Received: 15 September 2022 | Revised: 21 March 2023 | Accepted: 4 June 2023

DOI: 10.1002/we.2846

RESEARCH ARTICLE

WILEY

A Bayesian reliability analysis exploring the effect of scheduled maintenance on wind turbine time to failure

Fraser Anderson[1] | **Rafael Dawid**[2] | **David McMillan**[2] | **David García-Cava**[1]

[1]Institute for Energy & Infrastructure, University of Edinburgh, Edinburgh, UK

[2]Electronic and Electrical Engineering, Strathclyde University, Glasgow, UK

Correspondence
Fraser Anderson, Institute for Energy & Infrastructure, University of Edinburgh, Edinburgh, UK.
Email: F.J.Anderson@sms.ed.ac.uk

Funding information
UK's Engineering and Physical Sciences Research Council via the Wind & Marine Energy Systems Centre for Doctoral Training, Grant/Award Number: EP/L016680/1

Abstract

This article presents a Bayesian reliability modelling approach for wind turbines that incorporates the effect of time-dependent variables. Namely, the technique is used to explore the effect of annual services on wind turbine failure intensity through time for turbines within a currently operational wind farm. In the operator's experience, turbines seemed to fail more frequently after scheduled maintenance was performed; however, this is an unexplored effect in the literature. Additionally, the effects of seasonality, year of operation and position in the array on failure intensity are explored. These features were included in a Cox-like model formulation which allows for time-dependent covariates. Inference was performed via Bayes rule. Results show a spike in failure intensity reaching 1.57 times the baseline in the six days directly proceeding annual servicing, after which failure intensity is reduced compared to baseline. Also observed is a significant year-on-year reduction of failure intensity since the introduction of the site's data management system in 2018, a clear preference for modelling time to failure via a Weibull distribution and a dependence on location in the array with respect to the prominent wind direction. Results also show the benefit of employing a Bayesian regime, which provides easily interpretable uncertainty quantification.

KEYWORDS

annual services, O&M, offshore wind, reliability modelling

1 | INTRODUCTION

For the purposes of lowering operations and maintenance (O&M) costs for offshore wind farms (OWFs), many operational decision making tools have been created with OWF operators in mind.[1] Central to many of these tools is failure modelling. However, failure modelling is often hampered by issues with consistent and reliable data collection. Reder et al[2] and International Energy Agency (IEA) Wind Task 33[3] explore this issue, in particular highlighting that failure data are generally hard to come by and can be of a poor quality when they are available. Such issues may be addressed by more advanced data management systems at OWFs. In turn, more advanced data management systems might lead to better failure modelling and better operational efficiency.

One of the potential benefits of the industry incorporating better data processing practices is that it allows researchers to draw inferences between relevant covariates and wind turbine (WT) failure intensity. Information about what factors are detrimental to turbine performance is valuable to wind farm stakeholders as it allows the potential for a more effective operational strategy and more accurate project financing. This is especially true for offshore sites, which in responding to WT failures contend with weather constrictions on safe access and expensive vessel charters. Particularly, uncertainties in reliability or maintainability estimates are becoming a key focus of investigation for O&M researchers, as shown by recent works by Seyr and Muskulus,[4] Scheu et al[5] and Dao et al.[6]

This study is motivated by WT reliability analysis. An extensive dataset, provided by a currently operational wind farm, acts as the basis. Insights through conversations with the operator act as another motivator. Particularly, a phenomenon was observed by the operational team where more failures were occurring directly after scheduled maintenance works, which is a hitherto unquantified effect for WTs. Regarding methodology, we focus on the frequency of corrective maintenance actions via extensions of the traditional methods, namely, Poisson processes and Weibull processes.

We extend the typical Poisson process-type approach by two means. First, we propose a transition of traditional reliability analysis methods into the Bayesian regime. This has the advantages characteristic of all Bayesian methodologies, namely, their suitability for small and incomplete datasets,[7] their explicit treatment of uncertainty and support for decision analysis[8] and the ability to combine different sources of knowledge or datasets.[9] Second, we propose a data-modelling approach which has been employed by numerous recurrent time-to-event data analyses, capable of capturing the effect of non-corrective works on the likelihood of proceeding failures. This was inspired by anecdotal evidence of increased turbine failures after annual services from the operator of a large OWF. In this study we establish and quantify this relationship, which is a novel consideration for wind farm O&M modellers. Time since annual service is therefore considered as an impacting covariate which has not been explored by previous studies. The effects of seasonality, year of operation and position in the array are also considered as covariates in the model.

The rest of the paper is structured as follows. Section 2 provides a review of reliability modelling for WTs, covering the definition and measurement of relevant key performance indicators (KPIs), reliability modelling practices commonly employed by operational decision making tools and factors which have been considered in the literature. Section 3 is an outline of the methodology employed, covering data preprocessing, the definition of the Bayesian reliability model and the model selection procedure. Section 4 presents the results of the model selection criteria and covariate values. Section 5 presents a discussion of the results and their implications. Section 6 is a conclusion.

2 | RELIABILITY MODELLING OF WTS

In the context of wind energy, reliability broadly describes the ability of the WT to produce power over its design life and within its wind limits. Addressing this concept via the analysis of operational data usually involves the calculation of certain KPIs. As summarized by Gonzalez et al,[10] the wind industry is chiefly concerned with two questions concering WT reliability: 'How often does a turbine fail?' and 'Which WT downtimes are associated with which failure?'. They go on to describe four common reliability KPIs:

1. Mean time between failures (MTBF) and failure rate
2. Mean time to repair (MTTR) and repair rate
3. Mean time to failure (MTTF)
4. Availability

Of these, MTBF and failure rate are the indicators of interest for this study. While many reliability analyses focus on a component breakdown of turbine failures, we focus on the WT system as a whole. In this respect, we are restricted by the limitations of the dataset available, which does not readily allow for breakdown of failures into a subcomponent taxonomy. This distinction is important, as not all failures are equal in terms of resources required to repair them and downtime associated with that repair. Component replacements that require a jack-up vessel—most notably the gearbox, main bearing, blades and generator—are very costly. Information on them is therefore valued most by decision makers. For a more thorough discussion of WT component failures and their consequences, see Artigao et al.[11]

This study relies on the assumption that a third question is of interest to researchers, namely, 'Under what operational conditions does a turbine fail?'. This is fundamentally what this study to answer by introducing novel covariates to reliability models.

2.1 | Failures, failure rates and MTBF

There is no standard definition of a WT failure in the wind industry. Here, we are consistent with the definition used by the current authors in Anderson et al[12]: A failure is defined by a downtime event accompanied by a recorded unscheduled visit to a turbine (a visit signifying the transfer of a technician crew from vessel to turbine). This is the same definition as Kaidis et al.[13] Some studies have additional criteria, for example, Carroll et al,[14] who define a failure as 'a visit to a turbine, outside of a scheduled operation, in which material is consumed'. Some studies impose a limit on downtime, (e.g.) only downtimes lasting longer than an hour are counted as failures.[13] Other studies rely more heavily on alarm logs for their definitions and include remote resets,[2] in which case failure rates tend to be higher.

Since this study is restricted to turbine-level failures, we impose two additional conditions to our failure definition:

1. A lower limit on downtime of 1 h for the event to be considered a failure.
2. Downtime events with recorded unscheduled visits to the same turbine occurring within 24 h of each other are assumed to be the same failure.

WT failure rate describes the number of failures for a given fleet of turbines per unit time, usually presented as number of failures per turbine per year.[15] Failure rate is a particularly critical KPI for wind farm operators to consider as it is a key O&M cost driver. For onshore sites, numerous reliability analyses have been published—see Pfaffel et al.'s review of WT performance and reliability for details.[16] For offshore sites published reliability figures are sparse. Table 1 summarizes those average failure rate estimates for offshore WTs that are available in the literature. For this particular wind farm and failure definition, the failure rate was calculated to be 7.31 failures per year. Figure 1A shows how these failures are distributed. Notably, this study has a lower failure rate than those previously recorded for turbines with similar power ratings. In the case of SPARTA, it may be that their definition of 'repair rate' differs significantly than the proposed 'failure rate' of this study; however, since it is not defined in the publication, it is difficult to comment on.

The focus of this study is in modelling time between failures rather than failure rates. This represents something of a shift away from the counting processes usually employed in WT reliability modelling to time-to-event data, which is akin to the common reliability metric MTBF or just time between failures. Figure 1B shows a histogram plot of time between failures for the dataset utilized in this study, with the MTBF shown by the dotted line. The metric is evidently exponentially distributed, justifying the use of the exponential family of models to represent it.

2.2 | Reliability models, Poisson and Weibull processes

A central part to many operational decision making tools is the representation of WT failure, more formally known as failure modelling. It is vital to capture the frequency and probability of turbine failures accurately, as strategy for responding to them will dictate maintenance costs. A popular option for doing so is via so-called classical reliability analysis, which relies on historical failure data to predict time to failure. Another option is via structural reliability models, which are comprehensively reviewed by Jiang et al.[18] In contrast to classical reliability models, the structured option aims to derive probability of failures for specific components based on the load characteristics and material properties (among other variables) of the component. In this study, the focus is on operational efficiency improvement through analysis of operational data. It is restricted to

TABLE 1 Summary of published offshore WT failure rates.

Study	Failure rate	Power rating (MW)	No. of turbines
Anderson et al.	7.31	2–4	N/A
OWEZ[15]	7.82	3	36
Carroll et al.[14]	8.27	2-4	350
SPARTA[17a]	15.84	N/A	1045

[a]Repair rate rather than a failure rate.

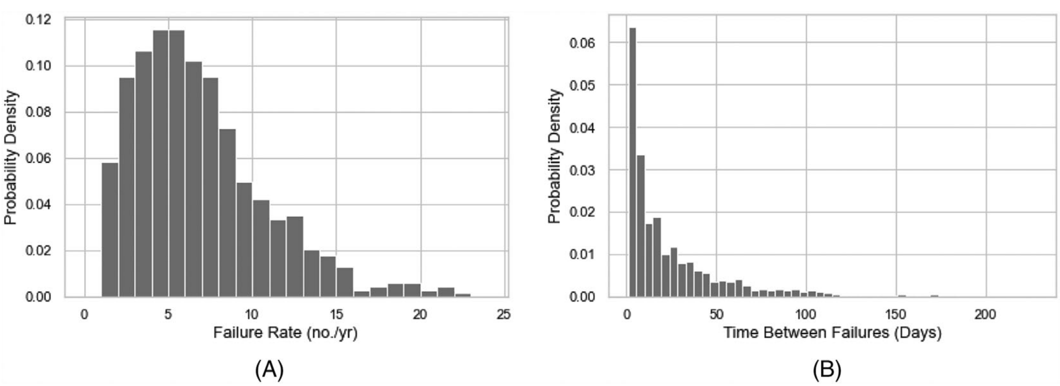

FIGURE 1 Histograms for (A) WT failure rates and (B) WT time between failures for the period covering July 2018–July 2021. Mean values for both metrics shown by dashed lines.

turbine-level failures. For those two reasons, we deem a classical reliability analysis approach more suitable. If the aim were to inform the design standards of different components (most notably the blades), structural reliability analysis should be considered.

Many studies have in common their reliance on historic failure data to model failures of WTs via classical reliability analysis. However, there is plenty of licence for different interpretations of that data. In simple cases, failures are modelled deterministically.[1] Turbines are assumed to fail after a constant period of time, taken as the MTBF. While this is broadly indicative of failure behaviour, a recent study by Scheu et al[5] has shown the importance of considering the uncertainty of failure rate around a mean value. It is therefore advisable to adopt a probabilistic method, where failures are assumed to occur at random intervals as described by some probability distribution function (PDF), which can be fit to the data. Most popular of these is the Poisson process, of which there are two variants: the homogeneous and non-homogeneous Poisson process (HPP and NHPP, respectively). The HPP assumes that time between failures is Poisson distributed according to a constant mean failure rate throughout time. The NHPP is an extension of this regime, where the mean failure rate is assumed to be time-dependent. The usefulness of this assumption is demonstrated by Slimacek and Lindqvist,[19] who use it to model seasonal effects, and numerous studies which use a power law process to model wear-in, wear-out and serial defect effects.[20]

Another popular method of modelling WTs is modelling time to failure via a Weibull distribution. Like the NHPP, Weibull time-to-failure modelling allows for changing failure intensity through time. In the case where covariates are not included in the model (or where the effect of covariates is constant), the Weibull formulation leads to a failure intensity that is monotonically increasing or decreasing through time.[21]

Such methods might be categorized under the broader heading of reliability analysis in the field of engineering[22] and survival analysis in the fields of biology and medicine.[23] There is a depth of theory in survival analysis in the data science research community which has seen broad application in these fields. However, extensions of the NHPP which permeate survival analysis literature are yet to be fully utilized in the reliability analysis of WTs. To the authors' best knowledge, there are two studies which have made use of such extensions thus far. The first is by Slimacek and Lindqvist,[19] who extended a NHPP by including a frailty model to capture heterogeneity unexplained by model covariates. They also employed the model covariates to explore the effect of 4 factors on turbine reliability, namely: harshness of local environment, turbine concept, date of installation and seasonality. The second is by Ozturk et al,[24] who also explored the impact model covariates affecting turbine reliability. Their list was slightly more comprehensive, encompassing: climatic regions, elevation location, distance to coast, mean annual wind speed, turbine age, turbine type, number of previous failures and scheduled maintenance history. However, for the most part they used simpler non-parametric methods, which differ from the semi-parametric methods which are closely related to NHPPs.

2.3 | Model covariates

Reliability models that are typically used in operational decision making tools for OWFs rarely incorporate model covariates. They tend to assume the same probability of failure distribution for all turbines in the farm, usually under the assumption of a power law process NHPP incorporating wear-in and wear-out effects. There are, however, many studies that attempt to quantify the impact that various features have on WT failure rate via other means. These features, or covariates, can be broadly categorized into the sub-headings *Environmental*, *Operational* and *Design*. Table 2 presents summary of the covariates which have been considered in the literature. Environmental covariates are those which (with the exception of tidal access restrictions) have a direct effect on the turbine loads and operational conditions. Structural reliability analysis methods may often be more suitable for modelling these effects. Operational and design covariates have an indirect effect on reliability of components. We propose that classical reliability analyses are more suited to the inclusion of these features.

It should also be noted that several of the variables listed in Table 2 are correlated, and care should be taken in reliability models to capture this correlation.

2.3.1 | Environmental

Several studies have explored the effect of meteorological conditions on WT failure rate. Reder and Melero[34] stress the potential benefit of altering the constant failure rate used to estimate WT reliability by environmental covariates. In doing so, operators might alter their maintenance strategy to react to, for instance, seasonal periodicity in wind speed. The relationship between wind speed and WT reliability was established relatively early on in the industry's development. Tavner et al[25,26] contributed the majority of this early research, identifying a significant cross-correlation between failure rates and wind speed. Faulstich et al[27] corroborated these early findings. Early researchers also identified correlations between temperature, humidity and proximity to coast and failure rate, implying that the relationship was between reliability and weather, as opposed to solely wind speed.[26] This work was followed up by Wilson and McMillan,[28] who quantified the impact of weather effects on failure rates by employing a Bayesian technique—namely, they used Markov chains and Monte Carlo simulations. They went a step further by quantifying the effect that wind speed-dependent failure rates have on operational expenditure. Later, Reder et al[29] devised a methodology for analysing

TABLE 2 Summary of covariate effects on WT reliability.

Factor considered	Factor category	References
Wind speed	Environmental	Tavner et al.,[25,26] Faulstich et al.,[27] Wilson and McMillan,[28] Reder et al.[29]
Turbulence	Environmental	Tavner and Tavner[30]
Temperature	Environmental	Tavner et al.,[26] Wilson and McMillan,[28] Reder et al.[29]
Humidity	Environmental	Tavner et al.,[26] Wilson and McMillan,[28] Reder et al.[29]
Seasonality	Environmental	Slimacek and Lindqvist,[19] Reder et al.,[29] SPARTA[17]
Harshness of environment	Environmental	Slimacek and Lindqvist[19]
Koppen–Geiger climatic regions	Environmental	Ozturk et al.[24]
Elevation location	Environmental	Ozturk et al.[24]
Distance to coast	Environmental	Faulstich et al[27] and Ozturk et al.[24]
Tidal access restrictions	Environmental	Anderson et al.[12]
Position in the array	Design	Anderson et al.[12]
Number of previous failures	Operational	Ozturk et al.[24]
Scheduled maintenance history	Operational	Ozturk et al.[24]
WT concept	Design	Faulstich et al.,[31] Slimacek and Lindqvist,[19] Ozturk et al.,[24] Arabian-Hoseynabadi et al.,[32] Carroll et al.[33]
WT manufacturer	Design	Slimacek and Lindqvist[19]

environmental conditions ahead of turbine failures in more detail. They also explored an alternative approach using a naive Bayesian network to predict WT failures.[34] Seasonality is explored by a number of studies. Slimacek and Lindqvist[19] and Reder et al[29] both observed higher failure rates in winter months, where wind conditions are harsher. However, Slimacek and Lindqvist[19] also observed a less prominent peak in June and July. Presumably, the summer months offer the most favourable access conditions, meaning that operators use this time to carry out the majority of non-critical maintenance work. This trend was also observed by the SPARTA initiative,[17] who recorded a higher repair rate of turbines in the summer months. The current authors also explored the effects of tidal access restrictions on WT reliability and maintainability metrics.[12] While no disparity was observed between tidally restricted and non-tidally restricted turbines, a statistically significant effect was observed for location in the array with respect to the prominent wind direction.

2.3.2 | Operational

In the wider context of operational decision making tools, the covariates which fit under the *Operational* sub-heading are a key focal point for researchers. In theory, it is conceivable that maintenance actions taken today will have a consequence on WT reliability in the future. There are two prominent factors to consider in this regard:

1. Scheduled maintenance campaigns are used as preventative measures aiming to improve reliability. They therefore might be assumed to have a significant impact on WT failure rates.
2. Modern WTs are equipped with advanced condition monitoring tools for failure prevention.

In reality, however, the impact of operational parameters such as these on failure rates can be difficult to quantify. State-of-the-art decision making tools have various methods of handling them, however drawing inferences from the operational data itself seen relatively little scrutiny in reliability models.

Studies that have scrutinized the impact of scheduled maintenance in the literature tend to optimize the time between preventative maintenance actions. Carlos et al[35] derived an optimal 113 days between scheduled maintenance visits using a Monte Carlo simulation method, using a Spanish failure database. Zhong et al[36] devised a decision making tool based on a nondominated sorting genetic algorithm which allows the user to balance the objectives of turbine reliability and maintenance cost. Besnard et al[37] used a stochastic optimization model to show that preforming service maintenance tasks opportunistically during periods of low wind could save 32% of the transportation and production losses when compared to more conventional approaches. Byon et al[38] used a partially observed Markov decision process to derive an optimal preventative maintenance policy. Pattison et al[39] include annual service campaigns in their novel architecture and system for the provision of reliability-centred maintenance (RCM) such that cost-effective PM management could be achieved. To the authors' best knowledge, no evidence is presented in the literature for increased failure intensity directly after scheduled maintenance works, nor any attempts to model this effect.

Condition monitoring studies represent the main academic contribution to the field of WT O&M.[40] For detailed examples, see Artigao et al,[11] who present a comprehensive review of condition monitoring tools for WTs. Despite this prevalence of CM studies in the literature, attempts to quantify the impact of the developed tools prove to be difficult. The earliest was by McMillan and Ault,[41] who used discrete-time Markov chains to evaluate effectiveness of a CM system on various metrics such as failure rate, O&M costs and availability. The work of May et al[42] followed up, using a hidden Markov model to estimate model reduced failure types, false alarms, detection rates and 6-month failure warnings form CM systems. The dataset made available for this study is not detailed enough to investigate this factor; however, it is conceivable that the efficacy of condition monitoring tools on reducing turbine failure rate could be measured provided some preliminary alarm code analysis.

2.3.3 | Design

By *Design*, we refer to those factors which are determined by the turbine model—that is, turbine concept or manufacturer. Faulstich et al[31] investigated (among other factors) turbine size, which was shown to increase failure rates. This finding is corroborated by Slimacek and Lindqvist.[19] They also compared turbine manufacturers, finding that differences in failure intensity were insignificant. Turbine concept has been explored by Carroll et al,[33] who compared the reliability of WTs with doubly fed induction generators (DFIG) and permanent magnet generator (PMG) drive trains for onshore turbines, finding that the DFIG has approximately 40% more failures than the PMG. Arabian-Hoseynabadi et al[32] compared geared and direct drive turbines but were unable to establish that one was more reliable that the other. Ozturk et al[43] also compared geared and direct drive turbines via a failure modes, effects and criticality analysis (FMECA), finding significant differences in cost criticality in various subcomponents.

3 | METHODOLOGY

3.1 | Reliability analysis

Modelling time between failures facilitates the use of hazard scale models. Within a hazard scale formulation, we can represent the hazard function for a given turbine *i* at time *t* from our designated point of origin via a regression model:

$$h_i(t) = h_0(t)e^{\eta_i(t)}, \tag{1}$$

where $h_0(t)$ is the baseline hazard rate and $\boldsymbol{\eta}_i(t)$ is some linear predictor describing the effect of the chosen covariates on turbine *i* at time *t*. The baseline hazard is therefore the estimated hazard rate in the absence of covariate effects. Typically, WT reliability modelling relies on exponential models:

$$h_i(t) = \lambda_i(t);\ \lambda_i(t) = ae^{\eta_i(t)}. \tag{2}$$

Here, when covariate effects are assumed to be time invariant, we arrive at a constant hazard rate defined by *a*. Note the similarity to the formulation used by Slimacek and Lindqvist[19] to define their HPP. In some instances,[44,45] researchers opt for Weibull-distributed failures. In this case,

$$h_i(t) = \gamma t^{\gamma-1}e^{\eta_i(t)}, \tag{3}$$

where $\gamma > 0$ is a shape parameter. We explore both models by comparing Bayesian fit metrics.

The above model formulations are akin to the popular Cox model.[46] The model we present here, however, differs by three means: via time-dependent covariate effects, frailty (random) effects and Bayesian estimation, which are elaborated on in the following subsections.

3.1.1 | Covariate effects

Covariate effects are wrapped up in the linear predictor $\boldsymbol{\eta}_i(t)$:

$$\boldsymbol{\eta}_i(t) = \boldsymbol{\beta}^T(t)\boldsymbol{X}_i(t), \tag{4}$$

where $\boldsymbol{X}_i(t) = [1, x_{i,1}(t), \ldots, x_{i,P}(t)]$ is a vector of covariates with $x_{i,p}(t)$ being the observed value of the pth covariate for the ith turbine at time t. $\boldsymbol{\beta}(t) = [\beta_0, \beta_1(t), \ldots, \beta_P(t)]$ is a vector of coefficient values in which β_0 is an intercept term and $\beta_p(t)$ is the coefficient for the pth variable. Individual covariates contribute an *effect* via the term $\exp(\beta_p(t))$, otherwise known as the hazard ratio. This represents a relative increase or decrease in the baseline hazard due to the inclusion of a specific parameter. A unit increase in the covariate corresponds to a unit increase in the hazard, proportional to the hazard ratio.

Typically within WT reliability analyses, $\beta_p(t)$ is time invariant, that is,

$$\beta_p(t) = \theta_{p0}, \tag{5}$$

such that θ_{p0} is a constant effect. This is commonly referred to as the proportional hazards assumption and is key in formulating the Cox model. Note that this assumption is implicitly held by Slimacek and Lindqvist[19] and any previous studies that employ covariate effects.[44] Some covariates may violate this assumption, however, by having a time dependence. Such variables will be referred to as having a time-varying effect, since their hazard ratio is a function of time. In this study, we model the time dependence via a B-spline, or basis spline, so-called because it combines a number of weighted basis functions to represent the given relationship. Thus, we model the covariate effect as

$$\beta_p(t) = \theta_{p0} + \sum_{l=1}^{L} \theta_{pl} B_l(t; \boldsymbol{k}, \delta), \tag{6}$$

where θ_{p0} is a constant, $B_l(t; \boldsymbol{k}, \delta)$ is the lth basis term of a degree δ B-spline function subject to a vector of knot locations k and θ_{pl} is the corresponding lth B-spline coefficient. Such a formulation utilizing B-splines has been shown to be effective for modelling time-varying effects, as shown by (e.g.) Perperoglou,[47] Andrinopoulou et al[48] and Gao et al.[49] Their effectiveness, however, depends on two hyper-parameters, namely, δ and the knot vector $\boldsymbol{k}$. Regarding the choice of δ, cubic basis functions (i.e., $\delta = 3$) are most popular in the literature.[50] Given the property of continuous first and second derivatives, they provide smooth interpolation when compared to linear and quadratic basis functions.[51] They also avoid some of the issues associated with higher order polynomials, for example, Runge's problem.

Knot locations present a more complex problem. Improper selection of knot locations can lead to poor predictions via either overfitting (in the case where there are too many splines) or underfitting (in the case where too few knot locations).[52] The literature presents two main ways of handling this issue. The first employs some form of penalty against overfitting, as is common to many machine learning methodologies. This involves the selection of another smoothing parameter or penalty weight, which is optimized to achieve the best 'smoothness' between adjacent splines.[47,53] We opt for the alternative approach, whereby several configurations of knot locations are considered and compared via a model selection criteria,[54] such that a model is selected parsimoniously. More information on model comparison is given in Section 3.3.3.

3.1.2 | Hierarchical reliability

WT reliability is difficult to properly quantify. There are myriad different covariates that have an influence. On top of this, the complete set of influential environmental, operational or design parameters that might be considered to improve the accuracy of any reliability model is rarely available to researchers—or indeed even operators. Often there is an additional heterogeneity between systems which is unexplained by model covariates. In a previous study, we have explored the use of hierarchical models in addressing issues of heterogeneity within a limited dataset. Indeed, they proved to be useful in quantifying WT reliability and in turn for assessing strategy to address failures. In survival or reliability analysis, this hierarchical nature is commonly accounted for by including *frailties* in the model.[55] In effect this clusters observations into groups sharing a common characteristic; here, we include a turbine-by-turbine frailty. This means that any ith event belonging to any jth given turbine are correlated, which can be modelled via a random effects term in the linear predictor:

$$\boldsymbol{\eta}_{ij}(t) = \boldsymbol{\beta}^T \boldsymbol{X}_{ij}(t) + \boldsymbol{b}_j^T \boldsymbol{Z}_{ij}, \tag{7}$$

where $\boldsymbol{X}_{ij}$ has been given a subscript to denote events belonging to the turbine j, $\boldsymbol{Z}_{ij}$ is a vector of covariates for the ith event for the jth turbine and b_j is the associated vector of turbine-specific random effect parameters with covariance matrix $Cov(\boldsymbol{b}_j) = \boldsymbol{\Sigma}_b$ and expected value $E(\boldsymbol{b}_j) = 0$.

Including frailty effects not only has the potential to improve the accuracy of the model but addresses how to alter the traditional Cox-like parameterization to include recurrent events for individual turbines. As explained by Amorim and Cai,[56] there two options for datasets of this type. The first is to employ a Markov process, where future events depend only on the immediate past. The other is via shared random effects or frailties, as we have done here. This is a natural choice for WTs: Failure rates may differ significantly even among fleets of the same model, and it is often difficult to attribute any discrepancies to observable covariates.

3.2 | Bayesian reliability analysis

3.2.1 | Priors

The specified model has a number of parameters for which prior distributions must be defined. For smaller datasets, setting informative priors would be useful for allowing statistically relevant inferences. As it stands, we argue that the dataset upon which this analysis is based is large enough to allow weakly-informative priors. Additionally, in the absence of structured expert opinion it is difficult to define informative priors for many parameters. The time-varying effect of maintenance works on failure rate, for instance, is a novel covariate for which no data points exist. Neither is the effect accounted for by state-of-the-art decision tools. The covariance matrices that characterize the frailties of individual turbines would vary from farm-to-farm.

Having determined that priors will be weakly-informative, we set the following assumptions:

1. *Intercept value* β_0. Models contain an intercept term in the linear predictor which partly characterizes the baseline hazard. It takes the form of a diffuse normal distribution such that $\beta_0 \sim N(\mu,\sigma)$, where $\mu = 0$ and $\sigma = 5$.
2. *Time-invariant regression coefficients* θ_{p0}. As above, we use a normal distribution such that $\theta_{p0} \sim N(\mu,\sigma)$, where $\mu = 0$ and $\sigma = 5$.
3. *Time-varying B-spline coefficients* θ_{pl}. We describe the B-spline coefficients via a random walk where $\theta_{p,1} \sim N(0,1)$ and $\theta_{p,m} \sim N(\theta_{p,m-1},\tau_p)$ for $m = 2,\ldots,M$, where M is the total number of basis terms. Using this specification also requires a prior for τ_p, which we define as an exponential such that $\tau_p \sim \exp(\lambda)$, where $\lambda = 5$. τ_p acts as a smoothing parameter for the B-spline which can be used as a form of regularization. Thus, the exponential form of prior is used to encourage low values and prevent overfitting.
4. *Wiebull scale parameter* γ. Where we employ a Weibull model, the scale parameter is assumed to follow a half normal distribution, such that $\gamma \sim N(\mu,\sigma)$, where $\mu = 0, \sigma = 1$ and $\gamma > 0$.
5. *Covariance matrix*. The covariance matrix is made up of a series of decompositions. First, the covariance matrix $\Sigma_{\boldsymbol{b}}$ is decomposed into a correlation matrix $\boldsymbol{\Omega}$ and a vector of variances. The variances are in turn decomposed into the product of a simplex vector $\boldsymbol{\pi}$ and the trace of the matrix. The trace is the product of the order of the matrix and the square of a scale parameter ϵ. The Lewandowski–Kurowicka–Joe (LKJ) distribution[57] is used as a prior for $\boldsymbol{\Omega}$—this is dependent on a regularization parameter ζ. We use the default $\zeta = 1$, meaning that the prior distribution is jointly uniform over all possible correlation matrices. A symmetric Dirichlet prior is used for π, which has a single concentration parameter $\phi > 0$. Again, we use the default $\phi = 1$, meaning jointly uniform probability for all simplexes. A gamma distribution is specified for ϵ, such that $\epsilon \sim Gamma(\alpha,\beta), \alpha = 1, \beta = 1$.

Figure 2 summarizes the relationships between model variables for the Weibull model.

3.2.2 | Posterior calculation

Posterior calculation is where we condition our model on observed data via Bayes's rule. It is a key difference from the likelihood methods as previously used for inference by Slimacek and Lindqvist.[19] Given a set of observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ assumed to be independent given the parameters $\boldsymbol{\theta}$, a set of prior distributions $p(\boldsymbol{\theta})$, the posterior distribution is derived by

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{8}$$

where the non-zero probability of observation $p(\boldsymbol{y})$ of $\boldsymbol{y}$ is given by

$$p(\boldsymbol{y}) = \int p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{9}$$

a quantity that often doesn't readily allow for analytical solutions. For this reason we use Markov chain Monte Carlo (MCMC), a popular inference method for building Bayesian models.[58] We use a form of Hamiltonian Monte Carlo (HMC) called the No-U-Turn-Sampler (NUTS),[59] which

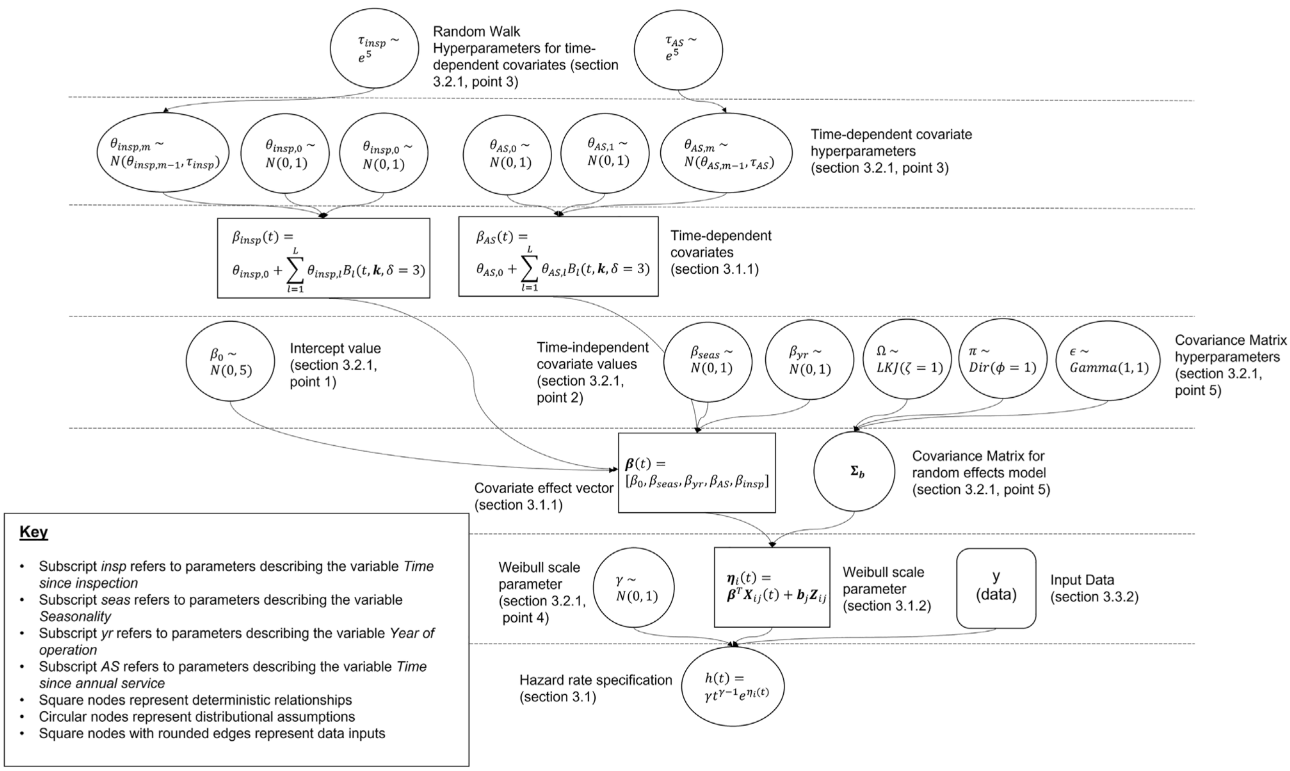

FIGURE 2 Diagram showing the relationships between model parameters and hyper-parameters as described in Section 3.2.1, along with their associated prior specifications.

benefits practitioners in not having to hand-tune HMC via a parameter which defines the number of steps in the simulation. We use 1000 iterations as an initial training period for the algorithm. These training samples are then discarded, and we go on to sample 1000 draws from the target posterior distribution.

3.2.3 | Model comparison

We use leave-one-out (LOO) cross-validation (CV) to undertake model comparison.[60] LOO was developed specifically for estimating out-of-sample predictive accuracy from the posterior sample of fitted Bayesian models. Consider a dataset $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, which has been generated by some theoretical true generating mechanism for $\mathbf{y}, p_t(\mathbf{y})$. Also consider a new dataset $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)$. $\tilde{y}$ is independent from $\mathbf{y}$ but is also assumed to be generated by $p_t(\mathbf{y})$. In the Bayesian setting, we arrive at a posterior predictive distribution for the new dataset $\tilde{\mathbf{y}}$ fitted on the original dataset $\mathbf{y}$:

$$p(\tilde{\mathbf{y}}|\mathbf{y}) = \int p(\tilde{y}_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}. \tag{10}$$

Evaluating a given set of models $M_k \in \{M_1, M_2, \ldots, M_z\}$ involves estimating the expected log point-wise predictive density (ELPD), which acts as a measure of predictive accuracy for the dataset $\tilde{\mathbf{y}}$. The ELDP is given by[61]

$$ELDP(M_k|\mathbf{y}) = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log\, p_k(\tilde{y}_i|y) d\tilde{y}_i, \tag{11}$$

where $p_k(\tilde{y}_i|y)$ is the posterior predictive density for the model M_k. In practice. however, what we refer to as the *true* data generating process is unknown. We therefore must resort to some means of approximation for Equation (11). In many machine learning methodologies, out-of-sample test data are used for a similar purpose. In the Bayesian framework, CV[62] has become popular. CV involves splitting the data into K parts which are, respectively, used as out-of-sample validation sets for the model fit with the remaining data. In LOO CV, we estimate the predictive accuracy of the n data points taken one at a time, such that $K = n$. This is given by

$$ELPD_{LOO}(M_k|y) = \sum_{i=1}^{n} \log p_{M_k}(y_i|y_{i-1}), \tag{12}$$

where

$$\log p_{M_k}(y_i|y_{i-1}) = \log \int p_{M_k}(y_i|\theta) p_{M_k}(\theta|y_{i-1}) d\theta \tag{13}$$

is the LOO predictive log density for the ith data point of the model M_k. In practice, we are only concerned with the comparison of two models, so we estimate the difference in their expected predictive power. For two models M_1 and M_2 (e.g., representing models with different knot vectors $\boldsymbol{k}$), we estimate the difference in their expected predictive accuracy as

$$\begin{aligned} \widehat{elpd}_{LOO}(M_1, M_2|y) &= \widehat{elpd}_{LOO}(M_1|y) - \widehat{elpd}_{LOO}(M_2|y) \\ &= \sum_{i=1}^{n} (\log p_{M_1}(y_i|y_{i-1}) - \log p_{M_2}(y_i|y_{i-1})) \\ &= \sum_{i=1}^{n} \widehat{elpd}_{LOO,i}(M_1, M_2|y). \end{aligned} \tag{14}$$

Now that there is a model comparison criterion in place, we employ a two stage step-wise procedure for model building as developed by Hofner et al.[63] This consists of the following steps:

1. *Starting model definition*. The starting model consists simply of the baseline hazard rate with turbine frailties. There are two choices for baseline hazard rate, namely, exponential and Weibull. This is the initial model selection problem. Note that frailties must be included from the outset to account for recurrent events among the same turbine.
2. *Initial choice set*. Define a set of variables to be, respectively, added to the model. These are defined in Section 3.3.2. The primary problem is in selecting the number of knots for time-varying effects. We do this by increasing the number of knots to the point where the increase in model accuracy is insignificant. If any respective variable improves the model's predictive accuracy it is included.
3. *Backwards selection*. Perform a backward deletion step on the current model, that is, estimate all hazard regression models obtained from the current model by dropping one covariate at a time. If an improvement of the model comparison criterion can be achieved, make the reduced model with optimal model comparison criterion the working model.

The process is summarized by Figure 3.

3.3 | Data preprocessing

3.3.1 | Dataset description and identification of failures

The current analysis is based on the same historical database as is used by Anderson et al.[12,64] It consists of supervisory control and data acquisition (SCADA), weather and operational data describing the performance, meteorological conditions and maintenance actions performed for a large group of modern, geared WTs with a multi-MW power rating. The farm utilizes a typical condition-based maintenance strategy and is located around an hour-long trip from its operational base by CTV.

Of particular importance to this study is the *Downtime Catalogue*, which was created by cross-referencing technician work orders and SCADA data, such that every period of turbine downtime is associated with a corresponding set of maintenance tasks. There are some important characteristics about *Downtime Catalogue* which should be highlighted for interpretation of results going forward.

1. Entries consist of periods of turbine downtime coinciding with maintenance actions. 'Failures' are selected by downtime events which have a coinciding maintenance action labelled as 'corrective'. Corrective actions are distinct from inspections and balance of plant activities: they correspond directly to a repair of the turbine. However, specific failure modes or components affected cannot be ascertained at this stage.

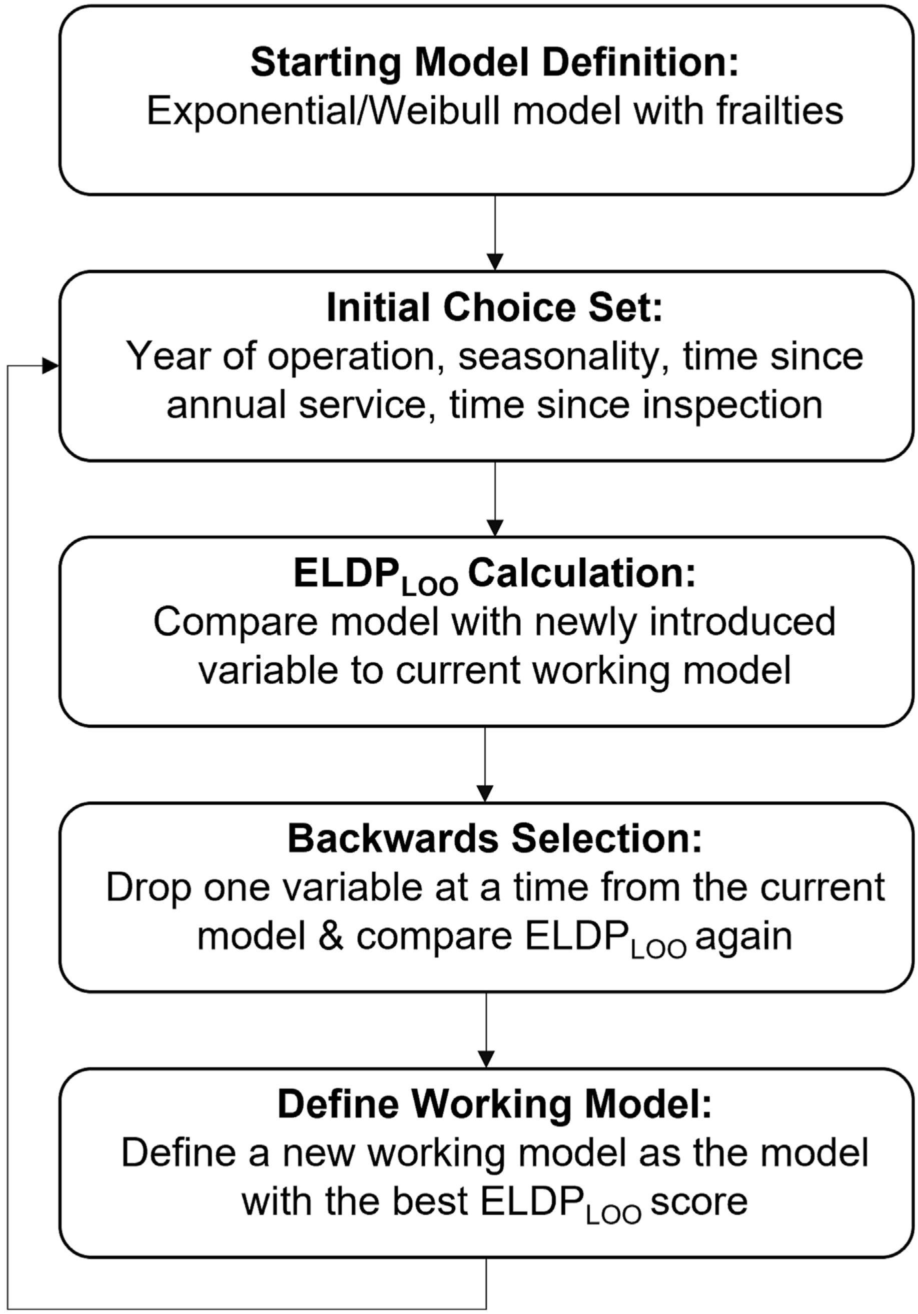

FIGURE 3 Flow chart summarizing the model selection criteria.

2. A 'failure' implies that something has gone wrong with the turbine and it can no longer produce power. In this study, the onset of turbine downtime is the qualifier for a failure. The time between the end of one corrective downtime event and the start of the next is what we measure. This is different from the time between corrective maintenance actions. Beyond ascertaining that a corrective maintenance action occurred within the period of downtime, the timing of them is irrelevant.
3. Using 'failure' in this context is justified in the absence of a standard failure definition in the industry. However, uncertainty remains because the onset of downtime may not be regarded as the onset of a failure by all researchers. In the absence of a detailed root-cause analysis, we deem a turbine shutdown a reasonable threshold to define a failure.
4. Restricting the identification of failures to the turbine-level potentially obfuscates dependencies that are only affect specific failure modes. There would therefore be a lot of value in scrutinizing particular failure modes, but it is outwith the scope of this study. On the other hand, restricting failures to turbine level avoids some uncertainty associated with small datasets; that is, there are far more samples at the turbine level than component level.

3.3.2 | Selected covariates

The key contribution of this study in in quantifying the effect of operational covariates on WT reliability. The time-varying effect of scheduled maintenance works on failure rates is a particularly novel consideration. The full list of covariates considered is as follows:

1. *Seasonality*. Simalarly to Slimacek and Lindqvist,[19] we employ seasonality as a model covariate. This is either coded by month of the year (i.e., 'Jan', 'Feb', …) such that each month has a constant covariate effect with respect to a reference month (which is arbitrarily chosen to be April) or season of the year under the same assumption (reference month is arbitrarily autumn).
2. *Year of operation*. Year of operation is of interest because WT failure behaviour is often assumed to vary over time, most frequently by a power law process. Failure intensity is often assumed to follow the bathtub curve,[65] which includes a wear-in period and wear-out period. One study by Stiesdal et al also included a serial defect period for Siemens turbines.[20] We assume that the data cover the normal operations period in the wind farms life; however, there are few results in the literature to conform that this is actually flat. Also of interest is the fact that the site deployed their advanced data management system in 2018, which in itself may have improved operational efficiency.
3. *Turbine location*. Turbine location may be assumed to effect reliability primarily due to the effects of turbulence. Indeed, we have explored this effect in a previous publication.[12] We include this effect via a frailty term, either using the turbine row or individual turbine names as categorical variables.
4. *Time since annual services*. It is common practice in the industry to have an annual campaign where a set of scheduled maintenance actions are carried out on turbines. The exact nature of these actions vary from one service provider to the next and depend largely on maintenance contract arrangements.[66] Contracts generally assign up to 60 downtime hours per year towards these services; however, a recent study by Anderson et al[12] estimates the number to be around 29 h/year. This generally includes tasks such as lubrication of mechanical parts (e.g., gear oil, hydraulic oil and greasing), measurement of part temperatures, a torque tensioning of bolts and basic inspection of parts within the nacelle. Analysis of the effect of annual services on failure rates was a key motivation for introducing time-dependent effects, as the data provider for this study reported allegorical evidence of repeated failures after annual services from technicians.
5. *Time since inspection*. Likewise, it is common to regularly inspect certain key components—for example, a visual inspection of blades for cracks or erosion. Again, we assume a time dependence for this variable; it is conceivable that a repair is more likely to take place soon after an inspection, but thereafter, the risk of failure might be assumed to decrease.

3.3.3 | Time-to-event dataset

The *Downtime Catalogue* acts as the basis of this study. However, it requires an additional transformation step in order to convert the information into a data format that is compatible with the methodology. We refer to this as the time-to-event dataset, a subset of which is shown in Table 3. Most important are the columns *tgap* and *failure*, upon which the model fundamentally depend. *tgap* represents the time since the last maintenance action, that is, the difference between *tstart* (the time stamp at which the turbine is restored to fully operational after the previous maintenance action) and *tstop* (time time stamp at which the turbine stops producing power due to some maintenance intervention). The rest of the columns are covariate values. Time is recorded in units of days.

We propose that, at a turbine level, there are enough samples in the time-to-event dataset to set un-informative priors. However, 600 turbine years of data is a relatively small amount for a reliability analysis. As with most data analysis techniques, results would be more robust if the dataset were bigger. The dataset is made up of turbines of one concept from one OWF. Results should be interpreted with these facts in mind.

TABLE 3 First five rows of the time-to-event dataset.

ID	Turbine	Turbine row	tstart	tstop	tgap	AS	Inspection	Failure	Month	Year	Season
1	A1	A	0	15	15	0	0	1	7	2018	Summer
2	A1	A	29	49	20	0	1	1	8	2018	Summer
3	A1	A	50	87	37	0	0	1	9	2018	Autumn
4	A1	A	99	115	16	1	0	1	10	2018	Autumn
5	A1	A	115	117	2	0	0	1	10	2018	Autumn

4 | RESULTS

4.1 | Model comparison and knot locations

Table 4 summarizes the results of the step-wise model comparison procedure. Immediately from the starting model definition, we see that a Weibull model significantly outperforms the exponential model in representing the time-to-failure dataset. From visual inspection of Figure 4, it is evident that the Weibull model also matches more accurately the non-parametric Kaplan–Meier estimate.[67] Note that we re-check this comparison at each stage of the procedure, and results are consistent. This in itself is an interesting result. According to Seyr and Muskulus,[1] WT failures are most commonly modelled exponentially via a Poisson process. Our results imply that a Weibull-distributed time to failure would be more accurate. Scheu et al[5] conclude that the difference in modelled availability between exponential and Weibull-distributed failures is as much as 10%. This distributional assumption can therefore have a significant impact on the outputs of cost models—it may benefit the research community to investigate this disparity further. Weibull versus Poisson process for the modelling of individual failure would be particularly useful, as individual failure modes have different consequences in terms of costs.

Seasonality has a lesser, though still significant effect on modelling performance, as does year of operation. Time since annual service also improves model performance—the degree to which it does so depends on the knot vector. We see a quite significant improvement when adding 5 internal knots (compared to 0), but the improvement quickly tails off as additional knots are added. The peak value is at 7, after which there is a very small depreciation in model performance. Given that all internal knot configurations containing greater than seven have a small $ELPD_{LOO}$ score compared to the model error (SE), they have similar predictive power and we should choose the least complex model. Modelling both time since annual servicing and time since inspection provides the most accurate results. Time since inspection is most accurately modelled with five internal knots—perhaps this is fewer than time since annual service since it is characterized by a less dramatic spike in the first few weeks after performing the task. The effects of increasing number of internal knot locations is shown in Figure 5.

4.2 | Covariate values

The effect of seasonality is shown in Figure 6. This is most likely indicative of the strategy employed at the site rather than direct wind speed effects on turbine reliability. The operator favours the low wind speed months of June and July to perform repairs. This corroborates the

TABLE 4 Summary of covariate effects on WT reliability.

Modelling group	Added variable	Modelling alternative	LOO-score (ELPD)	LOO-score (SE)
1	Baseline	Weibull	0	0
		Exponential	−588.3	32.8
2	Seasonality	Seasons	0	0
		Monthly	−11.9	6.4
		No covariate	−38.5	10.2
3	Year of operation	Yearly (constant effect)	0	0
		Yearly (year-by-year)	−2.0	0.9
		No covariate	−47.6	9.3
4	Time since annual service	TVE (7 internal knots)	0	0
		TVE (8 internal knots)	−0.4	0.3
		TVE (9 internal knots)	−0.6	0.4
		TVE (10 internal knots)	−1.1	0.5
		TVE (6 internal knots)	−4.1	0.7
		TVE (5 internal knots)	−4.6	1.1
		No covariate	−28.1	5.9
5	Time since inspection	TVE (5 internal knots)	0	0
		TVE (6 internal knots)	−1.3	0.3
		TVE (7 internal knots)	−1.4	0.4
		TVE (4 internal knots)	−3.4	0.8
		TVE (2 internal knots)	−9.2	1.9
		No covariate	−18.7	5.1

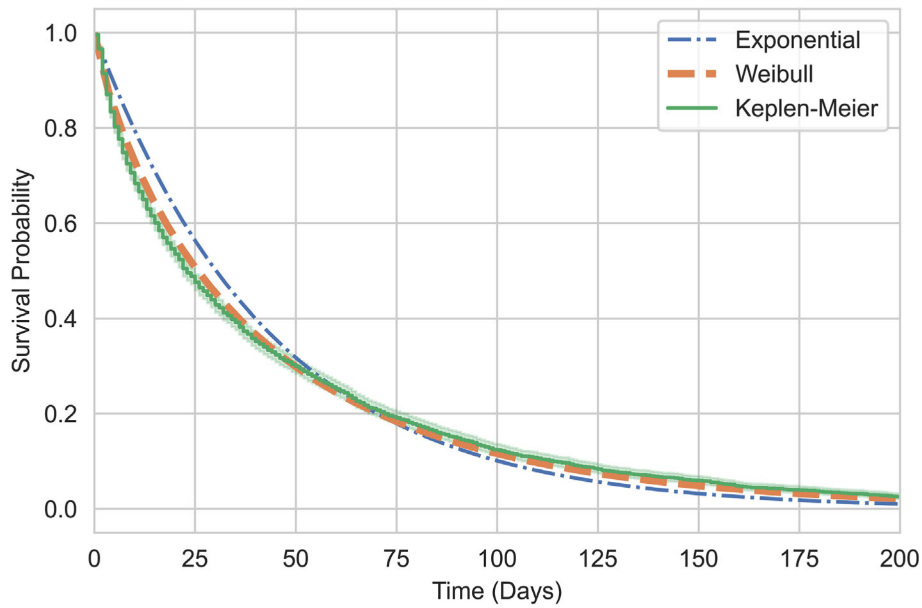

FIGURE 4 Survival curves for both baseline hazard rate distributions. For reference, the non-parametric Kaplan–Meier curve is also shown.

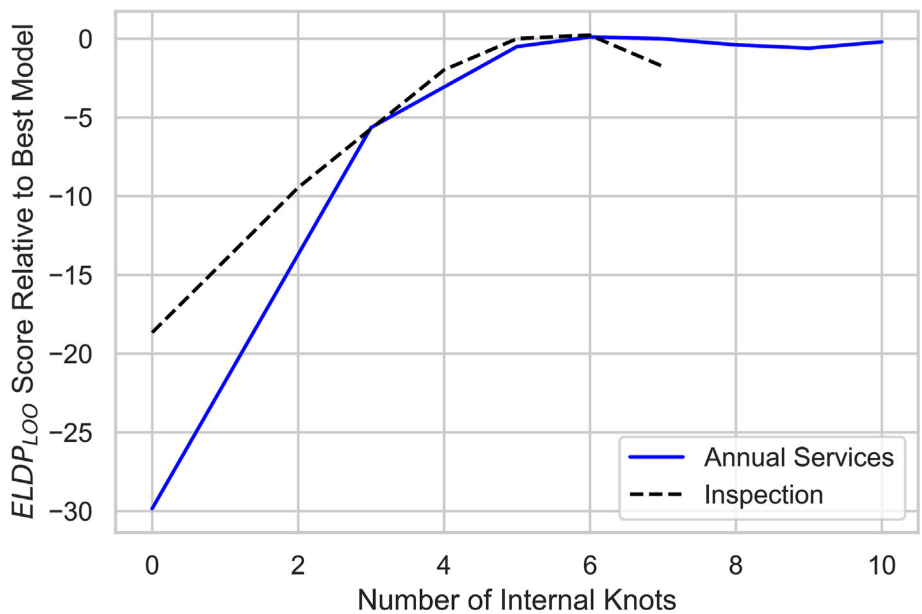

FIGURE 5 $ELPD_{LOO}$ values for models of varying internal knot locations.

seasonality pattern of repair rate presented by the SPARTA initiative.[17] It does not align with the results presented by Slimacek and Lindqvist,[19] who observe a similar uptick in repairs during summer, but increased failure rates throughout winter. The disparity may be indicative of improving operational efficiency with growing experience in the industry; however, it may also be attributed to different methods of modelling seasonality. The high hazard rate in autumn is unexpected given the patterns previously presented in the literature. It is unclear exactly what causes this—it has been suggested that there may be a 'fatigue effect' after the busy summer months, but this is difficult to verify.

The effect of year of operation is shown in Figure 7. The effect on the hazard rate is with reference to the first year for which data are available, 2018. There is a consistent and quite significant reduction of the hazard rate through time. What causes this is not clear. It may be indicative of increasing operational efficiency through learned experience of operating the site. It may imply that the bathtub curve describing infant turbine failure rates has a longer tail than commonly assumed. However, Carroll et al[14] found no evidence of increased infant failure rates, and we

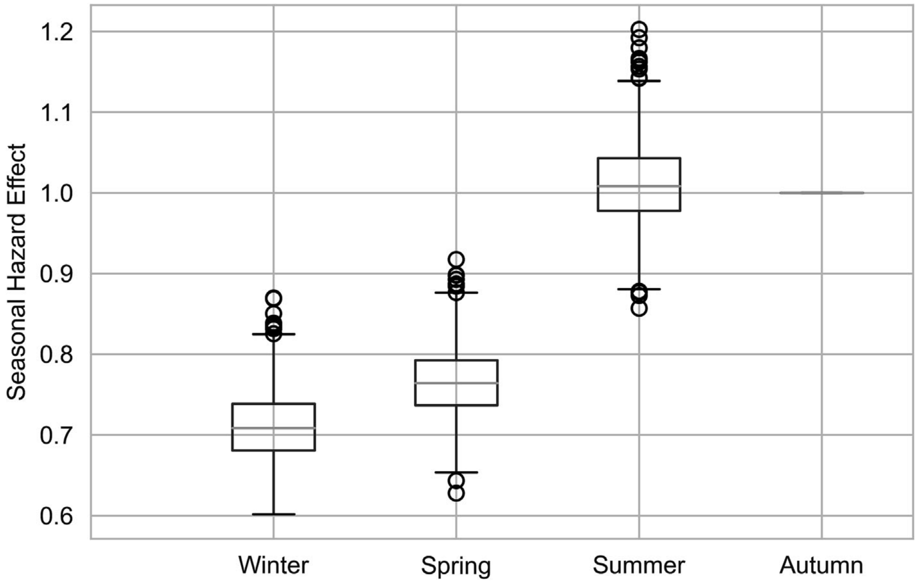

FIGURE 6 Effect of year of seasonality on WT hazard rate, relative to autumn.

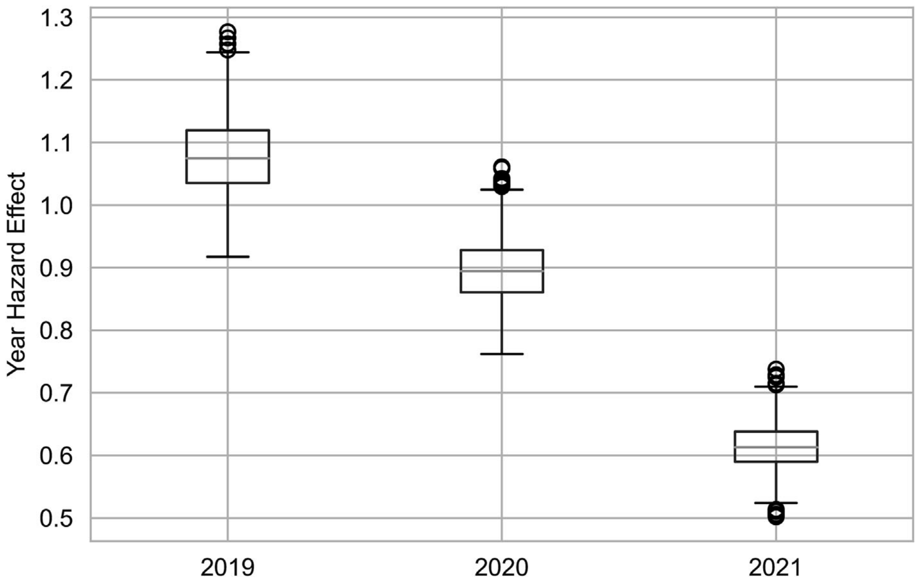

FIGURE 7 Effect of year of operation on WT hazard rate, relative to the year 2018. Since data collection commenced in 2018, there has been a consistent downward trend.

consider it to be unlikely. Interestingly, it coincides with the introduction of an advanced data management system to the site in 2018. It may therefore present evidence of the value of data management systems regarding operational efficiency. In the absence of more detailed fault data, it is difficult to draw any hard conclusions. Hazard rate does not necessarily translate directly to failure rate. It may be that repairs are carried out in a way that cause fewer stoppages or that there are fewer 'fault finding' missions.

Frailty effects are explored in Figure 8. Figure 8A shows how the means of all Monte Carlo simulations for each turbine are distributed—approximately normally around a median of 1. There are quite significant deviations from the mean. The most unreliable turbine is predicted to fail 1.65 times more that average, the most reliable 0.5 times the average. This is quite a significant heterogeneity in failure intensity that is rarely taken into account in costs models. Again, further research into how this effects cost modelling outputs may be useful for researchers in the field. Figure 8B shows how all of the frailty estimates (i.e., from every Monte Carlo simulation) are distributed, which is useful for checking the assumption that frailties are indeed gamma distributed. Evidently, the random effect estimates can be accurately approximated by a gamma distribution, as the fitted gamma curve fits the histogram very well.

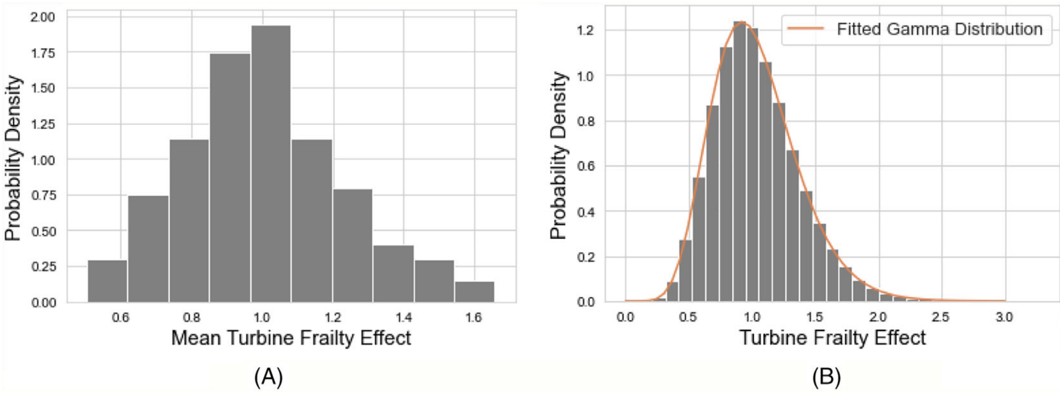

FIGURE 8 Histograms of (A) the mean frailty effect for each turbine and (B) the turbine frailty effect over all Monte Carlo simulations. The assumption of gamma-distributed random effects is a reasonable one.

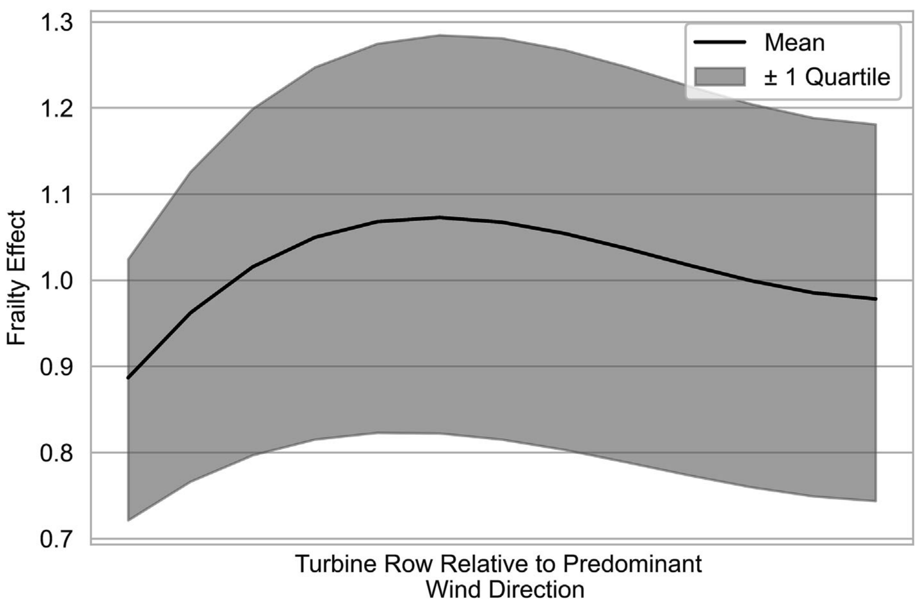

FIGURE 9 Fitted cubic function approximating frailty effects' relationship to turbine position in the array with respect to the prominent wind direction.

Frailties are also useful for exploring the effect of position in the array, as has previously been done by Anderson et al.[12] Figure 9 explores the effect that turbine row has on frailty values. It does so by fitting a cubic spline function to the mean frailty values of turbines in each row, as well as the 25th and 75th percentile values for each row. Values towards the left of the plot on the *x* axis represent rows which are towards the front of the array with respect to the prominent wind direction. The first row is characterized by a relatively low failure intensity and relatively low uncertainty in the failure intensity. The next foremost rows are characterized by both an increasing hazard rate and increasing uncertainty in the hazard rate. This trend continues to a point in the middle of the array. After this, successive rows retain high uncertainty in hazard rates, with mean values decreasing towards the back of the array. Turbines downwind of others are subject to higher turbulence, which has been shown to be detrimental to WT reliability.[30] At the same time, downwind turbines operate in lower speeds, so we see a decreasing median trend from the rows in the middle to the rows at the back. Interestingly, these 'middle' rows are characterized by the highest heterogeneity in reliability.

Figure 10 shows the time-dependent covariate *time since annual service*. It is characterized by a sharp initial upward trend peaking at 2 days. The maximum value for mean hazard ratio estimate at this point is 1.57. Turbine failure intensity is higher for the first 6 days after the service, after which there is a reduced failure intensity until 137 days after. Beyond this point, failure intensity is estimated to increase, as might be expected for turbines far away from their annual service. Lower values along the time axis have lower uncertainty in their estimate as the majority of turbines will most likely require corrective maintenance again in the next few weeks than in the next few months. Hence, estimates further along the *x* axis are more uncertain as there are fewer samples. The results back up the theory of the operational team—initially, annual services can be thought to lead to more failures, after which they prevent failures as they are supposed to. However, the effect is short lasting, and the higher failure intensity may not justify any change in strategy.

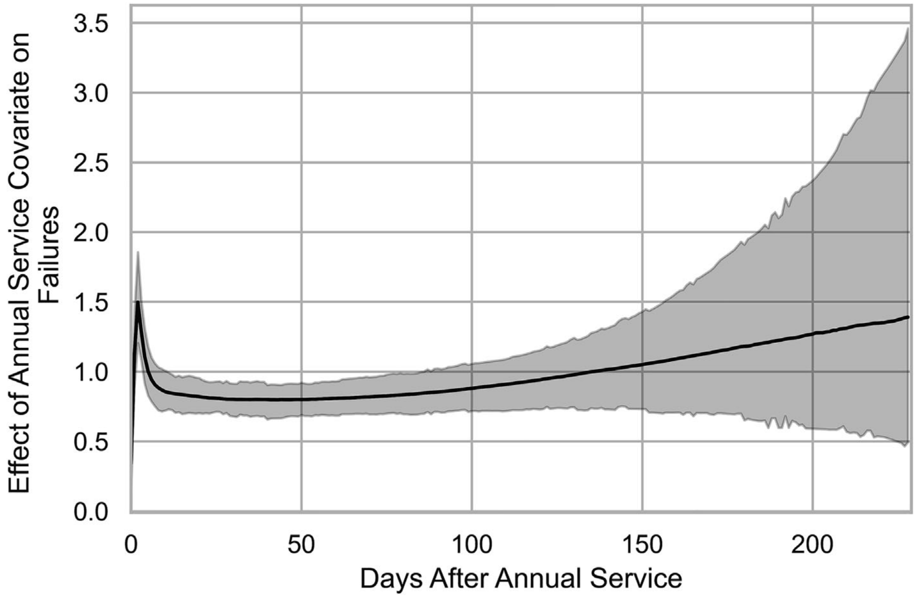

FIGURE 10 Time-dependent relationship between annual servicing and wind turbine failure rate. A hazard ratio <1 signifies an increased failure rate, <1 decreased.

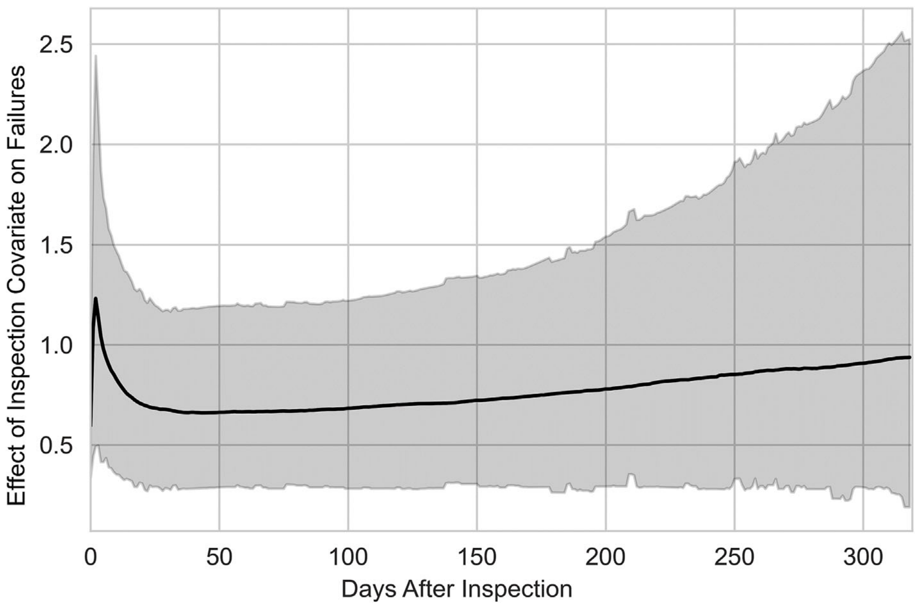

FIGURE 11 Time-dependent relationship between inspection and wind turbine failure rate. A hazard ratio <1 signifies an increased failure rate, <1 decreased.

Figure 11 shows the time-dependent covariate *time since inspection*. The shape is similar to *time since annual service*: There is an initial peak in failure intensity (where inspections report a problem leading to a corrective maintenance action), after which the hazard rate is below 1. Again, the initial peak is at 2 days after inspection, with a slightly smaller hazard ratio of 1.30. 6 days after servicing, the hazard rate falls below 1. It remains below 1 throughout the entire year, rising gradually as time goes on.

5 | DISCUSSION

5.1 | Reflections on methodology

There are two ways to consider the utility of this study. The first has to do with the context and quality of the input data, the second to do with the methodology itself. Regarding the input data: the data-table *Downtime Catalogue* results from an operational database provided by a currently operational OWF which employs an advanced data management system. Given the scarcity of reliability data available for offshore WTs (see Reder et al[2] for details), this is valuable to the research community. However, the dataset is limited in that it is not broken down into a component or subcomponent taxonomy. This means that no conclusions can be made on the time to failure of individual components, the effect of seasonality on individual components or the effect of year of operation on individual components. Since some modern cost modelling tools consider reliability at the component level, such an analysis would be useful to the research community. This is especially pertinent for the time-varying effect of inspections. Figure 11 does not reveal anything particularly useful—as expected, the inspection reveals the need for corrective action and so there are likely to be corrective actions following. However, future work plotting Figure 9 on a component by component basis with a view to seeing how inspections of a given component correlate to that component's failure rate would be more valuable. It may also be of value to consider the time-varying effect of annual services on specific components, such that operators can consider changes in how scheduled maintenance is carried out. That being said, the approach we have used here can be easily adapted to a subcomponent taxonomy or indeed can be applied to repairable systems other than WTs. Also, uncovering this relationship at the turbine level is also valuable, as it has previously not been formally quantified.

Regarding the methodology itself, the alterations we propose to the traditional NHPP are novel in the context of WT reliability. Given the growing recognition in the wind research community that more care must be taken uncertainty handling, extending the NHPP into the Bayesian regime is a natural extension. The advantages of implicit uncertainty handling by Bayesian models are well documented and often cited.[68] Interpreting uncertainty via Bayesian interval estimates, which are interpreted as the interval containing the true parameter with some probability, are generally seen as more intuitive than confidence intervals, which are interpreted as the range of values containing the true parameter a certain percentage of the time given a large sample approximation.

Inclusion of time-dependent variables is the second extension. In this case, it proved to be a useful one, as it allowed us to explore the effect of varying failure intensity proceeding scheduled maintenance works which was posited by the operational team. There is one important caveat in employing time-dependent variables, though. Namely, the selection of knot locations in B-splines takes care, as different assumptions might lead to quite significantly different results. The parsimonious model selection methodology we used was relatively computationally expensive, considering the usual run time of survival models.

The final methodological point has to do with failure definition. As noted by Leahy et al,[69] there is no standard definition of a failure in the wind industry. The choice of failure definition often depends on the data available to the researcher. Here, we used downtime as a qualifier for a failure, when that downtime coincided with a maintenance action recorded as corrective in the database. In the absence of a standard failure definition in the industry, this choice of failure definition becomes a model hyper-parameter in itself which may significantly affect the results. Of particular importance is the timing of failure onset, which is not necessarily the onset of downtime recorded here (especially for minor repairs). However, for many operational datasets, it would be difficult to decipher any failure onset time beyond the onset of downtime.

5.2 | Future analysis based on results

We propose that there are three ways to gain further insight from the results presented. The first is by the use of a cost modelling tool, for instance, the one developed by Dinwoodie et al.[70] To begin with, a comparison of the different baseline distributional assumptions could be informative. This would build on the work of Scheu et al,[5] which investigated different baseline hazard rate assumption. Two factors would provide novelty: the fact that the results presented here are based on real world data from a currently operational wind farm and an investigation into not only availability but the impact on levelized cost of energy (LCoE) or some other financial metric. The results of such a study would further highlight the importance of modelling statistical uncertainty accurately.

There could also be value in a similar study where the impact of heterogeneity in turbine reliability is investigated. Given that mean values of turbine-by-turbine frailty are approximately normally distributed, this may not have such a significant impact as if there were gamma-distributed random effects with a long tail. For instance, it would be interesting to see whether inclusion of random effects in cost models would significantly effect the financial outputs, especially for a large OWF.

The second avenue for further research is in optimizing the timing of scheduled maintenance campaigns and the resources dedicated to them. Figure 10 has three stages to consider. First, there is the initial spike in failure intensity which occurs during the first few weeks after a scheduled maintenance work. This may not inspire any drastic changes to when services are performed, as typically annual service campaigns are carried out in summer where lost production is low and access favourable. However, if operators could quantify the additional risk in terms of corrective maintenance, they could make a more informed decision about how many technicians and vessels they require.

The third avenue is in including more detailed data about exactly what work was carried out on the turbine. By doing so, the proposed methodology would be able to uncover dependencies between the selected covariates and specific failure modes. These data were not available for research presented in this article; however, it will be available for future works.

6 | CONCLUSION

This work presents a Bayesian reliability analysis of WTs operating within a large OWF. We employ the typical Poisson process and Weibull-distributed time-to-failure assumptions. However, we transitioned from the frequentest to the Bayesian regime, which has the primary benefit that it allows for built-in and easily interpretable uncertainty quantification. We propose a model which incorporates time-dependent variables in response to the operational team's reasoning that turbine failures are more frequent following annual services. This was indeed the observed behaviour, with a higher turbine failure intensity for 6 days after annual service works take place. There is a peak failure intensity reaching 1.57 times an average 2 days after servicing. A similar though less significant effect was observed with the covariate representing time since inspection. The higher failure rate lasted for 6 days after the maintenance work was carried out, reaching a peak of 1.302 days after. We also observed seasonality effects similar to those presented by the SPARTA initiative and a consistent year-on-year reduction of failure intensity. We propose that this may be partly due to the wind farm employing more advanced data management and control room software in 2018. Significant turbine frailties were observed. Plotting these by turbine row, a dependence of reliability on turbine position with respect to the prominent wind direction is observed. We suggest three avenues for future work. The first explores the effect of different hazard rate assumptions on the outputs of some cost modelling tool, in particular the different baseline hazard distributions and the effect of turbine-by-turbine random effects. The second utilizes the time-dependent hazard ratio of the variables *time since annual service* and *time since inspection* to optimize dynamic resource allocation. Quantifying financial risk due to scheduled maintenance works could be particularly beneficial for operations. The third involves incorporating more detailed work procedure data, so that correlations between covariate effects and specific failure modes could be uncovered.

ACKNOWLEDGEMENTS

This work was supported through the UK's Engineering and Physical Sciences Research Council via the Wind & Marine Energy Systems Centre for Doctoral Training, Grant EP/L016680/1.

DATA AVAILABILITY STATEMENT

The data used in this paper is subject to a non-disclosure agreement. It is therefore not available to share.

ORCID

Fraser Anderson https://orcid.org/0000-0001-9512-5209

REFERENCES

1. Seyr H, Muskulus M. Decision support models for operations and maintenance for offshore wind farms: a review. *Appl Sci*. 2019;9:278.
2. Reder MD, Gonzalez E, Melero JJ. Wind turbine failures—tackling current problems in failure data analysis. *J Phys Conf Ser*. 2016;753:072027.
3. Hahn B, Welte T, Faulstich S, Bangalore P, Boussion C, Harrison K. Recommended practices for wind farm data collection and reliability assessment for O&M optimization. *Energy Procedia*; 2017;137:358-365.
4. Seyr H, Muskulus M. Interaction of repair time distributions with a weather model. In: 29th International Congress on Condition Monitoring and Diagnostics Engineering Management (COMADEM 2016). COMADEM International; 2017.
5. Scheu MN, Kolios A, Fischer T, Brennan F. Influence of statistical uncertainty of component reliability estimations on offshore wind farm availability. *Reliab Eng Syst Saf*. 2017;168:28-39.
6. Dao CD, Kazemtabrizi B, Crabtree CJ. Offshore wind turbine reliability and operational simulation under uncertainties. *Wind Energy*. 2020;23(10): 1919-1938.
7. Scutari M. Bayesian network models for incomplete and dynamic data. *Statistica Neerlandica*. 2020;74(3):397-419. doi:10.1111/stan.12197
8. Mattei P-A. A parsimonious tour of Bayesian model uncertainty; 2019.
9. Wilson AG, Fronczyk KM. Bayesian reliability: combining information. *Qual Eng*. 2017;29(1):119-129. doi:10.1080/08982112.2016.1211889
10. Gonzalez E, Nanos EM, Seyr H, et al. Key performance indicators for wind farm operation and maintenance. *Energy Procedia*. 2017;137:559-570.
11. Artigao E, Martín-Martínez S, Honrubia-Escribano A, Gómez-Lázaro E. Wind turbine reliability: a comprehensive review towards effective condition monitoring development. *Appl Energy*. 2018;228:1569-1583.
12. Anderson F, Dawid R, García Cava D, McMillan D. Operational metrics for an offshore wind farm & their relation to turbine access restrictions and position in the array, Vol. 2018; 2021:012002.
13. Kaidis C, Uzunoglu B, Amoiralis F. Wind turbine reliability estimation for different assemblies and failure severity categories. *IET Renew Power Gener*. 2015;9(8):892-899.
14. Carroll J, McDonald A, McMillan D. Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*. 2019;19: 1107-1119.
15. Cevasco D, Kolios A, Koukoura S. Reliability, availability, maintainability data review for the identification of trends in offshore wind energy application. *Renew Sustain Energy Rev*. 2021;136:110414.
16. Pfaffel S, Faulstich S, Rohrig K. Performance and reliability of wind turbines: a review. *Energies*. 2017;10:1904.
17. SPARTA. Portfolio Review 2016/17; 2017.
18. Jiang Z, Hu W, Dong W, Gao Z, Ren Z. Structural reliability analysis of wind turbines: a review. *Energies*. 2017;10(12). doi:10.3390/en10122099
19. Slimacek V, Lindqvist B. Reliability of wind turbines modeled by a Poisson process with covariates, unobserved heterogeneity and seasonality. *Wind Energy*. 2016;19:1991-2002.

20. Stiesdal H, Madsen PH. Design for reliability. In: Proceedings of the Copenhagen Offshore Wind International Conference. CiteSeer Citeseer; 2005.
21. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions*, Vol. 2. John Wiley & Sons; 1995;289.
22. Rodríguez-Borbón MI, Rodríguez-Medina MA, Rodríguez-Picón LA, Alvarado-Iniesta A, Sha N. Reliability estimation for accelerated life tests based on a Cox proportional hazard model with error effect. *Qual Reliab Eng Int*. 2017;33(7):1407-1416. doi:10.1002/qre.2113
23. Emmert-Streib F, Dehmer M. Introduction to survival analysis in practice. *Machine Learn Knowl Extraction*. 2019;1(3):1013-1038. doi:10.3390/make1030058
24. Ozturk S, Fthenakis V, Faulstich S. Assessing the factors impacting on the reliability of wind turbines via survival analysis—a case study. *Energies*. 2018;11(11):3034.
25. Tavner P, Edwards C, Brinkman A, Spinato F. Influence of wind speed on wind turbine reliability. *Wind Eng*. 2006;30(1):55-72. doi:10.1260/030952406777641441
26. Tavner PJ, Gindele R, Faulstich S, Whittle G, Greenwood D. Study of weather and location effects on wind turbine failure rates. *Wind Energy*. 2010;16:175-187. doi:10.1002/we.538
27. Faulstich S, Lyding P, Tavner PJ. Effects of wind speed on wind turbine availability. In: Proceedings of the 2011 European Wind Energy Conference; 2011.
28. Wilson G, McMillan D. Assessing wind farm reliability using weather dependent failure rates, Vol. 524; 2014:012181.
29. Reder M, Yürüşen NY, Melero JJ. Data-driven learning framework for associating weather conditions and wind turbine failures. *Reliab Eng Syst Saf*. 2018;169:554-569.
30. Tavner P, Tavner PJ. *Offshore Wind Turbines Reliability, Availability and Maintenance*, Renewable Energy Series, vol. 13. Institution of Engineering and Technology; 2012.
31. Faulstich S, Lyding P, Hahn B. Component reliability ranking with respect to WT concept and external environmental conditions. Upwind Deliverable WP7.3; 2010.
32. Arabian-Hoseynabadi H, Tavner PJ, Oraee H. Reliability comparison of direct-drive and geared-drive wind turbine concepts. *Wind Energy*. 2010;13(1):62-73. doi:10.1002/we.357
33. Carroll J, McDonald A, Mcmillan D. Reliability comparison of wind turbines with DFIG and PMG drive trains. *IEEE Trans Energy Convers*. 2015;30:663-670. doi:10.1109/TEC.2014.2367243
34. Reder M, Melero JJ. A Bayesian approach for predicting wind turbine failures based on meteorological conditions. *J Phys Conf Ser*. 2018;1037(6):062003.
35. Carlos S, Sánchez A, Martorell S, Martín I. Onshore wind farms maintenance optimization using a stochastic model. *Math Comput Model*. 2013;57(7):1884-1890.
36. Zhong S, Pantelous AA, Beer M, Zhou J. Constrained non-linear multi-objective optimisation of preventive maintenance scheduling for offshore wind farms. *Mech Syst Signal Process*. 2018;104:347-369. doi:10.1016/j.ymssp.2017.10.035
37. Besnard F, Patriksson M, Strömberg A-B, Wojciechowski A, Fischer K, Bertling L. A stochastic model for opportunistic maintenance planning of offshore wind farms. In: 2011 IEEE Trondheim PowerTech. IEEE; 2011:1-8.
38. Byon E, Ntaimo L, Ding Y. Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Trans Reliab*. 2010;59(2):393-404.
39. Pattison D, Segovia Garcia M, Xie W, et al. Intelligent integrated maintenance for wind power generation. *Wind Energy*. 2016;19(3):547-562. doi:10.1002/we.1850
40. El-Thalji I, Liyanage JP. On the operation and maintenance practices of wind power asset: a status review and observations. *J Qual Maint Eng*. 2012;18(3):232-266.
41. McMillan D, Ault GW. Quantification of condition monitoring benefit for offshore wind turbines. *Wind Eng*. 2007;31(4):267-285. doi:10.1260/030952407783123060
42. May A, McMillan D, Thöns S. Economic analysis of condition monitoring systems for offshore wind turbine sub-systems. *IET Renew Power Gener*. 2015;9(8):900-907.
43. Ozturk S, Fthenakis V, Faulstich S. Failure modes, effects and criticality analysis for wind turbines considering climatic regions and comparing geared and direct drive wind turbines. *Energies*. 2018;11(9):2317.
44. Rinaldi G, Thies PR, Walker R, Johanning L. A decision support model to optimise the operation and maintenance strategies of an offshore renewable energy farm. *Ocean Eng*. 2017;145:250-262.
45. Abdollahzadeh H, Atashgar K, Abbasi M. Multi-objective opportunistic maintenance optimization of a wind farm considering limited number of maintenance groups. *Ocean Eng*. 2016;88:247-261.
46. Cox DR. Regression models and life-tables. *J R Stat Soc Series B (Methodol)*. 1972;34(2):187-220.
47. Perperoglou A. Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Stat Med*. 2014;33(1):170-180.
48. Andrinopoulou E-R, Eilers PHC, Takkenberg JJM, Rizopoulos D. Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*. 2018;74(2):685-693.
49. Gao D, Grunwald GK, Rumsfeld JS, Schooley L, MacKenzie T, Shroyer ALW. Time-varying risk factors for long-term mortality after coronary artery bypass graft surgery. *Ann Thorac Surg*. 2006;81(3):793-799. doi:10.1016/j.athoracsur.2005.08.005
50. Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press; 1993.
51. Prasad A, Manmohan A, Shanmugam PK, Kothari DP. Application of cubic spline interpolation technique in power systems: a review. In: Truong YK-N, Sarfraz M, eds. *Topics in Aplines and Applications*. IntechOpen; 2018.
52. Djeundje VB, Crook J. Dynamic survival models with varying coefficients for credit risks. *Eur J Oper Res*. 2019;275(1):319-333.
53. Eilers PH, Marx BD. Flexible smoothing with b-splines and penalties. *Stat Sci*. 1996;11(2):89-102.
54. Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. *Technometrics*. 1989;31(1):3-21.
55. Wienke A. *Frailty Models in Survival Analysis*. Chapman & Hall/CRC; 2010.
56. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2014;44(1):324-333.
57. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. *J Multivar Anal*. 2009;100(9):1989-2001.

58. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall/CBC; 2013.
59. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014;15(1): 1593-1623.
60. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2016;27(5): 1413-1432.
61. Sivula T, Magnusson M, Matamoros AA, Vehtari A. Uncertainty in Bayesian leave-one-out cross-validation based model comparison; 2020.
62. Geisser S. The predictive sample reuse method with applications. *J Am Stat Assoc*. 1975;70(350):320-328. doi:10.1080/01621459.1975.10479865
63. Hofner B, Kneib T, Hartl W, Küchenhoff H. Building Cox-type structured hazard regression models with time-varying effects. *Stat Model*. 2011;11(1): 3-24. doi:10.1177/1471082X1001100102
64. Anderson F, McMillan D, Dawid R, Garcia Cava D. A Bayesian hierarchical assessment of night shift working for offshore wind farms. *Wind Energy*. 2023;26(4):402-421.
65. Spinato F, Tavner PJ, van Bussel G, Koutoulakos E. Reliability of wind turbine subassemblies. *IET Renew Power Gener*. 2009;3(4):387-401.
66. Hawker G, Mcmillan D. The impact of maintenance contract arrangements on the yield of offshore wind power plants. *Proc Inst Mech Eng Part O: J Risk Reliab*. 2015;229(5):394-402. doi:10.1177/1748006X15594693
67. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481. doi:10.1080/01621459.1958.10501452
68. Congon P. *Bayesian Statistical Modelling*. John Wiley & Sons; 2006.
69. Leahy K, Gallagher C, O'Donovan P, O'Sullivan DTJ. Issues with data quality for wind turbine condition monitoring and reliability analyses. *Ocean Eng*. 2020;209:107381.
70. Dinwoodie I, Endrerud OE, Hofmann M, Martin R, Sperstad I. Reference cases for verification of operation and maintenance simulation models for offshore wind farms. *Wind Engineering*. 2015;39(1):1-14.

How to cite this article: Anderson F, Dawid R, McMillan D, García-Cava D. A Bayesian reliability analysis exploring the effect of scheduled maintenance on wind turbine time to failure. *Wind Energy*. 2023;1-21. doi:10.1002/we.2846