Wind Energy

WILEY

RESEARCH ARTICLE OPEN ACCESS

Field-Data-Based Wind Turbine Reliability Modelling: Quantifying Effects of Operating Age, Design and Technological Development

Julia Walgern[1,2] | Fraser Anderson[1,3] | Athanasios Kolios[2,4] | Katharina Fischer[1]

[1]Fraunhofer Institute for Wind Energy Systems IWES, Hannover, Germany | [2]University of Strathclyde, Glasgow, UK | [3]Fraunhofer UK Research Ltd., Glasgow, UK | [4]Department of Wind & Energy Systems, Technical University of Denmark, Roskilde, Denmark

Correspondence: Julia Walgern (julia.walgern@iwes.fraunhofer.de)

Received: 29 March 2025 | **Revised:** 10 December 2025 | **Accepted:** 23 February 2026

Keywords: failure rate | field data | maintenance reports | Nelson–Aalen plot | non-homogeneous Poisson process | reliability modelling | wind turbines

ABSTRACT

As wind energy continues to expand, ensuring the reliability of wind turbines is critical for optimising operational efficiency and minimising downtime. Based on maintenance data from over 1,000 onshore and offshore wind turbines covering more than 4200 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, whereas non-homogeneous Poisson process regression models are developed to quantify the effect of design and technological development, incorporating a range of covariates. Results reveal that whereas 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 generally 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.

1 | Introduction

In 2023, global wind energy capacity surpassed 1000 GW due to new onshore and offshore installations. Whereas 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 levelised cost of energy [2]. 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 (WTs) and their components. Existing research on WT reliability predominantly relies on outdated and limited datasets, often focusing on annual average failure rates (e.g., [3–9]).

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 [10].

Carroll et al. [11] and SPARTA [12] 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 various factors influencing reliability, including the works of Carroll et al. [13], Ozturk et al. [14], Reder and Melero [15, 16], Slimacek and Lindqvist [17], Wilson and McMillan [18] and Walgern et al. [19], which explore different methodological approaches and environmental or design influences on WT failure patterns.

This study advances the field of WT reliability by providing a comprehensive analysis based on a large and representative set of field data. Utilising maintenance reports covering over 4200 operational years from both onshore and offshore WTs, 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., [20, 21]) is employed in this study. This research extends beyond previous work to the application of reliability modelling 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 WTs, 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.

2 | Methodology and Evaluated Dataset

2.1 | Methodology

2.1.1 | Field-Data Collection and Preprocessing

Maintenance reports documenting each turbine visit for over 1000 WTs 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) [22] and the State-Event-Cause-Code 'ZEUS' [23] were employed for data preprocessing and classification. An overview of the RDS-PP defined subsystems which have been used for clustering WT components can be found in Table 1. Detailed information about the preprocessing approach is provided in Walgern et al. [24, 25]. 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 (indicated by the combination of ZEUS codes 02-08-01 and 02-09-09-01).

2.1.2 | Reliability Modelling

The most common method for modelling WT reliability involves calculating the failure rates of components,

TABLE 1 | Utilised subsystems according to VGB PowerTech [22].

RDS-PP code	Denomination	RDS-PP code	Denomination	RDS-PP code	Denomination
CKJ	Environmental measuring system	MST	Generator transformer system	UAB	Transformer station, in case of onshore turbines
MDA_rotor	Rotor system	MUD	Nacelle	XFB	Equipotential bonding/ earthing system
MDA_pitch	Pitch system	MYA	Remote monitoring system	XFC	Lightning protection system
MDK	Drive train system	UMD	Tower system	XAM	Ventilation systems
MDL	Yaw system	WBA	Personnel rescue systems	MDV	Central lubrication system
MDX	Central hydraulic system	XGM	Fire extinguishing system	MSS	Compensation system
MDY	Control system	CKA	Fire alarm system	MUR	Common cooling system
MKA	Power generation system	XMM	Lifting gears	Y	Telephone system
MSC	Generator switching system	XSD	Obstacle warning system	G	Other
MSE	Converter system	BFA	Low-voltage electrical main supply system		

subsystems and the overall turbine. To date, studies have primarily reported average failure rates per year and per WT (e.g., [3, 4, 6, 26]). Studies such as Spinato et al. [8], Koukoura [19], Walgern et al. [27], and Anderson et al. [21] have demonstrated a strong dependence of average failure rates on WT size. Therefore, we proposed in Walgern et al. [25] that failure rates should always be expressed per rated capacity in MW and per year, as presented in SPARTA [12] and Fischer et al. [28]. 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.

We consider failure behaviour through time within the context of a power law process, which we use to model the baseline failure intensity λ_0 with respect to time t [29]:

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

where $\delta > 0$ represents the shape parameter and $\upsilon > 0$ represents the scale parameter. The power law process can therefore be used to model distinct phases of reliability trends that form the characteristic shape of the bathtub curve [30, 31]:

- Early failures that are characterised by a decreasing failure rate ($\delta < 1$)
- Constant failures that are described by a constant failure rate ($\delta = 1$)
- Deterioration failures that are defined by an increasing failure rate ($\delta > 1$)

To identify these distinct phases, we employ a Nelson–Aalen estimator. The Nelson–Aalen estimator calculates the non-parametric cumulative failure intensity Λ_0 of a given field dataset.

$$\Delta_0 = \sum_{t_i \leq t} \frac{d_i}{n_i} \tag{2}$$

where 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. 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 1).

The gradient of a straight line in such a double-logarithmic Nelson–Aalen plot corresponds to the shape parameter δ. The kinks where contiguous lines of different gradients meet can be used to approximate the end of one failure stage and the beginning of another. We identify the position of these kinks by plotting the Nelson–Aalen estimator and reading the corresponding operating age(s) on the x-axis.

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 $\delta = 1$, 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:

$$\lambda(t) = z\,\lambda_0(t)\exp\left(\beta_1 x_1 + \ \dots\ + \beta_n x_n\right) \tag{3}$$

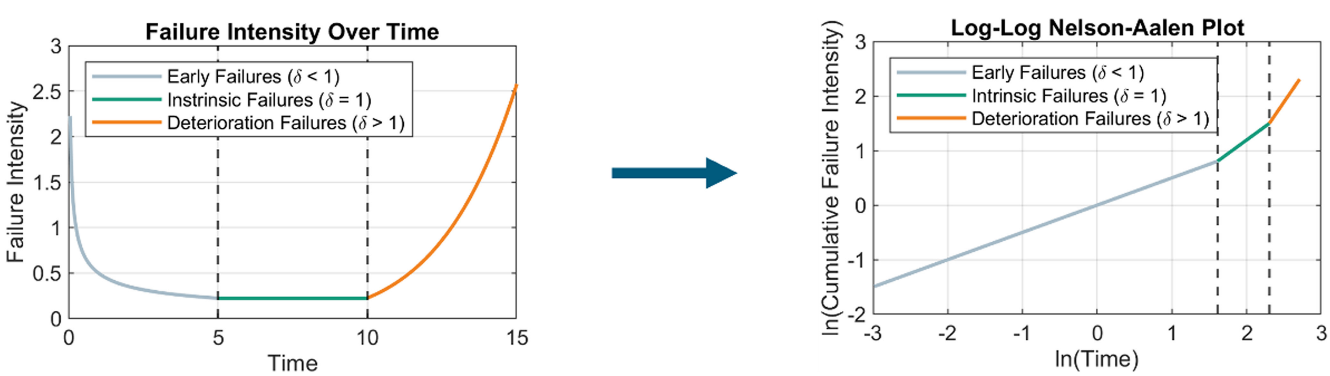

FIGURE 1 | Derivation of the shape parameter δ using Nelson–Aalen plots [32].

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 (PCA), which allows for the simultaneous inclusion of even highly correlated numerical covariates.
- A covariate selection procedure based on Collett [33], which enables the identification of covariates that have a significant effect on reliability. This is based on forward and backward selection: Each covariate is included in or excluded from the model in turn to see if it results in a better performing model (as indicated by a lower log-likelihood and a p-value < 0.05).
- 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, and we run through the covariate selection procedure for each subsample. This approach produces 'inclusion rate' plots, which indicate how often a given covariate is included in the model of a particular subsystem by the covariate selection procedure.

A detailed presentation of the methodology, using the power converter subsystem as an example, can be found in [21]. Results were obtained using Matlab Version R2023b and R Version 4.3.2. The R package frailtypack [34] was used to fit the NHPP regression models.

2.2 | Evaluated Dataset

The dataset used in this analysis is derived from maintenance reports of both onshore and offshore WTs, encompassing 1089 WTs with a total of over 4200 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 subsystems detailed in Table 1 according to VGB PowerTech [22].

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 and SCIG) across low and medium voltage levels, along with partially and fully rated converters. An overview of the essential dataset characteristics is provided in Table 2. In addition to the onshore/offshore distribution, the proportion of different subsystem technology concepts based on number of covered WT operating years are presented, underscoring the dataset's diversity and representativeness.

TABLE 2 | Overview of the essential dataset characteristics with split proportion based on number of covered WT operating years.

	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	
Split pitch system concepts	Electrical: 44%; hydraulic: 56%	
Split converter system concepts	Fully rated: 48%; partially rated: 52%	
Split drive train concepts	Geared: 79%; direct drive: 21%	

3 | Results and Discussion

In the reliability analysis presented in [25], 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 chose the rotor system, power generation system and yaw system for deeper, methodology-wise, more advanced analysis in the present paper as these subsystems are key for the power conversion process.

Although Walgern et al. [25] focus on reliability key performance indicators including average failure rates, repair times and required number of technicians, 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).

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

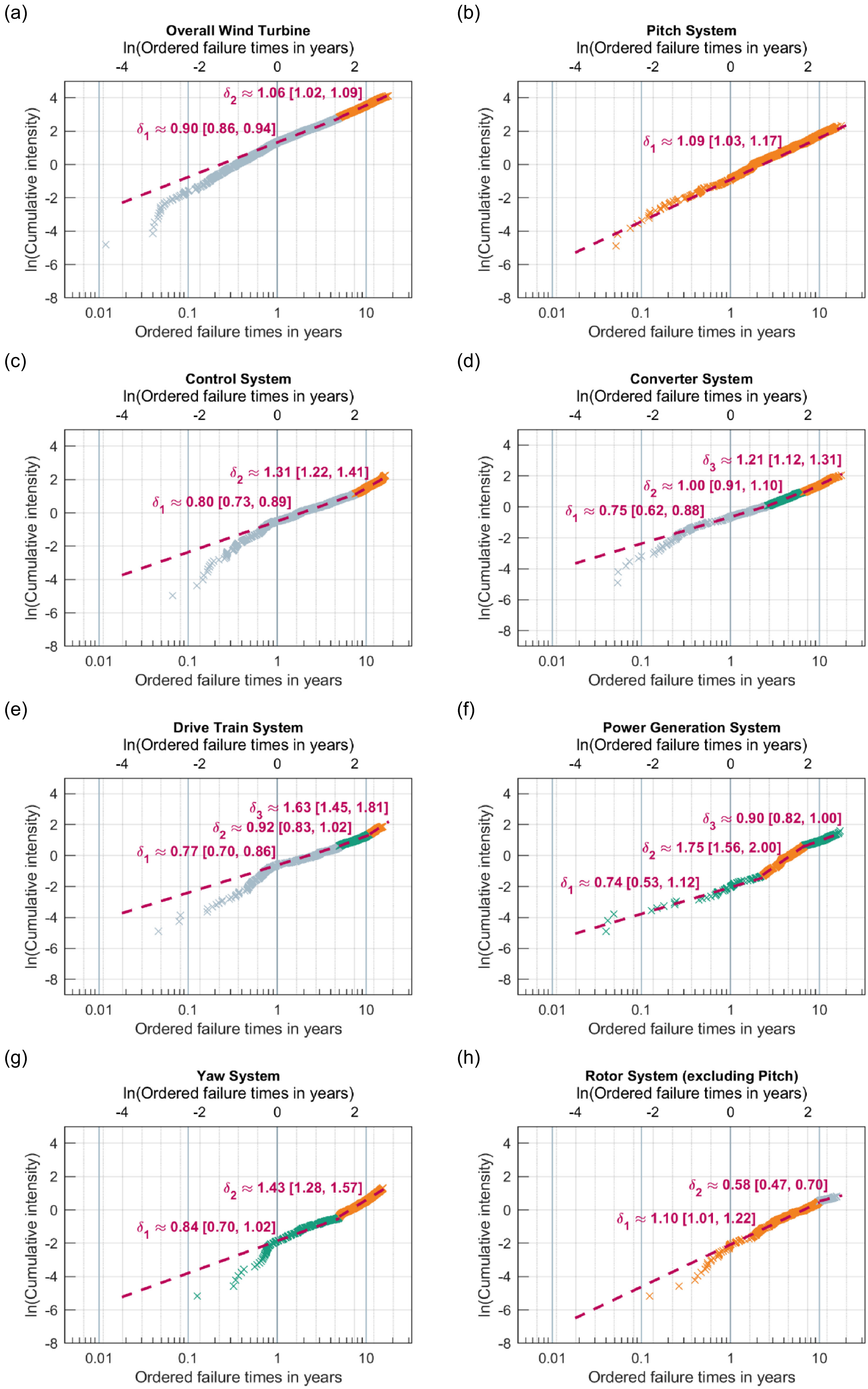

FIGURE 2 | Legend on next page.

FIGURE 2 | 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.

exhibit all three phases of the bathtub curve (cf. [21, 32]). As explained in Wilker [35], early failures are typically related to material or manufacturing defects, insufficient testing or inadequate mounting. Intrinsic failures are 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 ageing or wear-out, that is, 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 2 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 the case of the converter subsystem, early failure behaviour transitions to intrinsic failures after ~3 years and further to deterioration at a turbine age of ~6 years. In the case of the drive train subsystem, the early failure trend characterised by decreasing failure intensity is 5 years long, and the transition to deterioration is observed as late as around a WT operating age of 11 years. In contrast, the overall WT system and the control subsystem transition directly from early failures to deterioration failures, after 5 and 8 years, respectively, 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 at a WT operating age of ~5 years. The failure behaviour of the pitch and rotor subsystems is predominantly driven by deterioration. The trend of improving reliability observed from operation year 11 onwards 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 as early as after ~2 years and then reverting to a stable failure intensity at a WT operating age of 7 years. This could be associated with a learning curve as in the 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 Carroll et al. [11] and SPARTA [12] 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, SPARTA [12] 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.

Carroll et al. [11] 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 8 years of WT operation, allowing a comparison only with the first part of the long-term failure behaviour shown in Figure 2. 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.

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. Whereas highly correlated numerical covariates can be addressed through 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 Dormann et al. [36]—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 3, 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 Anderson et al. [21].

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 (2). Where a logarithmic inclusion of rated power improves the model fit, it implies that a root/linear/power function relationship is more accurate than the exponential relationship.

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 [21, 28], the rated capacity of the WT instead of the converter is utilised for the covariate 'rated capacity' in the present work.

TABLE 3 | 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

- 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 Dormann et al. [36], 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.

Where covariates are categorical, a reference and factor level must be defined. A reference level is a specific category that is used as a baseline for comparison in the regression analysis. All other categories (referred to as 'factor levels') are compared to this baseline. Table 4 provides an overview of the covariates, their types, value ranges and the subsystem models for which each covariate has been considered.

Figure 3 presents inclusion rate plots illustrating the outcomes of the covariate selection procedure for both the overall WT 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 'electrical/hydraulic' was considered exclusively in the analysis of the pitch subsystem, whereas 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 3.

TABLE 4 | Overview of covariates included in the analysis.

Covariate	**Included in analysis of subsystems**	**Type of covariate**	**Range of values (num.) or factor levels (categ.)**
Onshore/offshore	All subsystems	Categorical	Offshore, onshore
WT commissioning year	All subsystems	Numerical	1997–2019
WT rated capacity	All subsystems	Numerical	<1–9 MW
Electrical/hydraulic pitch system	Pitch subsystem	Categorical	Electrical, hydraulic
Fully/partially rated converter	Converter subsystem	Categorical	Fully rated, partially rated
Drive train concept	Drive train subsystem, power generation subsystem	Categorical	Geared, direct drive

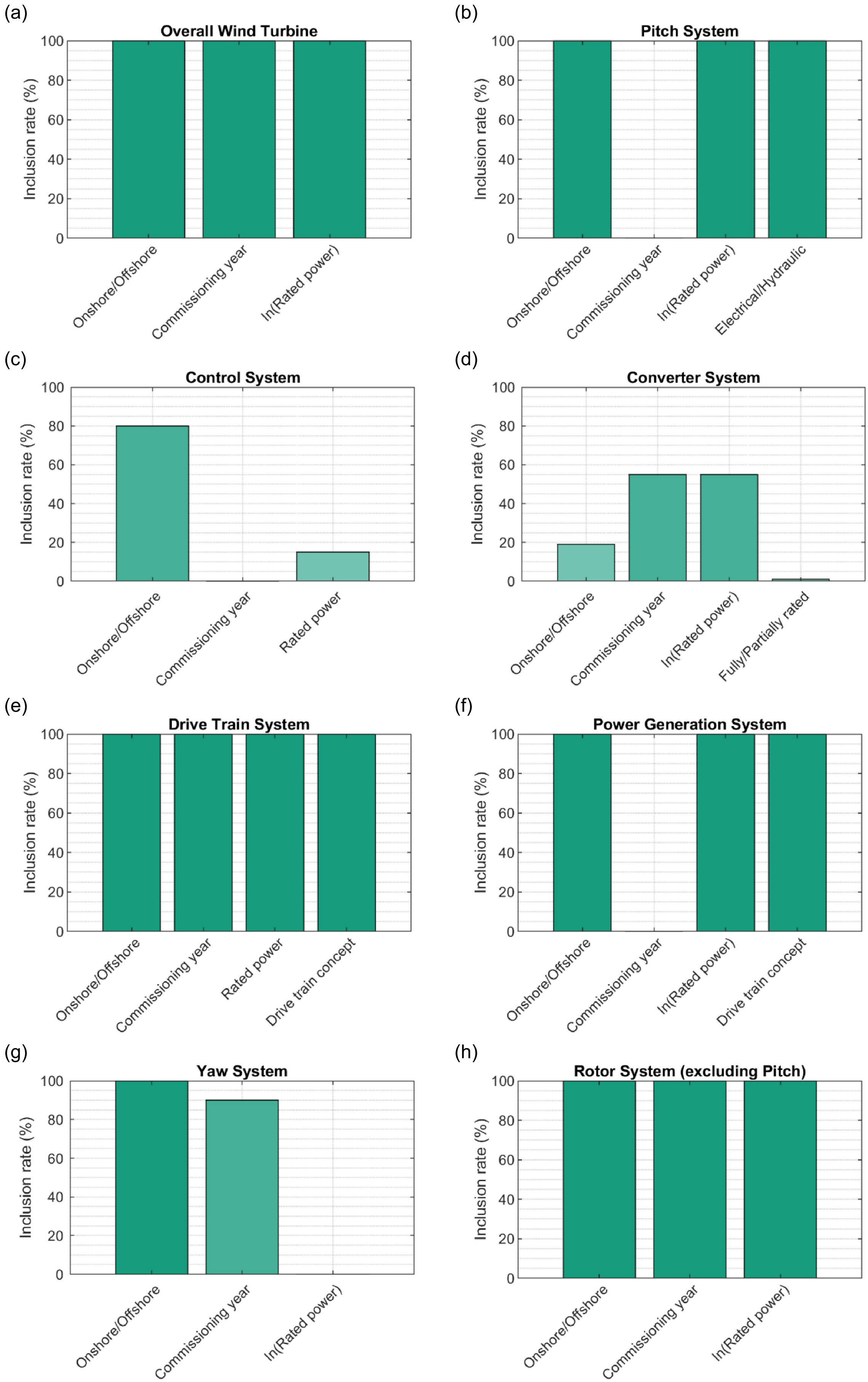

FIGURE 3 | Inclusion rate plots for the entire wind turbine system (a) and individual subsystems (b–h).

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., [21]). 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

TABLE 5 | 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)
Electrical/hydraulic	Electrical	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)

imply a lack of influence. As a result, only covariates identified as having a significant effect should be interpreted. Table 5 summarises the results of the final reliability models, presenting the estimated β coefficients and their respective confidence intervals for each covariate across the overall WT system and individual subsystems.

The β-coefficients shown in Table 5 are a result of regressing on the entire respective field dataset, with the set of covariates determined by the inclusion rate analysis. Where applicable, Table 5 also shows exp(β). This can be interpreted more directly as the relative effect of the covariate on the subsystem's failure intensity. It tells you by what quantity the failure intensity changes when the covariate increases by one unit. An $\exp(\beta) = 2$, for example, tells us that a change of the covariate in one unit will result in a doubling of the failure intensity.

For the overall WT 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. This implies that the drive trains specifically of larger turbines in our dataset are far less reliable than the rest.

Additionally, the estimated β coefficients associated with the selected covariates (see Figure 3 and Table 5) provide insights into both the direction and magnitude of their effect on overall WT and subsystem reliability, enabling a quantitative assessment of their influence.

For both the overall WT system and most of the subsystems where the covariate 'onshore/offshore' was found to have a significant effect on reliability, a beta coefficient $\beta > 0$ indicates that offshore turbines exhibited higher reliability compared to their onshore counterparts. This finding aligns with the results of Walgern et al. [25], 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 negative β coefficient. Whereas Walgern et al. [25] normalised failure rates by rated power assuming a general linear 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 WT system, identified exp(β) values in the range of 0.949–0.986, indicating a reduction in failure intensity by these factors per year of later commissioning show that 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 considerably in more recent turbine generations as indicated by $\exp(\beta) = 1.097$ (implying almost 10% higher failure intensity per year of later commissioning). This may be attributed to the industry's shift towards slimmer and more flexible blade designs, which are optimised for larger turbines but may introduce new reliability challenges affecting operational expenditure.

Whereas for the overall WT and most subsystems, 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 WT 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., [21]), the pitch subsystem (cf. [19, 37]), the drive train subsystem [27] and the overall WT system (e.g., [8, 12]).

Whereas Walgern et al. [19] concluded, based on annual average failure rates per WT, that electrical and hydraulic pitch systems exhibit similar reliability, the present study demonstrates that turbines equipped with hydraulic pitch systems exhibit considerably higher reliability than comparable ones with electrical pitch systems. Note that Walgern et al. [19] covered 1847 operational years and turbines of up to 6 MW, whereas the current dataset comprises more than 4200 operational years and includes turbines with rated capacities of up to 9 MW. With this significant increase in scale, and the more advanced methodology used in this study, the latter result is more reliable. This difference is indicated by the relative effect $\exp(\beta) = 0.646$ between the failure intensity in the reference case with electrical pitch system and the failure intensity in the case of the hydraulic pitch system.

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 WTs with and without gearboxes. Given that direct-drive turbines inherently have fewer components within this subsystem category defined by RDS-PP [22], it is unsurprising that they exhibit a much lower failure intensity compared to geared turbines, as indicated by a relative effect of $\exp(\beta) = 0.321$. Nevertheless, this finding aligns with SPARTA [12], which reported that direct-drive turbines experience fewer average monthly forced outages per MW and lower associated production losses.

Also for the power generation system, direct-drive turbines have been found to exhibit higher reliability. With $\exp(\beta) = 0.623$, the relative effect is clear but not as pronounced as in the case of the drive train subsystem. Although the superior reliability of the power generation system in direct-drive turbines may initially seem counterintuitive given that direct-drive concepts 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 slip-ring unit, which is subject to wear-out and as such a typical driver for maintenance interventions. Our result aligns with the findings of Carroll et al. [13], who reported that PMSG, including their auxiliary components such as cooling and lubrication systems, failed less frequently than DFIG-based configurations in turbines of identical capacity.

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. Although O&M simulations commonly rely on average failure rates per turbine (e.g., [38–40]), we suggest in Walgern et al. [25] 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 2 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 5 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, whereas 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 5).

Figure 4 illustrates the isolated effect of WT rated capacity on subsystem reliability for those subsystems where rated capacity has a significant effect. These simplified curves are derived as follows. Firstly, we assume all other covariates to be at a constant value. We take the β coefficients relating to each subsystem from Table 5. We want to calculate the ratio of the failure intensity (λ) at some arbitrary value of rated power (x_{RP}) to some baseline failure intensity (λ_{1MW}), which we define to be at a rated power of 1 MW. When there is a log transform of rated power (as is the case for the power generation, rotor, pitch and converter systems), this looks like

$$\frac{\lambda(x_{RP},t)}{\lambda_{1MW}(t)} = \exp\left(\beta_{RP}\log(x_{RP})\right) = x_{RP}^{\beta_{RP}} \tag{4}$$

When there is a linear inclusion of the rated power covariate (as is the case for the drive train system), the resulting equation is

$$\frac{\lambda(x_{RP},t)}{\lambda_{1MW}(t)} = \frac{\exp\left(\beta_{RP}x_{RP}\right)}{\exp\left(\beta_{RP}\right)} = \left[\exp\left(\beta_{RP}\right)\right]^{x_{RP}-1} \tag{5}$$

The generic scaling per WT assumes a 1–1 relationship between λ and λ_{1MW} and rated power, that is, a 5-MW turbine fails five times as often as a 1-MW turbine.

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 the case of the pitch and the converter subsystem, is close to linear for the rotor subsystem

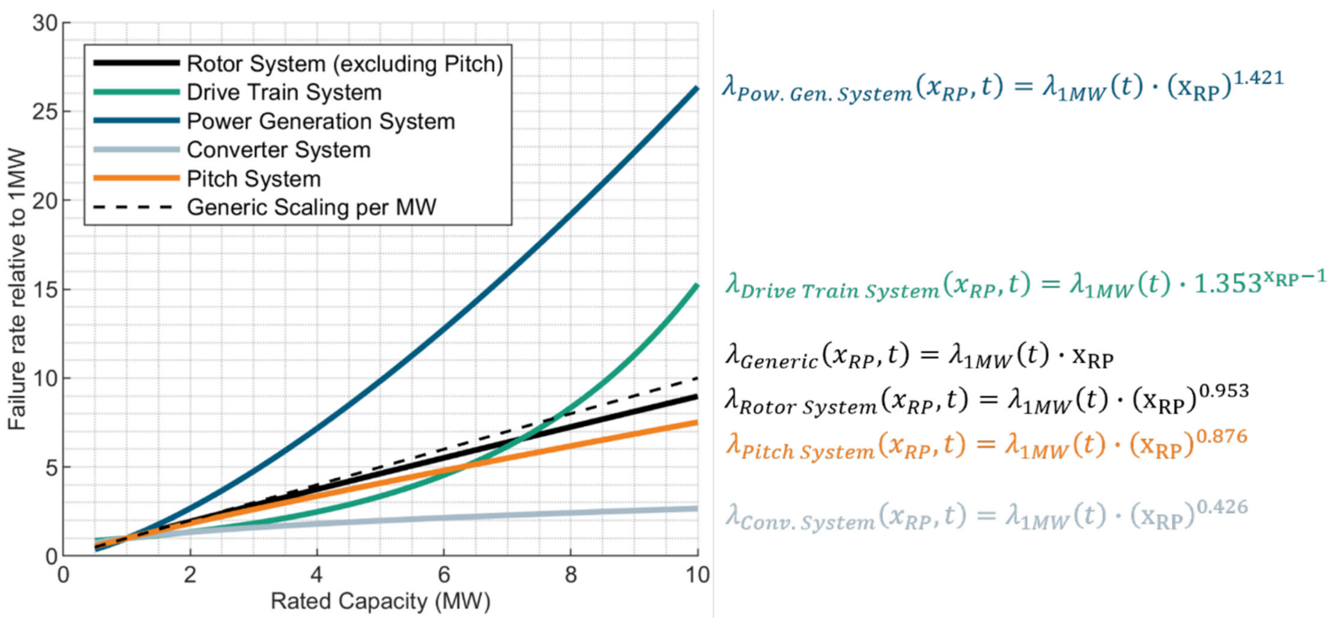

FIGURE 4 | Impact of rated capacity on wind turbine subsystem reliability.

and is represented by a power function for the power generation subsystem. The black dashed line in Figure 4 represents the outcome of generic scaling per MW, which assumes that a 5-MW turbine fails five times as often as a 1-MW turbine. 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 WT by a factor of $2^{0.592} \approx 1.51$ (cf. β values in Table 5), 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 4 solely depicts the effect of rated capacity, although 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 preferable to assuming constant failure rates across different turbine sizes, even though rated capacity does not significantly affect all subsystems.

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.

4 | Conclusions and Outlook

This study has provided a comprehensive analysis of WT reliability on a subsystem level, based on failure data from over 1000 WTs and more than 4200 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 NHPP in combination with a covariate selection procedure.

The results highlight distinct reliability trends over WT operating age across different subsystems, demonstrating that whereas some subsystems exhibit a classical bathtub curve, others transition directly from early failures to deterioration or show even solely a deterioration trend. 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 also been 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. [19]). 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. Although 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, for example, of the drive train system, into individual subassemblies or components enables more detailed reliability modelling and remains the 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 Pelka and Fischer [41] and Anderson et al. [32], both of which focus on the converter system. Pelka and Fischer [41] improved upon the assumption of constant covariates by introducing time-dependent covariates, revealing a previously unobserved dependence on electrical utilisation. Anderson et al. [32] 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.

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 the Engineering and Physical Sciences Research Council (EPSRC) through the Wind and Marine Energy Systems Centre for Doctoral Training under the grant number EP/S023801/1. Open Access funding enabled and organized by Projekt DEAL.

Funding

This work was supported by the German Federal Ministry for Economic Affairs and Climate Action (03EE3106) and the Engineering and Physical Sciences Research Council (EP/S023801/1).

Data Availability Statement

Research data are not shared.

References

1. Global Wind Energy Council, "GWEC | Global Wind Report 2024," Brussels: Global Wind Energy Council, (2024).

2. BVG Associates, "Guide to an Offshore Wind Farm," (2019).

3. J. Carroll, A. McDonald, and D. McMillan, "Failure Rate, Repair Time and Unscheduled O&M Cost Analysis of Offshore Wind Turbines," *Wind Energy* 19 (2016): 1107–1119.

4. J. B. Gayo, "Final Publishable Summary of Results of Project ReliaWind," (2011).

5. B. Hahn, M. Durstewitz, and K. Rohrig, "Reliability of Wind Turbines," in *Wind Energy*, ed. J. Peinke, P. Schaumann, and S. Barth (Springer, 2007), 329–332, https://doi.org/10.1007/978-3-540-33866-6_62.

6. 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)* 753, no. 7 (2016): 072027.

7. SPARTA, "Portfolio Review 2016: System Performance, Availability and Reliability Trend Analysis," SPARTA Project, (2017).

8. F. Spinato, P. Tavner, G. ban Bussel, and E. Koutoulakos, "Reliability of Wind Turbine Subassemblies," *IET Renewable Power Generation* 3, no. 4 (2009): 387–401.

9. P. Tavner, J. Xiang, and F. Spinato, "Reliability Analysis for Wind Turbines," *Wind Energy* 10 (2007): 1–18.

10. D. Cevasco, S. Koukoura, and A. J. Kolios, "Reliability, Availability, Maintainability Data Review for the Identification of Trends in Offshore Wind Energy Applications," *Renewable and Sustainable Energy Reviews* 136 (2021): 110414, https://doi.org/10.1016/j.rser.2020.110414.

11. J. Carroll, "Offshore Wind Turbine Sub-Assembly Failure Rates Through Time," Conference Proceedings of EWEA 2015, (2015).

12. SPARTA, "Portfolio Review 2020/21: System Performance, Availability and Reliability Trend Analysis," ORE Catapult, (2022).

13. J. Carroll, A. McDonald, and D. McMillan, "Reliability Comparison of Wind Turbines With DFIG and PMG Drive Trains," *IEEE Transactions on Energy Conversion* 30, no. 2 (2015): 663–670.

14. S. Ozturk, V. Fthenakis, and S. Faulstich, "Assessing the Factors Impacting on the Reliability of Wind Turbines via Survival Analysis – A Case Study," *Energies* 11 (2018): 3034.

15. M. Reder and J. J. Melero, "Modelling Wind Turbine Failures Based on Weather Conditions," *Journal of Physics: Conference Series* 926 (2017): 012012.

16. M. Reder and J. J. Melero, "Modelling the Effects of Environmental Conditions on Wind Turbine Failures," *Wind Energy* 21 (2018): 876–891.

17. V. Slimacek and B. H. Lindqvist, "Reliability of Wind Turbines Modeled by a Poisson Process With Covariates, Unobserved Heterogeneity and Seasonality," *Wind Energy* 19 (2016): 1991–2002.

18. G. Wilson and D. McMillan, "Assessing Wind Farm Reliability Using Weather Dependent Failure Rates," *Journal of Physics: Conference Series* 524 (2014): 012181.

19. 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* 9 (2023): 3273–3281.

20. K. Pelka and K. Fischer, "Field-Data-Based Reliability Analysis of Power Converters in Wind Turbines: Assessing the Effect of Explanatory Variables," *Wind Energy* 26, no. 3 (2023a): 310–324.

21. 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* 40 (2025a): 7286–7297.

22. VGB PowerTech, "VGB-Standard RDS-PP Application Guideline Part 32: Wind Power Plants," (2014).

23. FGW e.V. – Fördergesellschaft Windenergie und andere Erneuerbare Energien, "Technical Guidelines for Power Generating Units – State-Event-Cause Code for Power Generating Units (ZEUS)," FGW e.V. – Fördergesellschaft Windenergie und andere Erneuerbare Energien, (2013).

24. J. Walgern, K. Beckh, N. Hannes, et al., "Impact of Using Text Classifiers for Standardising Maintenance Data of Wind Turbines on Reliability Calculations," *IET Renewable Power Generation* 18, no. 15 (2024): 3463–3479.

25. J. Walgern, N. Stratmann, M. Horn, et al., "Reliability and O&M Key Performance Indicators of Onshore and Offshore Wind Turbines Based on Field-Data Analysis," preprint, *Wind Energy Science Discussion*, (2025), https://doi.org/10.5194/wes-2025-212.

26. 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* 2626, no. 1 (2023): 012025.

27. S. Koukoura, "Wind Turbine Gearbox Diagnostics Using Artificial Intelligence," University of Strathclyde, 2019.

28. K. Fischer, K. Pelka, and J. Walgern, "Trends and Influencing Factors in Power-Converter Reliability of Wind Turbines," in *Konferenz: PCIM Europe 2023 - International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management 09.05.2023-11.05.2023 in Nürnberg, Germany* (VDE, 2023), 1–10, https://doi.org/10.30420/566091068.

29. H. E. Ascher, "Repairable Systems Reliability," in *Encyclopedia of Statistics in Quality and Reliability* (Wiley, 2008).

30. G. Pulcini, "Modeling the Failure Data of a Repairable Equipment With Bathtub Type Failure Intensity," *Reliability Engineering & System Safety* 71, no. 2 (2001): 209–218.

31. S. Rigdon and A. Basu, *Statistical Methods for the Reliability of Repairable Systems* (John Wiley and Sons, 2000).

32. F. Anderson, J. Walgern, and K. Fischer, "Early and Deterioration Stage Wind Turbine Reliability Models: A Case Study for the Converter System," EERA DeepWind Conference, (2025b).

33. D. Collett, *Modelling Survival Data in Medical Research* (Chapman and Hall/CRC, 2023).

34. 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* 47 (2012): 1–28.

35. H. Wilker, *Weibull-Statistik in der Praxis*, 2nd ed. (Books on Demand, 2010).

36. C. F. Dormann, J. Elith, S. Bacher, et al., "Collinearity: A Review of Methods to Deal With It and a Simulation Study Evaluating Their Performance," *Ecography* 36 (2013): 27–46.

37. P. Padman, F. Vanni, E. Echavarria, K. Mortstock, and M. Wilkinson, "Benchmarking Pitch System Reliability and Reducing Cost of Energy Through Advanced Design," EWEA-Workshop "Analysis of Operating Wind Farms", (2016).

38. O. Donelly and J. Carroll, "Daylight Considerations for Offshore Wind Operations and Maintenance," *Journal of Physics: Conference Series* 2875 (2024): 012018.

39. O. Donnelly, J. Carroll, and M. Howland, "Analysing the Cost Impact of Failure Rates for the Next Generation of Offshore Wind Turbines," *Wind Energy* 27, no. 7 (2024): 695–710.

40. 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* 4, no. 3 (2024): 227–250.

41. K. Pelka and K. Fischer, "Modeling the Effect of Environmental and Operating Conditions on Power Converter Reliability in Wind Turbines With Time-Dependent Covariates," (2023b).