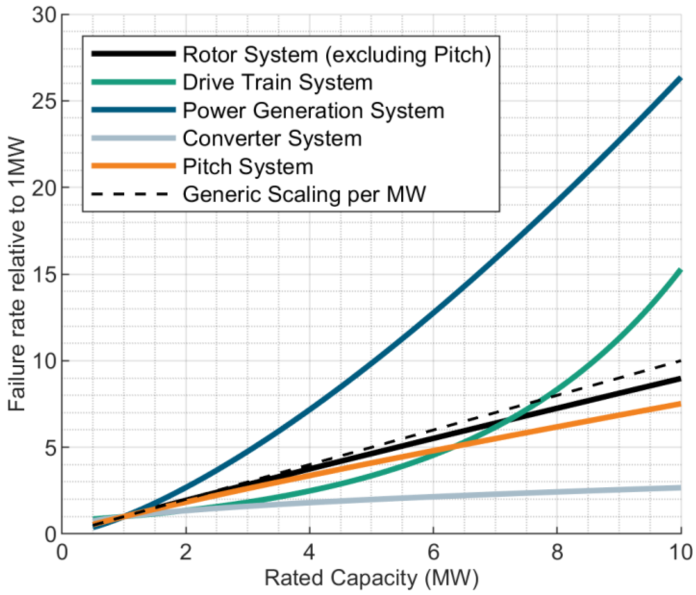

Figure 53. Impact of rated capacity on wind turbine subsystem reliability

The former examples highlight that the choice of reliability modelling methodology and depth significantly affects the accuracy of reliability assessments and, consequently, the value of O&M simulations and associated decision-making. Selecting an appropriate model is essential to capturing the complexity of WT failure behaviour and ensuring that O&M strategies are both cost-effective and operationally efficient.

10.4 Conclusions and outlook

This study has provided a comprehensive analysis of wind turbine (WT) reliability on subsystem level, based on failure data from over 1,000 WTs and more than 4,200 operational years. The dataset includes turbines with rated capacities of up to 9 MW and operating ages up to 18 years. Failure behaviour over time has been examined using Nelson-Aalen plots and the influence of covariates on reliability has been analysed through a non-homogeneous Poisson process (NHPP) in combination with a covariate selection procedure.

The results highlight distinct reliability trends over WT operating age across different subsystems, demonstrating that while some subsystems exhibit a classical bathtub curve, others transition directly from early failures to deterioration. These findings emphasise the necessity for time-dependent subsystem-specific reliability modelling rather than assuming a uniform failure behaviour over time across all components.

The results of NHPP regression and the related covariate selection procedure confirm that several covariates significantly influence WT and subsystem reliability. A later turbine commissioning year positively impacts reliability for most subsystems, indicating the effectiveness of technological advancements and design improvements over time. In contrast, a higher turbine rated power has a negative effect on reliability, confirming previous findings that larger turbines tend to experience higher failure intensities. These opposing trends underline the importance and advantages of NHPP modelling, which allows for the separation and quantification of individual covariate effects.

Reliability differences have been also observed between onshore and offshore turbines, with the subsystems of offshore turbines generally achieving higher reliability than their onshore counterparts, except for the yaw subsystem, where an opposite effect has been identified.

Additionally, subsystem design choices were found to play a crucial role in reliability outcomes. Hydraulic pitch systems demonstrated higher reliability compared to electrical pitch systems. This result highlights the importance of multivariate analysis, as a previous study based solely on average failure rates per turbine and a smaller, less representative dataset had indicated similar levels of reliability of hydraulic and electrical pitch systems (cf. [105]). Another finding of the present study with respect to design choices is that direct-drive turbines exhibited superior reliability in both the drive train and power generation subsystem. This can be attributed to the reduced number of components and failure modes associated with direct-drive configurations.

Furthermore, this study examined and compared different reliability modelling approaches, emphasising their impact on O&M simulations. While traditional assessments based on average failure rates per turbine or per MW of WT capacity provide a simple means of describing reliability, they do not account for time-dependent failure behaviour or key influencing factors. In contrast, NHPP regression modelling offers a more advanced approach by incorporating age-dependent failure intensities as well as covariate effects, leading to a more comprehensive understanding and representation of WT reliability. The choice of reliability modelling methodology plays a critical role in O&M simulations, as relying on simplified average failure rates may result in inaccurate cost estimations and suboptimal O&M strategies. In contrast, NHPP-based models enhance predictive accuracy, supporting more effective O&M planning and financial decision-making.

Overall, this study underscores the importance of continuously collecting and analysing large-scale field data to enhance WT reliability. The results provide valuable insights for manufacturers, operators and maintenance planners, enabling data-driven decisions for design optimisation, O&M strategies, and lifetime extension efforts.

In the present study, the described methodology has been applied with a limited set of covariates, focusing on the most critical subsystems. Further refinement of these subsystems, e.g. of the drive train system, into individual subassemblies or components enables more detailed reliability modelling and remains subject of ongoing work. Likewise, future regression models will incorporate additional covariates, particularly those related to operating conditions.

Further methodological advancements have been explored in previous studies, including [206] and [43], both of which focus on the converter system. [206] improved upon the assumption of constant covariates by introducing time-dependent covariates, revealing a previously unobserved dependence on electrical utilisation. [43] introduced an approach for fitting separate NHPP models to distinct phases of the bathtub curve. This enhances accuracy by allowing covariate effects to be assessed separately at different stages of a turbine’s operational life, taking into account that different failure mechanisms with different drivers and promoting factors dominate in the phases distinguished by their reliability trends.

Expanding this analysis with aforementioned methodologies holds further significant potential for optimising O&M strategies and improving root cause analysis, ultimately contributing to enhanced reliability and cost-effectiveness in wind energy operations.

10.5 Acknowledgements

The present work was mostly carried out within the research project "Reduction of uncertainties for continued operation of offshore wind farms combining reliability and yield analysis (RUN25+)" funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant number 03EE3106. The provision of comprehensive field data by project partners is gratefully acknowledged. Further financial support was received by EPSRC through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1.

11 Discussion: Impact of using different reliability models for O&M simulations

Reliability modelling is a crucial aspect of optimising O&M strategies for wind turbines. Different approaches exist and were developed within this thesis for quantifying reliability, ranging from simple average failure rates per turbine or per MW to more advanced statistical methods such as non-homogeneous Poisson process (NHPP) regression modelling. The choice of methodology significantly influences the accuracy of reliability assessments and, consequently, the effectiveness of O&M simulations and related decision-making.

11.1 Limitations of average failure rates

Traditional reliability assessments often rely on average failure rates per WT, more recently also on average failure rates per MW of turbine capacity. While these metrics provide a straightforward overview of failure behaviour, they come with inherent limitations. First, they assume a constant failure rate over time, disregarding the dynamic nature of WT reliability, which is characterised by early failures, constant failures, and deterioration phases. Second, these averages fail to capture dependencies on factors such as design characteristics, environmental conditions, or operational strategies. Consequently, using average failure rates in O&M simulations can lead to oversimplifications, potentially underestimating or overestimating failure risks, and resulting e.g. in suboptimal maintenance scheduling and spare parts planning.

11.2 Advantages of multivariate reliability modelling

More sophisticated approaches, such as Nelson-Aalen estimator and NHPP regression models, offer a deeper understanding of failure behaviour. Nelson-Aalen plots have been used within this thesis for a visual assessment of cumulative failure intensities over time, revealing distinct failure patterns across different subsystems (cf. **Figure 51**). This helps to differentiate between components exhibiting the classical bathtub curve and those transitioning directly from early failures to deterioration or following other failure patterns.

NHPP regression modelling further enhances reliability analysis by incorporating time-dependent failure behaviour and offering the possibility to include the effect of covariates such as rated power, commissioning year, drivetrain configuration, and site conditions. The ability to separate and quantify the effects of these factors allows for a more precise estimation of failure risks under varying conditions. For instance, NHPP modelling has demonstrated that while technological improvements over time have enhanced reliability, larger turbines tend to experience higher failure intensities (cf. Section 10.3.2). Such insights cannot be captured through simple averaging methods.

11.3 Impact on O&M simulations

This thesis developed two distinct approaches for reliability modelling based on failure data from more than 1,000 wind turbines, covering more than 4,200 operating years, with rated capacities of up to 9 MW and operating ages of up to 18 years. The first approach estimates average failure rates per MW of rated capacity per year (cf. Chapter 7), while the second employs multivariate reliability modelling to isolate and quantify the effect of factors such as design, technological advancements, rated capacity, and operating age (cf. Chapter 10). Both methodologies are compared to the reliability figures of Carroll et al. [12], which despite being based on 3 MW turbines and published already a decade ago, remain widely used for O&M simulations due to their comprehensive breakdown of subassembly-specific failure and repair rates, repair times, and spare part costs. The comparative analysis of the three reliability modelling approaches and their impact on O&M simulations is summarised in **Table 21**.

When reliability figures serve as input parameters for O&M simulations, the choice of modelling approach directly impacts the simulation results and thus accuracy of maintenance planning, cost estimation, and energy availability predictions.

The assessment of O&M simulation impacts in this chapter, along with the findings from Chapter 10, emphasise the importance of considering turbine size and time-dependent failure behaviour. Previous studies (e.g. [204], [157]) indicate that the reliability figures provided by Carroll et al. [12] are only applicable to smaller turbines (~3MW) and highlight the lack of field data of larger turbines. This thesis addressed this gap by developing reliability models based on field data of turbines up to 9 MW, currently being the largest turbines for which a certain operation history is available. Newer turbines with rated capacities of 11 MW as utilised e.g. for the Hollandse Kust wind farms or up to 14 MW as for Morray West have been operational since end of 2023 or will enter full operational phase in 2025, respectively.

While the average failure rate per MW approach provides a useful baseline, multivariate reliability modelling serves as a complementary extension offering deeper insights into failure behaviour over time and the influence of different covariates. In particular, NHPP regression models reveal that failure rates of many subsystems do indeed scale with turbine size, but not in a strictly proportional manner as suggested by the per-MW approach.

Therefore, using average failure rates for O&M simulations, particularly if not adjusted for time-dependent behaviour, turbine size or influencing factors, may lead to:

- Overestimation or underestimation of failure events, resulting in inefficient O&M strategy development
- Inaccurate spare parts management, potentially leading to costly downtimes if critical components are unavailable when needed
- Suboptimal long-term financial planning, as O&M cost projections may not reflect actual failure trends and influencing factors

In contrast, incorporating results from NHPP regression models allows O&M simulations to account for evolving failure risks over time, providing a more realistic representation of expected maintenance needs. The ability to integrate specific turbine characteristics and environmental conditions leads to improved field-data based maintenance strategies, reducing unexpected failures and optimising resource allocation. Overall, the transition from simple average failure rates to more sophisticated reliability models represents a crucial step toward improving the efficiency and cost-effectiveness of WT O&M strategies. However, if a more simplified approach is desired or necessary, scaling failure rates per MW is preferable to assuming constant failure rates across different turbine sizes, despite rated capacity not significantly affecting all subsystems. Furthermore, it is crucial to base O&M simulation inputs on datasets that include turbines of similar sizes to those being modelled, as extrapolating to substantially larger turbines introduces uncertainties and potential inaccuracies.

Table 21. Comparison of the two developed reliability modelling approaches ([199], [207]) with Carroll et al.'s method [12] and their impact on O&M simulations

	Average failure rates per WT and year [12]	**Average failure rates per MW and year** [199]	**Multivariate reliability modelling** [207]
Underlying dataset			
Number of WTs	~350	1089	
WT operational years	1768	4244	
Rated capacity (MW)	2-4	up to 9 MW	
Years of operation	0-8	0-18	
Available data period	~2004-2014	2006-2024	
Dataset diversity			
Onshore/ Offshore	offshore	onshore and offshore	
Number of OEMs	1	onshore: 9, offshore: 4	
Technical concepts	Geared WT with induction machine; hydraulic pitch system; LV converter	Drive train concepts: geared, direct drive, hybrid drive; Generator types: DFIG, EESG, PMSG, SCIG; electrical and hydraulic pitch systems; LV and MV converter technologies	
Aspects covered within reliability model			
Subsystems / Subassemblies	19 subassemblies	29 subsystems + overall WT	7 subsystems + overall WT
Classification of maintenance interventions	4 categories: Major Replacement, Major Repair, Minor Repair, No Cost Data	4 categories: Failures (corrective replacements), Major Component Replacements, Corrective Maintenance Rate (excl. Failures), Preventive Maintenance Rate	1 category: Failures (corrective replacements)
Confidence intervals	No	Yes	Yes
Turbine size	Fixed	Variable	Quantified with covariate coefficient
Failure behaviour over time	Provided exemplarily for overall WT and 2 subassemblies	Provided exemplarily for overall WT and 8 subsystems	Modelled as function
Technical concepts	1 technical concept covered	Average values of different technical concepts	Quantified with different covariate coefficients
Available O&M simulation input			
Corrective maintenance reliability figures	Yes	Yes	Yes
Preventive maintenance reliability figures	No	Yes	No
Repair times	Yes	Yes	No
Required technicians	Yes	Yes	No
Repair costs	Yes	No	No
Impact on O&M simulations			
Recommended application on turbine sizes of	3 MW [204]	Interpolation between 1-9 MW possible	Modelling between 1-9 MW possible
Corrective maintenance rate per year for a 3 MW turbine	8.3	8.2	Doubling the rated capacity increases the failure intensity of the overall WT by a factor of 1.51 (neglecting impact of other covariates; large deviations depending on subsystem)
Corrective maintenance rate per year for a 6 MW turbine	Not applicable	16.4	
Corrective maintenance rate per year for a 9 MW turbine	Not applicable	24.6	
Failure behaviour over time	Not applicable for overall O&M simulation	Not applicable for overall O&M simulation	Modelled as function
Technical concepts	Only applicable for same technical concept	Applicable for covered technical concepts	Modelled as function including different covariates

12 Conclusions

This chapter provides a summary of the preceding chapters, aligning with the thesis objectives, with a particular focus on those chapters based on published research papers (Section 12.1). Section 12.2 presents the thesis' contribution to knowledge, research and industry. The chapter concludes with a discussion of future research directions, an outlook, and final remarks (Sections 12.3 and 12.4).

12.1 Summary of the chapters

12.1.1 Investigation and classification of existing reliability figures and reliability assessment methods

Chapter 2 has examined O&M strategies, modelling approaches, and available input data for O&M simulations which had been published in the past. Based on this analysis, key research gaps have been identified. Despite the availability of numerous O&M simulation tools, both commercial and academic ones, continuous development is necessary to integrate emerging technologies, strategies, and research advancements. The output of these tools is highly dependent on the quality of input data, which remains a critical limitation. Reviews of past studies indicate that approximately 20 initiatives have published reliability statistics, yet many datasets are outdated or lack comprehensive coverage of modern turbine technologies. Existing reliability studies suffer from further shortcomings, including limited subsystem coverage, reliance on non-standardised failure rate definitions, and difficulties in extrapolating data to larger turbines due to failure rates being presented per turbine and year rather than per MW of rated turbine capacity. Simple average failure rates, though easy to interpret, fail to account for time-dependent reliability trends, design variations, and operational influences.

Chapter 3 has described the field-data based approach followed within this thesis to address these gaps. Next to failure data analysis of different kinds, this includes standardised data preprocessing – ensuring consistency and applicability – and a digitalisation and classification workflow to accelerate the process from data collection to publication, overcoming delays associated with manual failure classification. Within the scope of the thesis, failure data from a diverse and modern WT fleet has been collected and evaluated. The underlying dataset covers a total of 1335 onshore and offshore wind turbines covering 5539 WT operational years with turbines sizes of up to 9 MW.

12.1.2 Development of a framework for economic feasibility studies of offshore wind farms

Chapter 4 has evaluated the economic viability of extending the operational service-life of offshore wind farms in German waters. To facilitate this analysis, a comprehensive Economic Life Cycle Simulation and Assessment (ELSA) framework has been developed, integrating capital and operational expenditures (CAPEX and OPEX), revenue models, and site-specific deployment factors. By categorising existing offshore wind farms based on size and key characteristics, the study has provided a structured assessment of extended operation feasibility.

The results indicate that service-life extension is economically viable for most German offshore wind farms, emphasising the impact of wake effects on annual energy production, the variability of O&M costs, and the significance of CAPEX and financial modelling. Several assumptions and limitations have been identified which must be addressed to enhance the robustness of the findings:

- Suitability of input data: The analysis heavily depends on input data from literature and industry interviews. More direct input derived from operational wind farms, particularly O&M cost data and maintenance records, would improve the accuracy of the assessment.
- Reliability modelling: The scarcity of publicly available reliability data poses a challenge, as failure rates vary depending on technology type, turbine maturity, and operational age. Enhanced reliability models based on real-world maintenance data of currently deployed turbines, incorporating component aging effects and further impacts, would reduce uncertainties in service-life extension projections.
- Maintenance logistics: Deployment location and supply chain maturity significantly influence maintenance. Continuous updates to O&M simulation tools are necessary to reflect evolving logistical concepts.
- Component lifetime considerations: The study assumes a fixed nominal service life for wind turbines without adjusting for the lifetime of major components (e.g., blades, gearbox, generator). This may lead to unrealistic cost estimations if critical failures occur near the end of service life. Higher-fidelity reliability models should be introduced to assess the economic implications of major component replacements in the context of extended operation.

Overall, the study provides valuable insights into the economic potential of extending offshore wind farm service life, offering a foundation for future research and decision-making in offshore wind energy asset management. This thesis has specifically addressed the limitations related to data input quality to improve the accuracy of future O&M simulations.

12.1.3 Assessment of challenges deriving reliability metrics and impact analysis of differently applied methods for preprocessing and digitalisation of maintenance reports

Chapters 5 and 6 have addressed the challenges posed by heterogeneous and unstructured wind turbine maintenance reports, which vary in information content and depth. To enhance the accuracy of O&M simulations and OPEX modelling, a digitalisation workflow for maintenance information has been proposed, incorporating optical character recognition (OCR), information extraction, and classification. While OCR and information extraction have demonstrated high accuracy, the focus has been placed on classifying maintained components using the guideline of RDS-PP [40], employing a support vector machine (SVM) for text classification. Key findings include:

- Text classifiers have achieved high micro F1 scores when trained on specific datasets but exhibited reduced performance across different wind farms, highlighting the need for context-specific training.
- Smaller, well-curated training datasets have proved to be cost- and resource-efficient while achieving competitive classifier performance, emphasising the importance of balancing manual labelling efforts with model effectiveness.

- Industry feedback has indicated diverse preferences for classifier configurations, necessitating customisable solutions.
- Although text classifiers have tended to overgeneralise, skewing KPI calculations, their integration has reduced efforts, with manual verification improving quality of results.
- The studies underscore the need for standardised maintenance reporting, as both automated and manual methods suffer from inconsistencies. Implementing designation systems like RDS-PP can enhance data accuracy and KPI quality.

By compiling comprehensive datasets that capture technical jargon as well as variations in maintenance reports and refining classification methodologies, this research has contributed to improving maintenance data processing in the wind energy sector. The proposed strategy for preprocessing maintenance records facilitates reliability modelling of currently operated turbines, while acknowledging the uncertainties introduced by data preprocessing practices and limiting those by careful implementation.

12.1.4 Development of reliability models of wind turbine subsystems and chosen components based on real-world O&M data

Chapter 7 has conducted a comprehensive reliability analysis of onshore and offshore wind turbines using an extensive dataset encompassing over 1,000 turbines. By leveraging real-world maintenance records, the research has provided insights into average failure rates, maintenance interventions, and key performance indicators essential for optimising O&M strategies. The analysis has revealed notable differences in failure rates between onshore and offshore WTs, with onshore turbines exhibiting higher average failure rates when normalised per megawatt. Across both environments, the most failure-prone subsystems include the control system, pitch system, and converter system. Additionally, failures in the drive train are predominant in onshore WTs, while lifting gears represent an additional important failure source offshore. The study has further identified temporal failure trends exemplarily for a specific WT type that align with the typical bathtub curve for the overall WT system, with subsystem-specific variations.

A key challenge in reliability studies is the comparability of failure data across different research efforts due to differences in turbine size, technology generations, and failure definitions. This study has contextualised its findings by comparing them with existing literature, illustrating how variation in failure definitions and KPI categorisations can influence reported failure rates. For example, discrepancies with [12] are reconciled by accounting for differences in the scope of corrective maintenance activities, while findings from SPARTA (cf. [29], [13]) have been found to align with this study in identifying critical subsystems and manufacturer-specific reliability variations.

Beyond failure rates, Chapter 7 has provided a detailed assessment of maintenance interventions, differentiating between corrective maintenance interventions with and without spare part replacements and preventive maintenance interventions. The study has highlighted that the power generation system and drive train system are the main contributors for jack-up vessel interventions offshore and has also assessed average repair times and technicians required for maintenance interventions.

This study has addressed an important gap in reliability research by incorporating data from modern WT technologies, whereas much of the existing literature relies on outdated datasets.

The findings contribute to a more accurate and up-to-date understanding of WT reliability, supporting advancements in O&M strategies and turbine design improvements.

Since the pitch system and power converter system have been identified as critical subsystems in both onshore and offshore wind turbines, Chapters 8 and 9 have provided a comprehensive analysis of component failures and their respective failure behaviours.

Chapter 8 has examined failure rates of two pitch system configurations – electrical and hydraulic – based on a large dataset of onshore assets. The analysis has considered turbine rating, seasonal effects, and manufacturer-specific reliability performance. For the electrical pitch system, the most failure-prone components include the battery pack, control/rectifier/inverter/thyristor, and motor protection relay/multifunction relay. In contrast, the hydraulic pitch system exhibits the highest failure rates in components directly related to the hydraulic mechanism, namely the hydraulic accumulator unit/oil tank, pitch cylinder, and hydraulic valve.

Chapter 9 has investigated the reliability of medium-voltage (MV) and low-voltage (LV) power converters in offshore wind turbines by analysing failure rates, seasonal failure patterns, and temporal failure trends using Nelson-Aalen plots. The results indicate that MV converters demonstrate a reliability advantage over LV converters, exhibiting lower failure rates per converter capacity, reduced susceptibility to climatic influences, and a more stable failure behaviour over time.

Chapter 10 has presented a comprehensive analysis of WT reliability at the subsystem level, leveraging failure data from over 1,000 WTs with more than 4,200 operating years. The dataset encompasses turbines with rated capacities of up to 9 MW, operating ages of up to 18 years and is identical to the one of Chapter 7. Failure behaviour over time has been assessed using Nelson-Aalen plots, while the influence of key covariates on reliability has been analysed through non-homogeneous Poisson process (NHPP) modelling combined with a covariate selection procedure.

The findings have revealed distinct subsystem-specific reliability trends, with some exhibiting a classical bathtub curve while others transition directly from early failures to deterioration. NHPP regression results have confirmed the significant impact of multiple covariates on WT reliability, including improved reliability for later commissioning years due to technological advancements, but also increased failure intensity with higher rated power. Offshore turbines have been found to generally demonstrate higher reliability than onshore ones, except for the yaw subsystem. Additionally, subsystem design choices were found to be crucial, with hydraulic pitch systems outperforming electrical ones and direct-drive turbines exhibiting greater reliability than geared-drive turbines in both the drive train and power generation subsystem.

These insights of Chapter 10 highlight the importance of advanced reliability modelling over simplistic failure rate assumptions and underscore the value of large-scale field-data collection. The results provide essential guidance for manufacturers, operators, and maintenance planners, supporting field-data-driven decisions for design optimisation and O&M strategies.

12.1.5 Discussion and evaluation of the impact using different reliability models for O&M simulations

Chapter 11, along with Section 10.3.3, has assessed the impact of using different reliability modelling approaches for O&M simulations and subsequent decision making. The two reliability modelling concepts of the thesis have been compared: a simpler method using average failure rates per MW and year and a more advanced multivariate approach utilising NHPP regression modelling. While average failure rates offer a straightforward baseline, they fail to account for time-dependent failure behaviour and influencing factors such as turbine size, technological advancements, and operational conditions.

The comparison highlights the limitations of traditional reliability assessments and demonstrates the advantages of NHPP regression models, which enable a more accurate representation of reliability. In particular, the results confirm that failure intensity increases with turbine size but not in a strictly proportional manner, as suggested by the per-MW approach.

When used as input for O&M simulations, the choice of reliability modelling methodology significantly impacts maintenance planning, spare parts logistics, and cost estimations. Relying on average failure rates for O&M simulations, without accounting for time-dependent behaviour, turbine size, or other influencing factors, can result in:

- Misestimation of maintenance interventions, leading to inefficient O&M strategy development
- Inaccurate spare parts planning, increasing the risk of costly downtimes due to component unavailability
- Suboptimal long-term financial planning, as O&M cost projections may not accurately capture actual failure trends and influencing factors

Compared to the widely used reliability figures of Carroll et al. [12], which were derived from smaller turbines (~3 MW), the models developed in this thesis provide updated failure rates for larger turbines of up to 9 MW, addressing a critical data gap. The research underscores the importance of integrating time-dependent and covariate-based failure modelling into O&M simulations to improve cost-effectiveness and operational efficiency. However, if a simplified approach is required, scaling failure rates per MW remains preferable to assuming constant failure rates across different turbine sizes. Finally, ensuring that O&M simulation inputs are based on datasets reflecting turbines of similar rated power is essential to maintaining accuracy and avoiding the risks associated with extrapolation to significantly larger turbines.

12.2 Thesis contributions to knowledge, research, and industry

The contributions of this thesis to both academic knowledge and industrial applications are outlined in the following section. Each of the defined objectives (Section 1.3), which have been successfully achieved, is evaluated in terms of its novelty, scientific soundness, and practical relevance, as summarised in **Table 22**. The findings, methodologies developed, and insights gained throughout this research have been disseminated through multiple publications in peer-reviewed scientific journals, as well as oral presentations at scientific conferences, as detailed in Section 1.4 and Appendix A.

Table 22. Contribution to knowledge, research and industry of this thesis' research

Objective	Novelty	Scientific soundness	Value / Stakeholders
Investigate and classify existing reliability figures for onshore and offshore wind turbines and derive suitable reliability assessment methods and metrics	- A comprehensive review of existing reliability data has been conducted, highlighting the shortcomings and limitations of previously published studies. - Advantages and disadvantages of different reliability assessment methods and metrics have been thoroughly investigated.	- A systematic literature review is conducted through a comprehensive analysis of existing studies and comparative assessment with similar review studies. - A transparent overview of prior work in the field of reliability modelling is given, presented both as an overarching review at the beginning of the thesis and as topic-specific evaluations within individual chapters.	- The review is not only relevant for researchers and academics but also provides valuable insights for wind farm developers and operators by highlighting the limitations of existing reliability studies and the necessity for further research.
Develop a framework for economic feasibility studies of offshore wind farms considering relevant input parameters to quantify their impact on output	- A comprehensive Economic Life Cycle Simulation and Assessment (ELSA) framework has been developed, integrating a cost-revenue model that accounts for both CAPEX and OPEX components, as well as revenue factors. - The framework incorporates wake and blockage effects and is used to evaluate all offshore wind farms in the German North and Baltic Seas.	- A comprehensive economic life cycle simulations and assessment framework has been developed to evaluate the economic feasibility of wind farm operation. - To assess the potential for extended operation of all German offshore wind farms, a classification of existing wind farms in German waters has been conducted based on their size and key characteristics. Model inputs have been sourced from literature and refined through stakeholder interviews. - The limitations of both the model and input data have been examined, with their impact on the results quantified.	- Beyond benefiting researchers, academics, and wind farm developers as well as operators, this study has provided relevant insights for the German Federal Maritime and Hydrographic Agency (BSH) in the context of the "Further Development of the Framework Conditions for the Planning of Offshore Turbines and Grid Connection Systems". The research has been conducted as part of an advisory project supporting the development of the Site Development Plan and was subsequently published as an annex to the plan[8], serving as a foundation for the decisions made within its framework.

[8] BSH - Flächenentwicklungsplan - Endbericht - Weiterentwicklung der Rahmenbedingungen zur Planung von Windenergieanlagen auf See und Netzanbindungssystemen

Objective	Novelty	Scientific soundness	Value / Stakeholders
Assess the challenges of deriving reliability metrics and evaluate the impact of differently applied methods for preprocessing and digitalisation of maintenance reports	- A digitalisation workflow has been developed to convert heterogeneous, unstructured, and non-standardised maintenance reports into a machine-readable data framework. This framework includes classified components maintained during turbine visits and standardised maintenance activities. - The feasibility of using text classifiers for preprocessing wind turbine maintenance reports has been evaluated, demonstrating their potential to significantly reduce manual data preprocessing efforts. - The influence of different classification methods on reliability key performance indicators has been analysed.	- Preliminary tests have been conducted using three different model architectures – support vector machine (SVM), convolutional neural network (CNN), and a fine-tuned transformer variant (XLM-RoBERTa). Following evaluation, the SVM approach has been selected for implementation due to its superior performance in this context and lower computational demands. - Classification models have been trained on 26 different test scenarios with varying training datasets to assess performance, including the applicability of derived text classifiers to new datasets from other wind farms. - Manually labelling has been compared to automated text classification, analysing the impact of different preprocessing approaches on failure rate calculations to better understand their advantages and limitations.	- This study has generated important insights for machine learning and text classification researchers by applying existing model architectures to the specific context of wind turbine maintenance records. The findings challenge common assumptions in machine learning for this application, contributing to the further refinement of text classification models. Additionally, the research offers practical value for wind farm operators by assessing the suitability of text classifiers, quantifying associated risks, and providing recommendations for the digitalisation of maintenance reports, thereby reducing manual processing efforts. Furthermore, the evaluation of industry interviews on the practical use of classifiers has uncovered an unexpected diversity in preferences for classifier configurations. These findings are particularly relevant for product developers aiming to commercialise text classification services.

Objective	Novelty	Scientific soundness	Value / Stakeholders
Develop reliability models of wind turbine subsystems and chosen components based on real-world O&M data	- A unique dataset has been compiled and utilised in this thesis, encompassing a total of 1335 onshore and offshore wind turbines with 5539 years of operation. The dataset spans from 2006 to 2024 and includes turbines with rated capacities of up to 9 MW, covering various turbine designs and manufacturers. Compared to datasets used in previous reliability studies, this dataset offers an unparalleled level of diversity, scale, and recency. - Leveraging real-world data from onshore and offshore wind turbines, this thesis has provided a comprehensive analysis of failure rates, repair times, and maintenance resource requirements. O&M simulation input has been developed for 29 different subsystems, covering major component replacements, as well as corrective and preventive maintenance interventions. - The failure behaviour of the entire wind turbine system and the seven most important subsystems has been analysed over time using Nelson-Aalen plots. The impact of various covariates on reliability has been assessed through a non-homogeneous Poisson process model, incorporating a systematic covariate selection procedure to identify factors with a significant effect. - Two in-depth reliability analyses have been conducted for the pitch and converter subsystems, identifying the most frequently failing components. A comparative assessment of electrical and hydraulic pitch systems has been performed, along with an evaluation of the influence of OEMs, wind turbine rated capacity, and seasonal patterns. Additionally, for the first time, the reliability of medium-voltage and low-voltage power converters has been systematically compared, providing detailed insights into the distribution of failures, seasonal trends, and failure behaviour over time.	- Datasets have been systematically curated following established standards and guidelines, including the Reference Designation System for Power Plants (RDS-PP) and the ZEUS state-event cause code, ensuring comparability and clarity of the results. - Different influences on failure rates have been systematically analysed. Given the strong correlation between failure rates and turbine rated capacity, failure rates have been presented per MW and year to enhance comparability and applicability of results. - Aggregated failure rates have been derived from a diverse fleet of turbines, encompassing multiple turbine types and manufacturers, ensuring the representativeness and robustness of the findings. - Confidence intervals for the failure rates have been computed to quantify the uncertainty stemming from the size of the datasets. - For the development of multivariate reliability models, a subsampling routine has been implemented to account for uncertainties in the covariate selection procedure. This approach generates inclusion rate plots, which quantify the likelihood of a given covariate significantly influencing the reliability of a specific subsystem.	- This thesis highlights the critical need for the continuous collection and analysis of large-scale field data to improve wind turbine reliability. - The findings offer essential insights for manufacturers, operators and maintenance planners, facilitating field-data based decisions for design optimisation, O&M strategies, and lifetime extension initiatives. Furthermore, the findings are of significant value to wind farm developers who need to make assumptions about reliability for OPEX modelling and business case evaluations, particularly for new market entrants with limited operational experience. - This study supports researchers, academics, and industry professionals, by providing updated reliability figures and O&M simulation inputs. It offers a viable alternative to the widely used but from today's perspective outdated study of Carroll et al. published in 2015, addressing the lack of more recent data sources.

Objective	Novelty	Scientific soundness	Value / Stakeholders
Evaluate the impact of using different reliability inputs for O&M simulations by comparing the developed reliability models with previous published ones	- A comparative analysis of three distinct reliability modelling approaches has been conducted: the two methods developed in this thesis – average failure rates per MW of rated turbine capacity per year, and multivariate reliability modelling to separate and quantify the effects of factors such as design, technological advancements, rated capacity, and operating age – alongside the widely used study by Carroll et al. - The impact of applying these different models in O&M simulations has been assessed. The limitations of average failure rates and the advantages of NHPP regression modelling have been discussed. Based on these findings, key factors that must be considered in O&M simulations have been identified, along with the potential consequences of neglecting these considerations. Additionally, recommendations are provided on deciding to use failure rates per WT or per MW when a simplified approach is desired.	- A thorough comparison of three distinct reliability modelling approaches has been conducted by systematically evaluating the models across five key subject areas, encompassing a total of 25 subcategories. - The recommendations provided are grounded in the results of the field-data based analyses derived from a large and representative dataset, ensuring a solid foundation for the conclusions drawn.	- A decision framework for selecting appropriate reliability models as input for O&M simulations has been presented, offering valuable guidance to academics, researchers, and industry practitioners. - The consequences of using average failure rates for O&M simulations, especially when not adjusted for time-dependent behaviour, turbine size, or other influencing factors, have been outlined. This supports stakeholders across the value chain in understanding the associated risks and limitations and helps wind farm developers and operators grasp the impact on O&M strategy development and long-term financial planning.

12.3 Future work and outlook

Building upon the findings of this thesis, several key areas for future research have been identified to enhance the applicability and accuracy of wind turbine reliability modelling.

A primary focus will be on continuously updating the dataset to ensure the relevance of the derived reliability figures and models also in the future. Given the rapid technological advancements in wind energy, maintaining a dataset that reflects the latest turbine generations is crucial for supporting effective O&M strategies of recent and future wind farms.

Further efforts will be directed toward a more detailed assessment of specific subsystems, following the approach already applied to the pitch and power converter systems. This will involve breaking down failures to the component level, enabling a more granular understanding of failure modes, failure mechanisms and their contributing factors. Additionally, while the current NHPP regression models have been applied with a limited set of covariates focused on critical subsystems, future work will extend this methodology to additional subsystems as well as components and integrate new covariates, particularly those related to operating conditions. This approach will go beyond time- and design-based failure modelling by incorporating environmental and load-dependent influences.

Further advancing reliability modelling methodologies is another key area of future research. Studies such as [206] and [43] have demonstrated the potential benefits of incorporating time-dependent covariates in NHPP regression models and modelling distinct failure behaviour of different phases of the bathtub curve. These approaches have provided new insights in power converter reliability, such as the influence of electrical utilisation on converter reliability primarily during the deterioration phase whereas environmental factors predominately influence early failures. Building on these findings, further methodological improvements will be explored, including refined strategies for integrating reliability models into O&M simulations.

Beyond methodological advancements, the research outcomes have practical implications for various applications, including wind-turbine design, wind farm development, lifetime extension strategies, and reliability control. A particularly promising avenue is the development of site-specific reliability models. The RUN25+[9] project aims to implement a novel two-stage Bayesian approach, combining prior knowledge from a large turbine fleet (A-priori reliability model) with wind-farm-specific operational and failure data (A-posteriori reliability model). This will allow for the generation of wind farm-specific reliability models that account for key trends in failure behaviour and relevant covariate effects while overcoming the limitations of small-scale datasets.

Overall, these future developments will further refine wind turbine reliability analysis and support the continued cost reduction and operational optimisation of wind energy systems.

12.4 Concluding remarks

The findings of this thesis highlight the critical importance of continuously collecting and analysing field data to improve wind turbine reliability modelling and operational decision-making. Ensuring that data is systematically gathered in a machine-readable and standardised

[9] RUN25+

format from the beginning of wind farm operation enhances efficiency and facilitates more valuable analyses.

To maximise the value of reliability assessments, wind farm developers and operators should prioritise data availability and clarify data ownership between operator and OEM early in the project development phase. Establishing clear protocols for data access and management will enable more comprehensive long-term analyses, ultimately supporting better-informed O&M strategies.

Furthermore, the choice of input parameters for O&M simulations and OPEX modelling has a significant influence on the overall business case of wind energy projects. Accurate and field-data based reliability figures are essential for optimising O&M strategies, reducing downtime, and improving cost efficiency. By leveraging high-quality, representative datasets, stakeholders can ensure that their reliability models reflect real-world conditions, leading to more effective decision-making and improved operational performance.

Appendix A – Additional dissemination activities

Additionally to the paper publications in conference proceedings and scientific journals listed in Section 1.4, parts of the research work were presented at the following conferences:

- Wind Energy Science Conference WESC 2021, Hannover: Julia Walgern, Paul Hentschel, Katharina Fischer "Reliability of electrical and hydraulic pitch systems in wind turbines"
- Ocean Energy and Maritime Transport Research Conference OEMT 2022, Glasgow: Julia Walgern "Analysis of Uncertainty and Impact on Reliability KPI Calculation using Text Classifiers for Standardising Maintenance Information of Wind Turbines"
- EERA DeepWind Conference 2023, Trondheim: Julia Walgern, David Baumgärtner, Johannes Fricke, Niklas Requate, Martin Dörenkämper, Tobias Meyer, Lukas Vollmer "Economic feasibility study for continued operation of German offshore wind farms"
- Wind Energy Science Conference WESC 2023, Glasgow: Julia Walgern, Karoline Pelka, Volker Berkhout, Linda Rülicke, Joshua Gelhaar, Timo Lichtenstein "Findable, Accessible, Interpretable, and Reusable Wind Energy Data Utilising a Data Trust Model Based on International Data Spaces Infrastructure"
- Forschungsverbund Erneuerbare Energien (FVEE)-Jahrestagung 2024 (Renewable Energy Research Association Conference 2024), Berlin: Julia Walgern "Offshore-Windenergie - Technologien für Gigawatt-Windparks“
- Wind Energy Science Conference WESC 2025, Nantes: Julia Walgern, Fraser Anderson, Katharina Fischer "Field-data based wind turbine reliability modelling: Quantifying effects of design, technological development, operating age and environmental and load conditions"

Further involvement in publications during the EngD but which were not directly used for this thesis is listed below:

- M. A. Lutz, K. Beckh, J. Kindermann, J. Schneider, J. Walgern, S. Pfaffel, S. Faulstich and A. Staak, 2021. "Digitalisierungsworkflow zur Strukturierung und Standardisierung von Instandhaltungsinformationen von Windenergieanlagen". Gesellschaft für Informatik, Lecture Note Informatics, pp. 229-249, 2021.
- Katharina Fischer, Karoline Pelka, Julia Walgern, 2023. "Trends and Influencing Factors in Power-Converter Reliability of Wind Turbines". PCIM Europe 2023, doi: 10.30420/566091068.
- Julia Walgern, 2025. "Offshore-Windenergie - Technologien für Gigawatt-Windparks“ Forschungsverbund Erneuerbare Energien (FVEE)-Jahrestagung 2024 (Renewable Energy Research Association Conference 2024), Tagungsband-Beitrag, doi: 10.5442/t2024.
- Fraser Anderson, Karoline Pelka, Julia Walgern, Timo Lichtenstein, Katharina Fischer, 2024. "Trends and Influencing Factors in Power-Converter Reliability of Wind Turbines: A Deepened Analysis" IEEE Transactions on Power Electronics, doi: 10.1109/TPEL.2025.3530163.

- Fraser Anderson, Julia Walgern, Katharina Fischer, 2025. “Early, Intrinsic and Deterioration Stage Wind Turbine Reliability Models: A Case Study for the Converter System”, Submitted to Journal of Physics (DeepWind 2025) for publication.

Furthermore, collaborative research and industry projects were conducted during the course of the EngD, though they were not directly incorporated into this thesis. The published final project reports are listed below:

- Christian Broer, Kirsten Dehning, Katharina Fischer, Sören Fröhling, Nando Kaminski, Benedikt Kostka, Sebastian Kremp, Timo Lichtenstein, Axel Mertens, Karoline Pelka, Jan-Hendrik Peters, Oliver Schilling, Bernd Tegtmeier, Jörg Thiele, Julia Walgern, Johannes Wenzel, Stefan Zimmermann, Christian Zorn, 2023. “ReCoWind – Zuverlässige Frequenzumrichter für Windenergieanlagen”, Final project report, doi: 10.24406/publica-1961.
- Marc-Alexander Lutz, Daniel Zahn, Katharina Beckh, Jörg Kindermann, Juliane Schneider, Julia Walgern, Andreas Kluge, Falko Feßer, Holger Thiemann, 2023. “DigMa – Digitalisierung von Instandhaltungsinformationen”, Final project report.
- Martin Dörenkämper, Tobias Meyer, David Baumgärtner, Johanna Borowski, Christian Deters, Enno Dietrich, Johannes Fricke, Florian Hans, Torben Jersch, Mareike Leimeister, Mohsen Neshati, Georg Pangalos, Tulio Quiroz, Gesa Quisdorf, Niklas Requate, Jonas Schmidt, Marco Schnackenberg, Sandra Schwegmann, Severin Spill, Philipp Thomas, Lukas Vollmer, Julia Walgern, Viktor Widerspan, 2022. “Weiterentwicklung der Rahmenbedingungen zur Planung von Windenergieanlagen auf See und Netzanbindungssystemen”, 2023. Final project report, doi: 10.24406/publica-2202.
- Alexander Arzt, Volker Berkhout, Joshua Gelhaar, Angelina Göbel-Knapp, Sebastian Haugk, Abderrahmane Khiat, Timo Lichtenstein, Tasneem Tazeen Rashid, Linda Rülicke, Waleed Shabbir, Julia Walgern, 2024. “FAIRWinDS – Findable, Accessible, Interpretable and Reusable Wind Energy Data in Data Spaces”, Final project report.
- Jannik Barthel, Katharina Fischer, Sören Fröhling, Karoline Pelka, Juliane Schneider, Bernd Tegtmeier, Jannes Vervoort, Julia Walgern, Christian Zorn, Victoria Zimmermann, Tamara Reck, Michael Jank, Bianca Böttge, Sandy Klengel, Elisabeth Giebel, Falk Naumann, Felix Kulenkampff, Sebastian Franz, Stefan Wagner, Frederic Sehr, Amrita Bohn, 2024. “Zuverlässige Umrichter für die regenerative Energieversorgung (power4re)”, Final project report.

Appendix B – Wind turbine clustering into representative generic WT models

The following aspects are taken into account in the procedure for grouping the turbines into three WT size classes: First, the entire power class spectrum of the existing WTs should be covered. Second, since the overall lifetime extension evaluation concerns the service life and thus the fatigue loads of the WTs, the focus of the classification is on mapping comparable loads. Therefore, comparable rotor diameters and hub heights are particularly relevant. Furthermore, technologies (direct vs. geared drive) should preferably match within one generic turbine type. Last, the representation of the different WT types by only three generic WTs necessarily involves deductions in the accuracy of the results per WF, since the generic WTs are only representative of, but not equal to, the real WTs and, moreover, are intended to represent several different WTs in each case. In order to reduce additional uncertainties inherent in the generic models, existing generic reference WT models that have already been used for many years should preferably be used as generic WTs, provided that they can represent the existing WTs, i.e., the above two aspects regarding grouping can be fulfilled.

Based on the requirements and aspects listed, the existing WTs are grouped into three size classes as colour-coded in **Table 23**. The weighted mean of rated power, rotor diameter and hub height is computed by using the number of turbines as weighting factor and is compared with the properties of the generic wind turbines. The smallest class thus comprises the SWT 2.3-93 and the 3.6 MW WTs. While the rotor diameter is comparable, the hub height and the power of the SWT 2.3 is significantly lower than the other turbines in the same class. Due to the low number of turbines, no additional generic model is created for this turbine. The second class includes all WTs with nominal outputs between 4.0 MW and 5.23 MW, plus WTs with 6.15 MW rated capacity. The intermediate WTs with 6.0 MW output are grouped in the highest class due to their significantly larger rotor diameters and higher hub heights. The third class thus includes, in addition to WTs with 6.0 MW output, all WTs with rated capacity from 6.3 MW up to 9.0 MW.

Table 23 also lists the power classes, rotor diameters and hub heights of the generic wind turbines belonging to the three classes. The following turbines are therefore used as generic wind turbine models:

- Lowest power class: 3.6 MW capacity, 120 m rotor diameter, 90.0 m hub height
- Since none of the existing generic reference wind turbines representatively depicts the characteristics of the lowest power class, a new generic wind turbine model is derived and created for this class from the existing reference wind turbines. The generation of this new generic wind turbine model is mainly done by applying scaling factors. To check the plausibility of the dynamic behaviour of the generic wind turbine model, selected load cases have been calculated based on the verification process developed and applied at Fraunhofer IWES [208]. In particular, this includes simulations with deterministic, stepwise increasing wind speeds as well as with turbulent wind fields.
- Medium power class: 5.0 MW capacity, 126 m rotor diameter, 90.0 m hub height
- The medium power class is represented by the NREL 5 MW reference wind turbine [209]. A generic wind turbine model exists for this reference wind turbine.
- Highest power class: 7.5 MW capacity, 164 m rotor diameter, 102.5 m hub height

- The highest power class is represented by the IWT-7.5-164 reference wind turbine [210]. The reference wind turbine was developed at Fraunhofer IWES and has already been used in various research projects. For the specified hub height, the monopile foundation structure from the SeaLOWT joint project [211] is used.

Table 23. Grouping of turbines in the existing German wind farms in three turbine size classes

Rated power (kW)	No.	Rotor diameter (m)			Hub height (m)			
		Min	Max	Mean	Min	Max	Mean	
2300	21	93.0	93.0	93.0	67.0	67.0	67.0	
3600	430	120.0	120.0	111.6	78.3	91.0	81.0	
4000	150	120.0	130.0	124.8	89.0	95.0	91.9	
5000	212	116.0	126.0	118.5	90.0	92.0	91.2	
5230	70	135.0	135.0	135.0	97.5	97.5	97.5	
6000	230	151.0	154.0	153.1	102.0	112.0	106.5	
6150	120	126.0	126.0	107.1	89.0	96.2	78.5	
6300	60	154.0	154.0	151.4	102.0	102.0	100.3	
6330	32	152.0	152.0	152.0	104.5	104.5	104.5	
7000	87	154.0	154.0	154.0	105.0	105.0	105.0	
8000	56	164.0	164.0	164.0	117.0	117.0	117.0	
8400	31	164.0	164.0	164.0	108.0	108.0	108.0	
9000	38	167.0	167.0	167.0	107.5	107.5	107.5	
3539.5				84.2			111.6	**Weighted Mean**
3600.0				*90.0*			*120.0*	***Generic model***
5007.4				89.4			119.8	**Weighted Mean**
5000.0				*90.0*			*126.0*	***Generic model***
6779.0				106.7			155.8	**Weighted Mean**
7500.0				*102.5*			*164.0*	***Generic model***

The power curves of the reference wind turbines NREL 5 MW and IWT-7.5-164 used for the medium and highest power classes as well as the power curve of the scaled 3.6 MW reference turbine are shown in **Figure 54**.

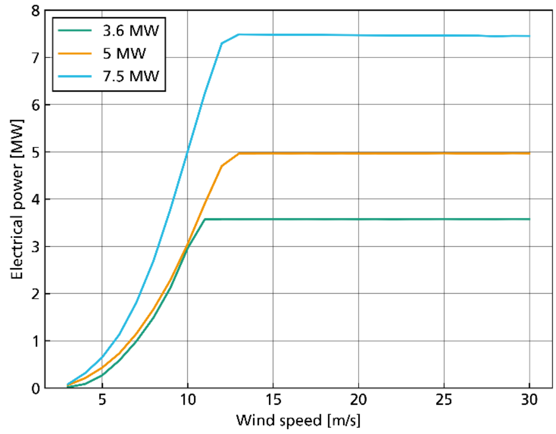

Figure 54. Power curves of the reference wind turbines

Appendix C – Structure of RDS-PP, exemplary classification and codes

C.1 Structure of RDS-PP

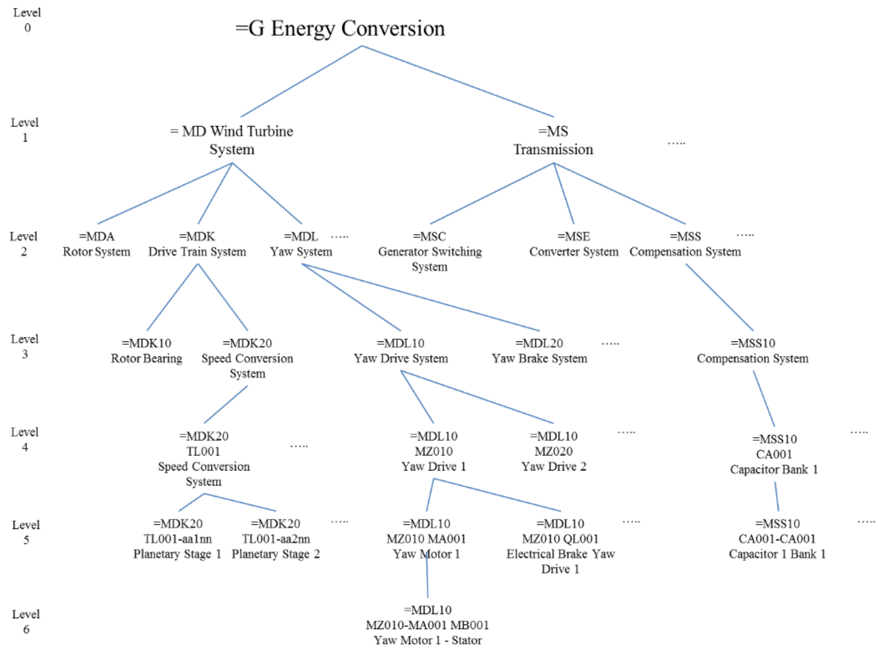

Figure 55. Overview of the hierarchical structure of RDS-PP ([40], [71])

C.2 Exemplary classification

An example for a possible classification of the TDoMM in WT service reports to RDS-PP® is given in **Table 24**.

Table 24. Example for the classification of the TDoMM according to RDS-PP

	Measure 1	Measure 2
TDoMM	Inspection of surface damage at azimuth system.	Failure due to icing at the rotor blade one. WT shut down.
RDS-PP® Label	Yaw System	Rotor Blade System 1
RDS-PP® ID	=MDL	=MDA11

C.3 Reference designation system RDS-PP

For each maintenance measure, the concerned components are classified using the reference designation system RDS-PP for wind turbines [40]. In **Table 25** all mentioned RDS-PP codes mentioned within Chapter 6 are summarised and translated.

Table 25. Summary and translation of all mentioned RDS-PP codes within Chapter 6

RDS-PP code	Translation
CKJ	Environmental measuring system
MDA	Rotor system (incl. pitch system)
MDA11	Rotor blade system 1
MDK	Drive train system (incl. main bearing and gearbox)
MDL	Yaw system
MDV	Central lubrication system
MDX	Central hydraulic system
MDY	Control system
MKA	Power generation system (incl. generator)
MSC	Generator switching system
MSE	Converter system
MSE10	Converter system overall, also denoted as “phase module” components including core power electronics (see [104])
MSE10 KF001	Control system converter system overall
MSE40	Heating/cooling converter systems
MSS	Compensation system
MST	Generator transformer system
MUD	Nacelle
UMD	Tower system
WBA	Personnel rescue systems
XGM	Fire extinguishing system
XMM	Lifting gears
XSD	Obstacle warning system
G	Overall system energy conversion

References

[1] Global Wind Energy Council, “GWEC | Global Wind Report 2024,” Global Wind Energy Council, Brussels, 2024.

[2] G. Costanzo and G. Brindley, “Wind energy in Europe - 2023 Statistics and the outlook for 2024-2030,” WindEurope, 2024.

[3] Federal Network Agency Germany, “Erneut dynamisches Gebotsverfahren bei Ausschreibung von Offshore-Windenergie,” 7 June 2024. [Online]. Available: https://www.bundesnetzagentur.de/SharedDocs/Pressemitteilungen/DE/2024/20240607_OffshoreBK6.html. [Accessed 25 September 2024].

[4] W. Pluta, “Siemens Gamesa baut Riesen-Windrad,” 11 February 2025. [Online]. Available: https://www.heise.de/news/Siemens-Gamesa-baut-21-5-Megawatt-Windrad-in-Daenemark-10278240.html?wt_mc=rss.red.ho.ho.rdf.beitrag.beitrag. [Accessed 12 February 2025].

[5] T. Stehly, P. Duffy and D. Mulas Hernando, “Cost of Wind Energy Review: 2024 Edition,” National Renewable Energy Laboratory, US, 2024.

[6] Federal Maritime and Hydrographic Agency, *Site Development Plan,* 2022.

[7] *ZGzÄWindSeeG,* 2022. Zweite Gesetz zur Änderung des Windenergie-auf-See-Gesetzes udn anderer Vorschriften. (Second amendment to the Wind Energy at Sea Act)

[8] R. Pandit and J. Wang, “A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights,” *IET Renewable Power Generation,* vol. 18, pp. 722-742, 2023.

[9] F. Spinato, P. Tavner, G. ban Bussel and E. Koutoulakos, “Reliability of wind turbine subassemblies,” *IET Renewable Power Generation,* vol. 3, no. 4, 2009.

[10] J. B. Gayo, “Final Publishable Summary of Results of Project ReliaWind,” 2011.

[11] S. Faulstich, B. Hahn and P. J. Tavner, “Wind turbine downtime and its importance for offshore deployment,” *Wind Energy,* vol. 14, no. 3, pp. 327-337, 2011.

[12] J. Carroll, A. McDonald and D. McMillan, “Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines,” *Wind Energy 19,* pp. 1107-1119, 2016.

[13] SPARTA, “Portfolio Review 2020/21: System Performance, Availability and Reliability Trend Analysis,” ORE Catapult, UK, 2022.

[14] H. Braam, L. Rademakers, T. Obdam and R. van d. Pieterman, “Operation and Maintenance Cost Estimator (OMCE) - Final Report,” 2009. [Online]. Available: http://resolver.tudelft.nl/uuid:ccd513a3-73d9-4fd8-bea2-947940f2f83b.

[15] M. Hofmann and I. Sperstad, “NOWIcob - A Tool for Reducing the Maintenance Costs of Offshore Wind Farms,” *Energy Procedia,* pp. 177-186, 2013.

[16] I. A. Dinwoodie and D. McMillan, “Operational strategies for offshore wind turbines to mitigate failure rate uncertainty on operational costs and revenue,” *IET Renewable Power Generation,* vol. 8, no. 4, pp. 347-441, 2014.

[17] M. Wiggert, L. Meyer-Lerbs, G. Wolken-Möhlmann and E. Dietrich, “Offshore TIMES - Offshore Transport, Inspection and Maintenance Software : Abschlussbericht,” Fraunhofer-Institut für Windenergiesysteme, Bremerhaven, 2018.

[18] M. Shafiee, F. Brennan and I. A. Espinosa, “A parametric whole life cost model for offshore wind farms,” *International Journal of Life Cycle Assessment,* pp. 961-975, 2016.

[19] A. Ioannou, A. Angus and F. Brennan, “A lifecycle techno-economic model of offshore wind energy for different entry and exit instances,” *Applied Energy,* pp. 406-424, 2018.

[20] B. Hahn, T. Welte, S. Faulstich, P. Bangalore, C. Boussion, K. Harrison, E. Miguelanez-Martin, F. O'Connor, L. Pettersson, C. Soraghan, C. Stock-Williams, J. Dalsgaard Sørensen, G. van Bussel

and J. Vatn, "Recommended practices for wind farm data collection and reliability assessment for o&m optimization," *Energy Procedia,* no. 137, pp. 358-365, 2017.

[21] A. Clifton, S. Barber, A. Bray, P. Enevoldsen, J. Fields, A. M. Sempreviva, L. Williams, J. Quick, M. Purdue, P. Totaro and Y. Ding, "Grand challenges in the digitalisation of wind energy," *Wind Energy Science,* vol. 8, pp. 947-974, 2023.

[22] S. Sørum, E. Smilden, J. Amdahl and A. Sørensen, "Active load mitigation to counter the fatigue damage contributions from unavailability in offshore wind turbines," in *ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering*, 2018.

[23] D. Bendlin, K. Hebig, G. Wolken-Möhlmann and J. Marx-Gómez, "Simulating the uncertain - present status of operation and maintenance simulation for offshore wind farms," Springer, 2018, pp. 1-15.

[24] Shoreline, "The only full lifecycle SaaS solution built for wind farm design, construction, O&M," 2025. [Online]. Available: https://shoreline.no/.

[25] P. Tavner, J. Xiang and F. Spinato, "Reliability Analysis for Wind Turbines," *Wind Energy,* vol. 10, pp. 1-18, 2007.

[26] Windstats, "Windstats survey," [Online]. Available: www.windstats.com (not available anymore).

[27] Landwirtschaftskammer Schleswig-Holstein, Germany, "LWK Survey," [Online]. Available: http://www.lwksh.de/cms/index.php?id=1743 (not available anymore).

[28] M. D. Reder, E. Gonzalez and J. J. Melero, "Wind Turbine Failures - Tackling current Problems in Failure Data Analysis," *Journal of Physics: Conference Series: The Science of Making Torque from Wind (TORQUE 2016),* vol. 753, no. 7, p. 072027, 2016.

[29] SPARTA, *Portfolio Review 2016: System Performance, Availability and Reliability Trend Analysis,* Northumberland, UK: SPARTA Project, 2017.

[30] F. Anderson, R. Dawid, D. McMillan and D. Garcia Cava, "On the Sensitivity of Wind Turbine Failure Rate Estimates to Failure Definitions," *Journal of Physics: Conference Series,* p. 2626, 2023.

[31] S. Pfaffel, S. Faulstich and K. Rohrig, "Performance and reliability of wind turbines: a review," *Energies,* vol. 10, no. 11, pp. 1-27, 2017.

[32] D. Cevasco, S. Koukoura and A. Kolios, "Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications," *Renewable and Sustainable Energy Reviews,* 2021.

[33] E. Artigao, S. Martin-Martinez, A. Ceña, A. Honrubia-Escribano and E. Gomez-Lazaro, "Failure rate and downtime survey of wind turbines located in Spain," *IET Renewable Power Generation,* vol. 15, pp. 225-236, 2021.

[34] N. Sarma, P. M. Tuohy, O. Özgönenel and Djurović, "Early life failure modes and downtime analysis of onshore type-III wind turbines in Turkey," *Electric Power Systems Research,* vol. 216, p. 108956, 2023.

[35] P. Padman, F. Vanni, E. Echavarria, K. Mortstock and M. Wilkinson, "Benchmarking Pitch System Reliability and Reducing Cost of Energy through Advanced Design," Bilbao, 2016.

[36] K. Fischer, K. Pelka, A. Bartschat, B. Tegtmeier, D. Coronado, C. Broer and J. Wenske, "Reliability of Power Converters in Wind Turbines: Exploratory Analysis of Failure and Operating Data From a Worldwide Turbine Fleet," *IEEE Transactions on Power Electronics,* vol. 34, no. 7, pp. 6332-6344, 07 2019.

[37] G. Pulcini, "Modeling the failure data of a repairable equipment with bathtub type failure intensity," *Reliability Engineering & System Safety,* vol. 71, no. 2, pp. 209-218, 2001.

[38] M. Reder and J. J. Melero, "Modelling Wind Turbine Failures based on Weather Conditions," *Journal of Physics: Conference Series Wind Europe Conference & Exhibition 2017,* vol. 926, p. 012012, 2017.

[39] M. Reder and J. J. Melero, "Modelling the effects of environmental conditions on wind turbine failures," *Wind Energy,* vol. 21, pp. 876-891, 2018.

[40] VGB PowerTech, *VGB-Standard RDS-PP Application Guideline Part 32: Wind Power Plants,* 2014.

[41] FGW e.V. - Fördergesellschaft Windenergie und andere Erneuerbare Energien, *Technical Guidelines for Power Generating Units - State-Event-Cause code for power generating units (ZEUS),* Berlin: FGW e.V. - Fördergesellschaft Windenergie und andere Erneuerbare Energien, 2013.

[42] H. E. Ascher, "Repairable systems reliability," in *Encyclopedia of Statistics in Quality and Reliability*, 2008.

[43] F. Anderson, J. Walgern and K. Fischer, "Early, Intrinsic and Deterioration Stage Wind Turbine Reliability Models: A Case Study for the Converter System," *submitted to Journal of Physics, DeepWind2025 for publication,* 2025.

[44] F. Anderson, J. Walgern and K. Fischer, *Early and Deterioration Stage Wind Turbine Reliability Models: A Case Study for the Converter System,* Trondheim: EERA DeepWInd, 2025.

[45] J. F. Lawless, "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association,* vol. 82, no. 399, pp. 808-815, 1987.

[46] K. Pelka and K. Fischer, "Field-data-based reliability analysis of power converters in wind turbines: Assessing the effect of explanatory variables," *Wind Energy,* vol. 26, no. 3, pp. 310-324, 2023.

[47] F. Anderson, K. Pelka, J. L. T. Walgern and K. Fischer, "Trends and influencing factors in power-converter reliability of wind turbines: A deepened analysis," *IEEE Transactions on Power Electronics,* pp. 1-12, 2025.

[48] Press and Information Office of the Federal Government, "Accelerated expansion - More wind energy at sea," 6 April 2022. [Online]. Available: https://www.bundesregierung.de/breg-en/issues/offshore-wind-energy-act-2024112.

[49] *WindSeeG,* 2016. Gesetz zur Entwicklung und Förderung der Windenergie auf See (Windenergie-auf-See-Gesetz, Wind Energy at Sea Act).

[50] BVG Associates, *Guide to an offshore wind farm,* 2019. [Online]. Available: https://www.thecrownestate.co.uk/media/2861/guide-to-offshore-wind-farm-2019.pdf.

[51] T. Bak, A. Graham, A. Sapronova, M. Florian, J. Dalsgaard Soerensen, T. Knudsen, P. Hou and Z. Chen, "Baseline layout and design of a 0.8 GW reference wind farm in the North Sea," *Wind Energies 20,* pp. 1665-1683, 2017.

[52] B. Jenkins, I. Belton, J. Carroll and D. McMillan, "Estimating the major replacement rates in next-generation offshore wind turbines using structured expert elicitation," *Journal of Physics: Conference Series,* vol. 2362, no. 012020, 2022.

[53] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner and X.-Y. Huang, "A Description of the Advanced Research WRF Model Version 4.3," NCAR, 2021.

[54] Bundesamt für Seeschifffahrt und Hydrographie BSH, "Maritime Spatial Plan 2021," 2021. [Online]. Available: https://www.bsh.de/EN/TOPICS/Offshore/Maritime_spatial_planning/Maritime_Spatial_Plan_2021/maritimespatialplan2021_node.html;jsessionid=1E42A4C84B0FA6DBAF20EB5F1579928B.live21301. [Accessed 2 November 2021].

[55] T. Rubert, D. McMillan and P. Niewczas, "A decision support tool to assist with lifetime extension of wind turbines," *Renewable Energy,* vol. 120, pp. 423-433, 2018.

[56] L. Ziegler, E. Gonzalez, T. Rubert, U. Smolka and J. J. Melero, "Lifetime extension of onshore wind turbines: A review covering Germany, Spain, Denmark, and the UK," *Renewable and Sustainable Energy Reviews,* vol. 82, pp. 1261-1271, 2018.

[57] J. S. Nielsen, N. K. Dimitrov and J. D. Sørensen, "Optimal Decision Making for Life Extension for Wind Turbines," *Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP,* p. 83, 2019.

[58] Fraunhofer Institute for Wind Energy Systems, “Weiterentwicklung der Rahmenbedingungen zur Planung von Windenergieanlagen auf See und Netzanbindungssystemen - Endbericht,” Bundesamt für Seeschifffahrt und Hydrographie BSH, 2023.

[59] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot and M. Bonavita, “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society,* pp. 1999-2049, July 2020.

[60] A. C. Fitch, J. B. Olson, J. K. Lundquist, J. Dudhia, A. K. Gupta, J. Michalakes and I. Barstad, “Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model,” *Monthly Weather Review,* vol. 140, no. 9, p. 017–3038, 2012.

[61] Fraunhofer Institute for Wind Energy Systems, “Farm optimization and extended yield evaluation software (FOXES),” [Online]. Available: https://fraunhoferiwes.github.io/foxes.docs/index.html. [Accessed 19 12 2022].

[62] Informationsportal Erneuerbare Energien, “EEG-Vergütung und Kapazitätszuweisung. Bundesministerium für Wirtschaft und Energie,” 2021. [Online]. Available: https://www.erneuerbare-energien.de/EE/Navigation/DE/Technologien/Windenergie-auf-See/Finanzierung/EEG-Verguetung/eeg-verguetung.html. [Accessed 02 11 2021].

[63] Bundesnetzagentur, “Marktstammdatenregister,” 2021. [Online]. Available: https://www.marktstammdatenregister.de/MaStR. [Accessed 8 November 2021].

[64] Bundesamt für Seeschifffahrt und Hydrographie (BSH), “GeoSeaPortal,” 2021. [Online]. Available: https://www.geoseaportal.de/mapapps/?lang=de. [Accessed 08 November 2021].

[65] Bundesamt für Kartographie und Geodäsie, “Verwaltungsgebiete 1:250000 (Ebenen), Stand 01.01. (VG250 01.01.),” 2021. [Online]. Available: https://gdz.bkg.bund.de/index.php/default/verwaltungsgebiete-1-250-000-ebenen-stand-01-01-vg250-ebenen-01-01.html. [Accessed 8 November 2021].

[66] Energi Data Service, *Elspot Prices.* [Online]. Available: https://www.energidataservice.dk/tso-electricity/Elspotprices

[67] S. Patzack, P. Svoboda, J. Wilms, G. Gerdes, A. Armeni and R. Borrmann, *Unterstützung zur Aufstellung und Fortschreibung des FEP. Zukünftige Rahmenbedingungen für die Auslegung von Offshore-Windparks und deren Netzanbindungssystemen. Endbericht.,* 2020.

[68] PEAK Wind, “[Update 2022] OPEX Benchmark – An insight into the operational expenditures of European offshore wind farms,” 2022. [Online]. Available: https://peak-wind.com/insights/opex-benchmark-an-insight-into-operational-expenditures-of-european-offshore-wind-farms/. [Accessed 02 December 2022].

[69] International Renewable Energy Agency, “Renewable power generation costs in 2019,” 2020. [Online]. Available: https://www.irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019.

[70] Danish Standards Foundation, *A Guide to RDS - Reference Designation Systems. TAG Numbers for Systems in Accordance with the ISO/IEC 81346 Standard Series",* 2020.

[71] M. A. Lutz, K. Beckh, J. Kindermann, J. Schneider, J. Walgern, S. Pfaffel, S. Faulstich and A. Staak, “Digitalisierungsworkflow zur Strukturierung und Standardisierung von Instandhaltungsinformationen von Windenergieanlagen,” *Gesellschaft für Informatik, Lecture Note Informatics,* pp. 229-249, 2021.

[72] M. P. Brundage, T. Sexton, M. Hodkiewicz, A. Dima and S. Lukens, “Technical language processing: Unlocking maintenance knowledge,” *Manufacturing Letters,* no. 27, pp. 42-46, 2021.

[73] M. Navinchandran, M. E. Sharp, M. P. Brundage and T. B. Sexton, “Discovering critical KPI factors from natural language in maintenance work orders,” *Journal of Intelligent Manufacturing,* 2021.

[74] Y. Gao, C. Woods, W. Liu, T. French and M. Hodkiewicz, “Pipeline for Machine Reading of Unstructured Maintenance Work Order Records,” *Proceedings of the 30th European Safety and*

Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference, pp. 1401-1408, 2020.

[75] A. Blanco-M., P. Marti-Puig, K. Gibert, J. Cusidó and J. Solé-Casals, “A Text-Mining Approach to Assess the Failure Condition of Wind Turbines Using Maintenance Service History,” *Energies,* vol. 12, no. 10, p. 1982, 2019.

[76] Q. Wang and C. Li, “Incident detection and classification in renewable energy news using pre-trained language models on deep neural networks,” *Journal of Computational Methods in Sciences and Engineering,* vol. 22, no. 1, pp. 57-76, 2022.

[77] K. Katterbauer, “An integrated NLP Big Data framework for enhancing wind turbine reliability,” 2021.

[78] D. Küçük and Y. Arslan, “Semi-automatic construction of a domain ontology for wind energy using wikipedia articles,” *Renewable Energy,* vol. 62, pp. 484-489, 2014.

[79] G. Ertek, X. Chi, A. N. Zhang and S. Asian, “Text mining analysis of wind turbine accidents: An ontology-based framework,” *IEEE International Conference on Big Data,* pp. 3233-3241, 2017.

[80] G. Ertek and L. Kailas, “Analyzing a decade of wind turbine accident news with topic modeling,” *Sustainability,* vol. 13, no. 22, p. 12757, 2021.

[81] G. L. Cash and M. Hatamian, “Optical character recognition by the method of moments,” *Computer Vision, Graphics, and Image Processing,* vol. 39, no. 3, pp. 291-310, 1987.

[82] V. Govindan and A. Shivaprasad, “Character recognition - a review,” *Pattern Recognition,* vol. 23, no. 7, pp. 671-683, 1990.

[83] S. Mori, H. Nishida and H. Yamada, “Optical character recognition,” in *A Wiley-Interscience publication*, New York and Weinheim, Wiley, 1999.

[84] D. Lopresti, “Optical character recognition errors and their effects on natural language processing,” *International Journal on Document Analysis and Recognition (IJDAR),* vol. 12, no. 3, pp. 141-151, 2009.

[85] ABBYY, [Online]. Available: https://www.abbyy.com.

[86] Kofax, [Online]. Available: https://www.kofax.com/Products/omnipage.

[87] Tesseract. [Online]. Available: https://github.com/tesseract-ocr/tesseract.

[88] C. Patel, A. Patel and D. Patel, “Optical character recognition by open source ocr tool tesseract: A case study,” *International Journal of Computer Applications,* vol. 55, no. 10, pp. 50-56, 2012.

[89] R. Grishman, “Information extraction: Techniques and challenges,” in *Information Extraction*, Berlin, Heidelberg, SpringerLink Bücher, Springer Berlin Heidelberg, 1997, pp. 10-27.

[90] D. Jurafsky and J. H. Martin, Speech and Language Processing (3rd ed. draft), Stanford University, 2024.

[91] M. T. Pazienza, *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology. Lecture Notes in Computer Science,* Berlin, Heidelberg: Springer Berlin Heidelberg, 1997.

[92] S. G. Small and L. Medsker, “Review of information extraction technologies and applications,” *Neural Computing and Applications,* vol. 25, no. 3-4, pp. 533-548, 2014.

[93] J. Oevermann, “Semantic pdf segmentation for legacy documents in technical documentation,” *Procedia Computer Science,* vol. 137, pp. 55-65, 2018.

[94] H. Chao and J. Fan, “Layout and content extraction for pdf documents,” in *Document Analysis Systems VI*, vol. 3163, Berlin, Heidelberg, Springer Belrin Heidelberg, 2004, pp. 213-224.

[95] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in *European conference on machine learning*, Berlin, Heidelberg, Springer Berlin Heidelberg, 1998, pp. 137-142.

[96] E. Leopold and J. Kindermann, “Text categorization with support vector machines. how to represent texts in input space?,” *Machine Learning,* vol. 46, no. 1/3, pp. 423-444, 2002.

[97] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, “Distributed representations of words and phrases and their compositionality,” vol. 26, Curran Associates, Inc., 2013.

[98] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT,* pp. 4171-4186, 2019.

[99] W.-C. Chang, H.-F. Yu, K. Zhong, Y. Yang and I. Dhillon, “X-bert: extreme multi-label text classification with using bidirectional encoder representations from transformers,” *arXiv preprint arXiv:1905.02331,* 2019.

[100] Z. Zhang, J. Liu and N. Razavian, “Bert-xml: Large scale automated icd coding using bert pretraining,” *arXiv preprint arXiv:2006.03685,* 2020.

[101] K. Leahy, C. Gallagher, P. O'Donovan and D. T. J. O'Sullivan, “Issues with data quality for wind turbine condition monitoring and reliability analyses,” *Energies,* vol. 12, no. 2, p. 201, 2019.

[102] M. Wilkinson, B. Hendriks, F. Spinato, E. Gomez, H. Bulacio, J. Roca, P. Tavner, Y. Feng and H. Long, “Methodology and results of the reliawind reliability field study,” *European Wind Energy Conference,* vol. 3, pp. 1984-2005, 2010.

[103] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes and D. Brown, “Text Classification Algorithms: A Survey,” *Information,* vol. 10, no. 4, p. 150, 2019.

[104] K. Fischer, K. Pelka and J. Walgern, “Trends and Influencing Factors in Power-Converter Reliability of Wind Turbines,” Nürnberg, 2023.

[105] J. Walgern, K. Fischer, P. Hentschel and A. Kolios, “Reliability of electrical and hydraulic pitch systems in wind turbines based on field-data analysis,” *Energy Reports,* vol. 9, pp. 3273-3281, 2023.

[106] J. Kenworthy, E. Hart, J. Stirling, A. Stock, J. Keller, Y. Guo, J. Brasseur and R. Evans, “Wind turbine main bearing rating lives as determined by IEC 61400-1 and ISO 281: A critical review and exploratory case study,” *Wind Energy,* pp. 1-19, 2023.

[107] M.-A. Lutz, J. Walgern, K. Beckh, J. Schneider, S. Faulstich and S. Pfaffel, “Digitalization Workflow for Automated Structuring and Standardization of Maintenance Information of Wind Turbines into Domain Standard as a Basis for Reliability KPI Calculation,” *IOP Journal of Physics Conference Series,* 2022.

[108] M.-A. Lutz, B. Schäfermeier, R. Sexton, M. Sharp, A. Dima, S. Faulstich and J. M. Aluri, “KPI Extraction from Maintenance Work Orders - A Comparison of Expert Labeling, Text Classification and AI-Assisted Tagging for Computing Failure Rates of Wind Turbines,” *Energies,* vol. 16, no. 7937, 2023.

[109] E. Ghazizadeh and P. Zhu, “A Systematic Literature Review of Natural Language Processing: Current State, Challenges and Risks,” *Advances in Intelligent Systems and Computing,* vol. 1288, pp. 634-647, 2021.

[110] Z. Zhang, E. Strubell and E. Hovy, “A Survey of Active Learning for Natural Language Processing,” *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022,* pp. 6166-6190, 2022.

[111] L. Shi, Y. Zhu, Y. Zhang and Z. Su, “Fault Diagnosis of Signal Equipment on the Lanzhou-Xinjiang High-Speed Railway Using Machine Learning for Natural Language Processing,” *Complexity,* no. art. no. 9126745, 2021.

[112] R. Siva Balan, K. Walia and K. Gupta, “A Systematic Review on POS Tagging,” *2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022,* pp. 1531-1536, 2022.

[113] D. Li, S. Luo, X. Zhang and F. Xu, “Review of Named Entity Recognition,” *Journal of Frontiers of Computer Science and Technology,* vol. 16, no. 9, pp. 1954-1968, 2022.

[114] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. Acharya, V. Makarenkov and S. Nahavandi, “A review of uncertainty

quantification in deep learning: Techniques, applications and challenges," *Information Fusion,* vol. 76, pp. 243-297, 2021.

[115] L. Breiman, "Random forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001.

[116] M. Amjad, A. Gelbukh, I. Voronkov and A. Saenko, "Comparison of Text Classification Methods Using Deep Learning Neural Networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13452 LNCS,* pp. 438-450, 2023.

[117] Z. Wang and Z. Zhang, "Research Convey on Text Classification Method based on Deep Learning," *7th International Conference on Intelligent Computing and Signal Processing, ICSP 2022,* pp. 285-288, 2022.

[118] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882,* 2014.

[119] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[120] J. Zhang, Y. Li, H. Li, P. Gao, D. Deng and J. Hao, "Text classification method based on improved long short term memory network," *Proceedings of SPIE - The International Society for Optical Engineering, 13105,* vol. art. no. 1310528, 2024.

[121] X. Yuemei, F. Zuwei and C. Han, "A Multi-Task Text Classification Model Based on Label Embedding of Attention Mechanism," *Data Analysis and Knowledge Discovery,* vol. 6, no. 2-3, pp. 105-116, 2022.

[122] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems,* vol. 30, 2017.

[123] N. Arabadzhieva-Kalcheva and I. Kovachev, "Comparison of BERT and XLNet accuracy with classical methods and algorithms in text classification," *Proceedings of the International Conference on Biomedical Innovations and Applications, BIA 2021,* pp. 74-76, 2021.

[124] T. Brown, M. B. N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal and D. ... Amodei, "Language models are few-shot learners," *Advances in Neural Information Processing Systems,* vol. 33, pp. 1877-1901, 2020.

[125] A. Elkhatat, K. Elsaid and S. Almeer, "Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text," *International Journal for Educational Integrity,* vol. 19, no. 1, p. art. no. 17, 2023.

[126] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen and V. ... Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692,* 2019.

[127] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena and P. J. ... Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research,* vol. 21, no. 140, pp. 1-67, 2020.

[128] P. Wang, A. Yang, R. Men, J. Lin, S. Bai, Z. Li, J. Ma, C. Zhou, J. Zhou and H. Yang, "OFA: Unifying Architectures, Tasks, and Modalities Through a Simple Sequence-to-Sequence Learning Framework," *Proceedings of Machine Learning Research,* vol. 162, pp. 23318-23340, 2022.

[129] D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia and D. Testuggine, "Supervised multimodal bitransformers for classifying images and text," *arXiv preprint arXiv:1909.02950,* 2019.

[130] A. Bhavani and B. Santhosh Kumar, "A Review of State Art of Text Classification Algorithms," *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021,* vol. art. no. 9418262, pp. 1484-1490, 2021.

[131] S. Vinod, M. Bouh, F. Hossain, P. Paul and A. Ahmed, "Advancements in Text Classification, A Comprehensive Review," *Humanitarian Technology Conference, R10-HTC,* pp. 679-684, 2023.

[132] Y. Sun, L. Wang, Y. Huang and Z. Zhang, “Multimodal Knowledge Graph Representation Learning with Entity Description,” *Proceedings of SPIE - The International Society for Optical Engineering, 12717,* p. art. no. 127171K, 2023.

[133] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez and M. ... Maleshkova, “Knowledge Graphs,” *arXiv preprint arXiv:2003.02320,* 2021.

[134] Q. Wang, B. Wang and L. Guo, “Knowledge Graph Embedding: A Survey of Approaches and Applications,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 29, no. 12, pp. 2724-2743, 2017.

[135] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research,* no. 12, pp. 2825-2830, 2011.

[136] J. C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,” *Advances in Large Classifiers, MIT Press,* 1999.

[137] H. P. Luhn, “The automatic creation of literature abstracts,” *IBM Journal of Research and Development,* vol. 2, no. 2, pp. 159-165, 1958.

[138] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval,” *Journal of Documentation,* vol. 28, no. 1, pp. 11-22, 1972.

[139] Y.-C. Lin, S.-A. Chen, J.-J. Liu and C.-J. Lin, “Linear Classifier: An Often-Forgotten Baseline for Text Classification,” *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* pp. 1876-1888, 2023.

[140] B. E. Boser, I. Guyon and V. Vapnik, “A training algorithm for optimal margin classifiers,” *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press,* pp. 144-152, 1992.

[141] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One,* vol. 10, no. 3, p. e0118432, 2015.

[142] J. Huang and C. Ling, “Using AUC and accuracy in evaluating learning algorithms,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 17, no. 3, pp. 299-310, 2005.

[143] Z. C. Lipton, C. Elkan and B. Naryanaswamy, “Thresholding classifiers to maximize F1 score,” *arXiv preprint arXiv:1402.1892,* 2014.

[144] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer and V. Stoyanov, “Unsupervised Cross-lingual Representation Learning at Scale,” *arXiv preprint arxiv:1911.02116,* 2019.

[145] Z. Yang, S. Batra, J. Stremmel and E. Halperin, “Surpassing GPT-4 Medical Coding with a Two-Stage Approach,” *arXiv preprint arXiv:2311.13735,* 2023.

[146] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So and J. Kang, “BioBERT: a pre-trained biomedical language representation model for biomedical text mining,” *Bioinformatics,* vol. 36, no. 4, pp. 1234-1240, 2020.

[147] X. Yang, Q. Yuan and C. Zhang, “FinBERT: A Pretrained Language Model for Financial Communications,” *arXiv preprint arXiv:2006.08097,* 2020.

[148] M. Martinez-Luengo, M. Shafiee and A. Kolios, “Data management for structural integrity assessment of offshore wind turbine support structures: data cleansing and missing data imputation,” *Ocean Engineering,* vol. 173, pp. 867-883, 2019.

[149] J. Wang, D. Liang, Z. Li, Z. Hong, X. Song and Y. Deng, “Data-Driven Reliability Assessment of Wind Power System Operation State,” Shanghai, China, 2023.

[150] P. W. Khan and Y.-C. Byun, “A Review of machine learning techniques for wind turbine’s fault detection, diagnosis, and prognosis,” *International Journal of Green Energy,* vol. 21, no. 4, pp. 771-786, 2024.

[151] H. Li and C. Guedes Soares, “Assessment of failure rates and reliability of floating offshore wind turbines,” *Reliability Engineering & System Safety,* vol. 228, p. 108777, 2022.

[152] E. Ayyildiz and M. Erdogan, “A comprehensive approach to evaluate risk mitigation strategies in offshore wind farms using spherical fuzzy decision making analysis,” *Ocean Engineering,* vol. 221, no. 1, p. 118881, 2024.

[153] P. J. Tavner, D. M. Greenwood, M. W. G. Whittle, R. Gindele, S. Faulstich and B. Hahn, “Study of weather and location effects on wind turbine failure rates,” *Wind Energy,* vol. 16, no. 2, pp. 175-187, 2012.

[154] Fraunhofer IEE, *WInD-Pool - Windenergie-Informations-Daten-Pool,* 2018.

[155] B. Hahn, M. Durstewitz and K. Rohrig, “Reliability of Wind Turbines,” *Wind Energy,* pp. 329-332, 2007.

[156] A. Turnbull, C. McKinnon, J. Carrol and A. McDonald, “On the Development of OffshoreWind Turbine Technology: An Assessment of Reliability Rates and Fault Detection Methods in a Changing Market,” *Energies,* vol. 15, p. 3180, 2022.

[157] O. Donnelly, J. Carroll and M. Howland, “Analysing the cost impact of failure rates for the next generation of offshore wind turbines,” *Wind Energy,* vol. 27, no. 7, pp. 695-710, 2024.

[158] M. N. Scheu, A. Kolios, T. Fischer and F. Brennan, “Influence of statistical uncertainty of component reliability estimations on offshore wind farm availability,” *Reliability Engineering & System Safety,* vol. 168, pp. 28-39, 2017.

[159] M. Leimeister and A. Kolios, “A review of reliability-based methods for risk analysis and their application in the offshore wind industry,” *Renewable and Sustainable Energy Reviews,* vol. 91, pp. 1065-1076, 2018.

[160] M. Kavakli and O. T. Gudmestad, “Analysis and Assessment of Onshore and Offshore Wind Turbines Failures,” *International Journal of Energy Production and Management,* vol. 8, pp. 27-34, 2023.

[161] J. Li, J. Bian, Z. Chuang, Y. Jiang and S. Leng, “Impact of pitch actuator fault on 10-MW semi-submersible floating wind turbine,” *Ocean Engineering,* vol. 254, p. 111375, 2022.

[162] J. C. Lopez and A. Kolios, “Risk-based maintenance strategy selection for wind turbine composite blades,” *Energy Reports,* vol. 8, pp. 5541-5561, 2022.

[163] K. Fischer, K. Pelka, S. Puls, M.-H. Poech, A. Mertens, A. Bartschat, B. Tegtmeier, C. Broer and J. Wenske, “Exploring the Causes of Power-Converter Failure in Wind Turbines based on Comprehensive Field-Data and Damage Analysis,” *Energies,* February 2019.

[164] M. N. Scheu, L. Tremps, U. Smolka, A. Kolios and F. Brennan, “A systematic Failure Mode Effects and Criticality Analysis for offshore wind turbine systems towards integrated condition based maintenance strategies,” *Ocean Engineering,* vol. 176, pp. 118-133, 2019.

[165] L. Tremps, B. Yeter and A. Kolios, “Review and analysis of the failure risk mitigation via monitoring for monopile offshore wind structures,” *Energy Reports,* vol. 11, pp. 5407-5420, 2024.

[166] J. Carroll, A. McDonald and D. McMillan, “Reliability Comparison of Wind Turbines With DFIG and PMG Drive Trains,” *IEEE Transactions on Energy Conversion,* vol. 30, no. 2, pp. 663-670, 2015.

[167] J. Carroll, S. Koukoura, A. McDonald, A. Charalambous, S. Weiss and S. McArthur, “Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques,” *Wind Energy,* vol. 22, no. 3, pp. 360-375, 2018.

[168] M. Dahane, M. Sahnoun, B. Bettayeb, D. Baudry and H. Boudhar, “Impact of spare parts remanufacturing on the operation and maintenance performance of offshore wind turbines: a multi-agent approach,” *Journal of Intelligent Manufacturing,* vol. 28, pp. 1531-1549, 2015.

[169] B. Le and J. Andrews, “Modelling wind turbine degradation and maintenance,” *Wind Energy,* vol. 19, no. 4, pp. 571-591, 2015.

[170] S. Rigdon and A. Basu, “Statistical Methods for the Reliability of Repairable Systems,” John Wiley and Sons, New York, 2000.

[171] A. Zaher, S. D. J. McArthur, D. G. Infield and Y. Patel, “Online wind turbine fault detection through automated SCADA data analysis,” *Wind Energy,* vol. 12, no. 6, pp. 574-593, 2009.

[172] A. Encalada-Dávilla, C. Tutiven, B. Puruncajas and Y. Vidal, “Wind Turbine Multi-Fault Detection based on SCADA Data via an AutoEncoder,” *Renewable Energies & Power Quality Journal,* vol. 19, no. 4, 2021.

[173] I. M. Black, M. Richmond and A. Kolios, “Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management,” *International Journal of Sustainable Energy,* vol. 40, no. 10, pp. 923-946, 2021.

[174] A. Kusiak and A. Verma, “A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines,” *IEEE Transactions on Sustainable Energy,* vol. 2, no. 1, pp. 87-96, 2011.

[175] A. Lorenzo-Espejo, A. Escudero-Santana, M.-L. Muñoz-Díaz and A. Robles-Velasco, “Machine Learning-Based Analysis of a Wind Turbine Manufacturing Operation: A Case Study,” *Sustainability,* vol. 14, no. 13, p. 7779, 2022.

[176] S. Jeong, E.-J. Kim, D. H. Shin, J.-W. Park and S.-H. Sim, “Data fusion-based damage identification for a monopile offshore wind turbine structure using wireless smart sensors,” *Ocean Engineering,* vol. 195, p. 106728, 2020.

[177] C. Dao, B. Kazemtabrizi and C. Crabtree, “Wind turbine reliability data review and impacts on levelised cost of energy,” *Wind Energy,* vol. 22, no. 12, pp. 1848-1871, 2019.

[178] E. Hart, B. Clarke, G. Nicholas, A. K. Amiri, J. Stirling, J. Carroll, R. Dwyer-Joyce, A. McDonald and H. Long, “A review of wind turbine main bearings: design, operation, modelling, damage mechanisms and fault detection,” *Wind Energy Science,* vol. 5, no. 1, pp. 105-124, 2020.

[179] L. J. Bain and M. Engelhardt, Statistical analysis of reliability and life testing models: theory and methods, Dekker, 1991.

[180] S. Koukoura, “Wind turbine gearbox diagnostics using artificial intelligence,” University of Strathclyde, Glasgow, 2019.

[181] The Crown Estate, “Jack-up vessel optimisation - Improving offshore wind performance through better use of jack-up vessels in the operations and maintenance phase,” The Crown Estate, London, 2014.

[182] A. Kolios, J. Walgern, S. Koukoura, R. Pandit and J. Chiachio-Ruano, “openO&M: Robust O&M open access tool for improving operation and maintenance of offshore wind turbines,” *Proceedings of the 29th European Safety and Reliability Conference, ESREL 2019,* pp. 629-635, 2019.

[183] E. Hau, Wind Turbines: Fundamentals, Technologies, Application, Economics, 3 ed., Heidelberg: Springer Berlin, 2013.

[184] A. Bartschat, C. Broer, D. F. K. Coronado, J. Kucka, A. Mertens, R. Meyer, M. Moriße, K. Pelka, B. Tegtmeier, S. Weber and J. Wenske, “Zuverlässige Leistungselektronik für Windenergieanlagen. Abschlussbericht zum Fraunhofer-Innovationscluster Leistungselektronik für regenerative Energieversorgung,” Fraunhofer-Verlag, 2018.

[185] J. Liniger, M. Soltani, H. C. Pedersen, J. Carroll and N. Sepehri, “Reliability based design of fluid power pitch systems for wind turbines,” *Wind Energy,* no. 20, pp. 1097-1110, 2017.

[186] European Centre for Medium-Range Weather Forecasts ECMWF, “Climate reanalysis | Copernicus,” 2022. [Online]. Available: https://climate.copernicus.eu/climate-reanalysis. [Accessed 1 July 2022].

[187] K. Fischer, M. Steffes, K. Pelka, B. Tegtmeier and M. Dörenkämper, “Humidity in Power Converters of Wind Turbines - Field Conditions and Their Relation with Failures,” *Energies,* March 2021.

[188] O. S. Senturk, L. Helle, S. Munk-Nielsen, P. Rodriguez and R. Teodorescu, “Medium voltage three-level converters for the grid connection of a multi-MW wind turbine,” *13th European Conference on Power Electronics and Applications,* pp. 1-8, 2009.

[189] R. Grinberg, S. Ebner and O. Apeldoorn, "Reliability in Medium Voltage Converters for Wind Turbines," in *European Wind Energy Association (EWEA) Offshore Conference*, Amsterdam, 2011.

[190] M. Boettcher and F. W. Fuchs, "Power electronic converters in wind energy systems - considerations of reliability and strategies for increasing availability," *Proceedings of the 14th European Conference on Power Electronics and Applications,* pp. 1-10, 2011.

[191] K. Fischer, K. Pelka, A. Bartschat, B. Tegtmeier, D. Coronado and C. Broer, "Reliability of Power Converters in Wind Turbines: Exploratory Analysis of Failure and Operating Data From a Worldwide Turbine Fleet," *IEEE Transactions on Power Electronics,* vol. 34, no. 7, pp. 6332-6344, 2019.

[192] Copernicus, "Climate reanalysis," [Online]. Available: https://climate.copernicus.eu/climate-reanalysis. [Accessed 21 August 2023].

[193] K. Fischer, T. Stalin, H. Ramberg, J. Wenske, G. Wetter and R. Karlsson, "Field-Experience Based Root-Cause Analysis of Power-Converter Failure in Wind Turbines," *IEEE Transactions on Power Electronics,* vol. 30, no. 5, pp. 2481-2492, 2015.

[194] J. Carroll, A. McDonald, O. Barrera Martin, D. McMillan and R. Bakhshi, "Offshore wind turbine sub-assembly failure rates through time," *Conference Proceedings of EWEA 2015,* 2015.

[195] S. Ozturk, V. Fthenakis and S. Faulstich, "Assessing the factors impacting on the reliability of wind turbines via survival analysis - a case study," *Energies,* vol. 11, p. 3034, 2018.

[196] G. Wilson and D. McMillan, "Assessing wind farm reliability using weather dependent failure rates," *Journal of physics: conference series,* vol. 524, p. 012181, 2014.

[197] V. Slimacek and B. H. Lindqvist, "Reliability of wind turbines modeled by a Poisson process with covariates, unobserved heterogeneity and seasonality," *Wind Energy,* vol. 19, pp. 1991-2002, 2016.

[198] J. Walgern, K. Beckh, N. Hannes, M. Horn, M.-A. Lutz, K. Fischer and A. Kolios, "Impact of Using Text Classifiers for Standardising Maintenance Data of Wind Turbines on Reliability Calculations," *IET Renewable Power Generation,* vol. 18, no. 15, pp. 3463-3479, 2024.

[199] J. Walgern, N. Stratmann, M. Horn, N. Then Wei Ying, M. Menzel, F. Anderson, A. Kolios and K. Fischer, "Reliability and O&M key performance indicators of onshore and offshore wind turbines based on field-data analysis," *submitted to Wind Energy for publication,* 2025.

[200] D. Collett, Modelling survival data in medical research, Chapman and Hall/CRC, 2023.

[201] V. Rondeau, Y. Marzroui and J. R. Gonzalez, "frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation," *Journal of Statistical Software,* vol. 47, pp. 1-28, 2012.

[202] H. Wilker, Weibull-Statistik in der Praxis, 2nd Edition, 2010.

[203] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. García Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. Skidmore, D. Zurell and S. Lautenbach, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography,* vol. 36, pp. 027-046, 2013.

[204] O. Donelly and J. Carroll, "Daylight Considerations for Offshore Wind Operations and Maintenance," *Journal of Physics: Conference Series,* vol. 2875, p. 012018, 2024.

[205] M. Vieira and D. Djurdjanovic, "Insights on the Optimization of Short- and Long-Term Maintenance Decisions for Floating Offshore Wind Using Nested Genetic Algorithms," *Wind,* vol. 4, no. 3, pp. 227-250, 2024.

[206] K. Pelka and K. Fischer, "Modeling the Effect of Environmental and Operating Conditions on Power Converter Reliability in Wind Turbines with Time-Dependent Covariates," in *Proceedings of the 33rd European Safety and Reliability Conference*, Southampton, 2023.

[207] J. Walgern, F. Anderson, A. Kolios and K. Fischer, "Field-data based wind turbine reliability modelling: Quantifying effects of operating age, design and technological development," *Submitted to Wind Energy for publication,* 2025.

[208] M. L. Huhn and W. Popko, “Best practice for verification of wind turbine numerical models,” *Journal of Physics: Conference Series,* vol. 1618, p. 52026, 2020.

[209] J. Jonkman, S. Butterfield, W. Musial and G. Scott, “Definition of a 5-MW Reference Wind Turbine for Offshore System Development,” Technical Report NREL/TP-500-38060. National Renewable Energy Laboratory, Golden, Colorado, 2009.

[210] W. Popko, P. Thomas, A. Sevinc, M. Rosemeier, M. Bätge and R. Braun, “IWES Wind Turbine IWT-7.5-164. Rev 4,” Fraunhofer-Inistitue for Wind Energy Systems IWES, Bremerhaven, 2018.

[211] W. Popko, A. Aghaei, B. Schürmann, U. Smolka, J. Tautz-Weinert and U. Wihlfahrt, “SeaLOWT - Impact of Sea Ice Loads on Global Dynamics of Offshore Wind Turbines,” Fraunhofer Institute for Wind Energy Systems IWES, Bremerhaven, 2021.

[212] J. Walgern, D. Baumgärtner, J. Fricke, N. Requate, A. Kolios, M. Dörenkämper, T. Meyer and L. Vollmer, “Economic feasibility study for continued operation of German offshore wind farms,” *IOP Journal of Physics Conference Series,* 2023.

[213] K. Fischer, F. Anderson and J. Walgern, “Medium-Voltage versus Low-Voltage Converter Reliability in Wind Turbines: A Field-Data Based Study,” *PCIM Europe 2025,* 2025.