Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Data-Driven Infrastructure Planning for Offshore Wind Farms

To cite this article: Isha Saxena *et al* 2024 *J. Phys.: Conf. Ser.* **2767** 062002

Journal of Physics: Conference Series **2767** (2024) 062002 doi:10.1088/1742-6596/2767/6/062002

Data-Driven Infrastructure Planning for Offshore Wind Farms

Isha Saxena, Behzad Kazemtabrizi, Matthias C.M. Troffaes, and Christopher J. Crabtree

Durham University, Durham, DH1 3LE, UK

E-mail: isha.saxena@durham.ac.uk

Abstract. Offshore wind farms are one of the major renewable energy resources that can help the UK to reach its net zero target. Under the 10 point plan of the green revolution, the UK is set to quadruple its wind energy production by increasing its offshore wind capacity to 40GW by 2030 [1]. Research needs to be conducted to study the failure and repair processes of wind turbines under various conditions as the current models make a simplifying assumption that the failure/repair rate remains constant over time. This research aims to create a more accurate model using SCADA data. In this research, different mathematical models are fitted to the time to failure and time to repair data of wind turbine components using frequentist methods (such as Maximum Likelihood Estimation) and Bayesian methods. Further analysis will be conducted using complex system analysis considering the failures of each electrical and mechanical component of the wind turbine. The aim of this project is to perform a more accurate reliability analysis that can help to further drive down costs of wind energy by potentially reducing the downtimes of the wind turbines.

1. Introduction

In 2022, the UK's offshore wind farms demonstrated their potential by supplying electricity to 11.5 million homes [2]. It was buoyed by government incentives like the Contract for Difference (CfD) scheme [2]. Despite operational and maintenance challenges, offshore wind remains pivotal in the UK's journey towards achieving net-zero emissions. The decreasing Levelised Cost of Energy (LCOE) signals its growing cost-effectiveness and investment attractiveness. For wind turbines to be economically viable, they must be reliable in order to provide clean, renewable energy. In this context, reliability refers to the probability of a system performing its expected task adequately for a designated period of time under specific operating conditions [3]. The performance and maintenance needs of wind turbines are heavily influenced by their times to failure and repair. Therefore, to estimate the total reliability of the wind turbine, it is crucial to comprehend the statistical characteristics of the time to failure of these components.

Currently, there are various tools available for the determination of unavailability of offshore wind farms in the early stages of a wind farm's life. The models utilised in these tools are tailored towards simplifying assumptions namely, they consider that all the turbine components to be repairable and the times to failure and repair exponentially distributed, i.e. they have density

$$f(t) = \frac{P(t \leq T \leq t + dt)}{dt} = \theta e^{-\theta t} \quad (1)$$

where θ is the parameter of the distribution (different for repair and failure), representing the rate. Note that if the rate is θ, then the mean of the distribution under these assumptions is $1/\theta$ [4, 5]. Some papers also take a different approach towards these calculations.

The exponential distribution assumption may lead to sub-optimal results when it comes to analysing turbine failures and repairs as random processes. This is because it assumes constant rates of failure and repair, which may not necessarily be true [6]. It effectively makes simplifying assumptions which can lead to a poor estimation of the Expected Energy Not Served (EENS). For this reason, some researchers have started adopting other types of distributions when analysing failure/repair times for wind turbines. For example, [7] uses two and three-parameter Weibull distributions and the parameters for these are calculated by using least squares and Maximum Likelihood Estimation (MLE). [8] also uses a modified version of the Weibull distribution for the reliability study. The references mentioned above highlight the necessity for a more robust uncertainty quantification process.

This research aims to improve the uncertainty quantification of failure and repair processes and, therefore, also of indicators such as EENS and LCOE. In the long term, this research can lead to more reliable maintenance schedules, resulting in cost reduction and better decision-making for investors. In the short term, it can aid in the planning and execution of maintenance activities. Consequently, this project aims to improve the estimation and uncertainty quantification of the costs at the beginning of construction of the wind farm. To this end, this paper aims to analyse a part of Supervisory Control and Data Acquisition (SCADA) data pertaining to a fleet of wind turbines taken over several years [9], and to frame a statistical model that more accurately models failures and repair times of wind turbines.

2. Background

2.1. Wind Turbine Models for Reliability Analysis

The overall lifetime of a wind turbine can be divided into three stages: early life failures region, steady state region, and wear-out region. This can be seen in Figure 1 which is commonly referred to as the bathtub curve. The bathtub curve shows how failure rate changes with the life of the wind turbine. The early life failure region is at the beginning of the life of the turbine. At this stage, the failures are initially high with a time-dependent failure rate which decreases in time until a constant failure rate is obtained in the steady-state region. Lastly, the failure rate increases towards the end of the lifetime of the wind turbine [5].

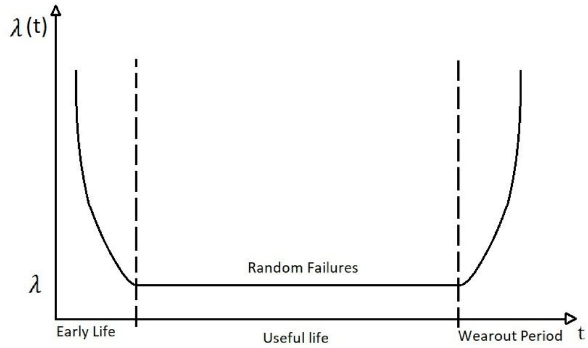

Figure 1. Bathtub curve

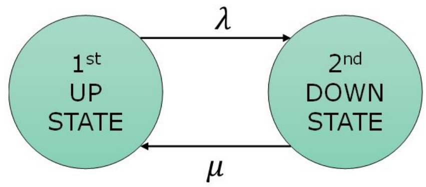

Figure 2. Two state Markov chain

The current models of reliability focus on the steady state region of the bathtub curve. A constant failure rate means that the failure time probability density function is exponential (see Eq. (1)). The same can be said for repair processes. In this case, failure and repair times are assumed to be distributed exponentially with a constant, albeit different parameters referred to as the failure and repair rate respectively. Consequently, the random failure/repair processes of a component with constant failure and repair rates can be modelled as a two-state Markov process as shown in Figure 2.

Extending the Markov assumption to a complex system like wind turbines with multiple components, reliability block diagrams are employed to assess their reliability by considering different failure scenarios and component interdependencies. These methods rely on the Markov assumption for modeling the system's failure and repair processes as multi-state Markov processes, with states representing failure and normal operation, assuming constant transition rates between these states [5].

In this paper, the aim is to relax the Markov assumption of having constant transition rates between different states as shown in Figure 2 when modelling the failure and repair processes of a complex system such as a wind turbine. The methodology by which the state transition rates are estimated is outlined in the Methodology section of this paper.

3. Methodology

The project's methodology adopts a comprehensive and advanced approach towards modeling the failure and repair distributions with the help of SCADA data. The purpose of this modelling is enhancing the reliability estimation and failure prediction of wind turbines by developing more accurate models for failure and repairs. These improved models will lead to more precise reliability assessments, ultimately impacting Operational Expenditure (OPEX) and the LCOE in the long run. It includes parameter estimation for the Probability Density Function (PDF) of times to failure distribution using Bayesian Parameter estimation and MLE. MLE is a comparatively simple method of parameter estimation whereas Bayesian statistics provides a full posterior distribution of the parameters, allowing for the quantification of parameter uncertainty. This approach allows for a deeper understanding of the failure and repair processes in repairable wind turbine components. The methodology is laid out below in a structured manner for clarity and coherence:

- DATA: SCADA data from wind turbines is used for the analysis. This SCADA data consists of the SCADA signals such as generator rotational speed, average wind speed, total active power, etc.
- PROCESSING: This data is processed alongside the wind turbine logs to determine the periods of availability and unavailability. The values of interest are time to repair and time to failure that are calculated once we know the duration of availability of the wind turbine.
- MODEL: A new statistical model characterising the times to failure for each turbine is prepared to understand the impact of the environment on the wind turbines' failures. The parameter evaluation for this model (e.g., failure rate, mean time to failure) is carried out by using either MLE or Bayesian inference.

3.1. Data Source

Our primary dataset is sourced from EDP Renewables [9], encompassing SCADA and log data from a wind farm of 16 turbines, each rated at 2MW, with a focus on a subset of 5 turbines. The data provides insights into the relationship between environmental conditions and turbine operational parameters, such as rotor speed, temperature, and net power production [9].

According to literature, it is initially assumed that the wind turbines experience an average of 8.5 failures per year with a mean downtime of 10 days per failure. These approximations are instrumental in shaping our prior parameter calculations and validating the models developed [10].

3.2. Data Analysis

The aim of the data analysis is to get failure and repair times from wind turbine observations. The data that is generally available comprises power, temperature, and wind speeds, rather than failure and repair times. Therefore, it is necessary to process the data first to convert it into sequences of times to failure and times to repair. This processed data will be used for a statistical analysis.

Interpreting when failures happen is inherently subjective to a certain degree, and this is reflected in our approach. The available data does not explicitly provide information on failure and repair times; however, it includes variables like power production, power curve of the wind turbine, and wind speed with corresponding timestamps. Using this data, we derive the times to failure and times to repair for both wind turbines.

For this, at first, a heuristic criteria is established to determine the turbine's operational status based on wind speed and power production. A turbine is classified as failed if the wind speed is high and the corresponding power production falls below 30% of the expected active power output. Conversely, a turbine is deemed to be in an unknown state if the wind speed is below cut-in or above cut-out and there is no power production. Otherwise, when the wind speed is high enough and the power production aligns closely with the power curve, then the wind turbine is considered to be in an available state.

This analysis results in 36 data points. This indicates 36 repairs in a year. These repairs could have been carried out for small or large failures. These times to failures are obtained in the unit of minutes and then converted to days.

3.3. Model Development

Let $T_1, \ldots, T_n$ be i.i.d. random variables representing the times to failure of a wind turbine, with values $t_1, \ldots, t_n$, each distributed according to the same density $P(t_i \mid \theta)$, where θ is an unknown model parameter. The prior density function of the model parameter θ is denoted by $P(\theta)$ and the posterior density as $P(\theta \mid t)$, where t denotes $(t_1, \ldots, t_n)$. To estimate the parameter θ, Maximum Likelihood Estimation and Bayesian estimation are used. The joint PDF for these n random variables, conditional on the parameter θ, is

$$P(t|\theta) = \prod_{i=1}^{n} P(t_i|\theta) \tag{2}$$

Two primary models are considered in this paper.

- **Model 1 (Weibull Model)**: In this model, it is assumed that failure times follow a Weibull distribution, therefore,

$$T_i \mid \lambda, k \sim \text{Weibull}(\lambda, k) \tag{3}$$

- **Model 2 (Exponential Model)**: This model assumes that failure times follow an exponential distribution, which is a special case of a Weibull distribution with $k = 1$. This is done to simplify the comparison of the parameter evaluation in both cases. Therefore,

$$T_i \mid \lambda \sim \text{Weibull}(\lambda, 1) \tag{4}$$

The PDF of the two-parameter Weibull distribution used here is:

$$P(t_i \mid \lambda, k) := \frac{kt^{k-1}}{\lambda^k} \exp\left[-\left(\frac{t}{\lambda}\right)^k\right] \tag{5}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. In this scenario, since the dataset is very small, the selected prior will highly impact the posterior. An inverse gamma function is selected to be a prior for a Weibull distribution as a likelihood function. This function belongs to a family of two-parameter distributions. We choose the prior distributions as follows:

- The scale parameter λ is modelled with an Inverse Gamma distribution, denoted as InverseGamma(α, β). Here, α denotes the shape parameter and β is the scale parameter of the prior. The mean of the prior for λ should approximately be equal to the prior expected failure time if k does not deviate too much from 1.
- The shape parameter k is a priori assumed to follow a Uniform distribution, expressed as Uniform(κ_*, κ^*).The mean of the prior for lambda should approximately be equal to the prior expected failure time if k does not deviate too much from 1.

Here, a two-parameter prior ensures a more accurate calculation, as it allows more flexibility and specificity.

Maximum Likelihood Estimation
Maximum Likelihood Estimation offers a distinct method for parameter estimation. The MLE of θ is determined by maximising the likelihood function $P(t \mid \theta)$, a process which can be efficiently computed using the log-likelihood function [11]. Mathematically, this is expressed as:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log P(t_i|\theta) \tag{6}$$

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta) \tag{7}$$

Bayesian Statistics for Parameter Estimation
In our analytical framework, Bayesian statistics play a crucial role in conducting a parameter estimation. The methodology involves selecting a prior distribution $P(\theta)$, which includes our initial understanding or hypothesis about the parameter θ. The likelihood function $P(t \mid \theta)$ quantifies the probability of observing the data given a particular value of the parameter. Through the application of Bayes's theorem, this prior belief is updated to a posterior distribution $P(\theta \mid t)$, effectively synthesising our initial knowledge with the empirical data. According to Bayes's theorem:

$$P(\theta \mid t) = \frac{P(t|\theta)P(\theta)}{P(t)} = \frac{P(t \mid \theta)P(\theta)}{\int P(t \mid \theta')P(\theta')d\theta'} \tag{8}$$

where $P(t)$ represents the marginal joint probability density function. To extract samples from this posterior distribution, we employ advanced computational techniques such as Markov Chain Monte Carlo (MCMC), which are integral for managing the complexities inherent in Bayesian analysis [11]. Some other techniques could be used here as well, such as, Sequential Monte Carlo and variational inference. All these techniques have their own pros and cons as compared to a conventional Bayesian statistical method. Bayesian statistical model provides an exact posterior distribution whereas, Sequential Monte Carlo and Variational Inference provide approximate values.

Markov Chain Monte Carlo in Bayesian Analysis Within the Bayesian modelling framework used in this paper, MCMC algorithms are useful for effectively sampling from the posterior distribution for computationally complex calculations, particularly those involving the integration of prior distributions and the likelihood derived from data.

The focus of this paper is the shape k and scale parameters λ of the Weibull distribution, treating them as independent variables a priori. This assumption simplifies the complexity of prior specification. These distributions reflect our initial hypotheses about the likely ranges and behaviours of the parameters. The test data visualised in our study, alongside these specified priors, forms the foundation for the Bayesian model.

MCMC provides a full distribution of model parameters, rather than single-point estimates, thereby offering a detailed perspective on parameter uncertainty suitable for research analysis.

The major difference between MLE and Bayesian estimation is that MLE provides a point estimate, which is an approximate single value that maximizes the likelihood function, whereas Bayesian estimation provides a probability distribution, known as the posterior distribution, which represents the uncertainty in the parameter estimates [11].

4. Hyperparameter Selection

The process of selecting hyperparameters relies on prior knowledge. Once the likelihood is specified, the next step involves assigning priors for the parameters of interest, in this case, k and λ. We assume that k and λ are a priori independent. To choose suitable hyperparameters, the prior knowledge of the wind turbine failure rate obtained from the literature is utilised. Specifically, the mean time to failure is used as a reference point for selecting the shape and scale parameters of the prior distribution. This reference point helps inform the choice and ensures that the prior distribution aligns with the expectations regarding the underlying process. Once initial hyperparameters are chosen, they are iteratively adjusted until a reasonable prior predictive distribution is achieved. This iterative process involves evaluating the prior predictive distribution and comparing it with historical data or expert knowledge. Adjustments are made to strike a balance between incorporating prior information and allowing flexibility for the data to influence the posterior inference. Figures 3 and 4 provide a visual representation of the prior predictive distributions of both Weibull and exponential distributions. The simulated data helps with the model validation by comparing the model's performance against historical data and expert knowledge. The value of shape parameter for Weibull function for this analysis is set to 2.5 and scale parameter is 56.25.

Overall, the process of selecting hyperparameters for the prior distribution for this paper involved a thoughtful consideration of prior knowledge, iterative adjustment, and validation against historical data. This approach ensured that the prior distribution effectively captures the beliefs about the parameters of interest and facilitates Bayesian inference.

5. Results

The results of processing the data are visualised in Figure 5. The data points are visually represented with colour coding. Here, blue is assigned to instances of low wind speed and low power production or an unknown state, while red represents failed states, and green is used to indicate an available state.

Next, to convert the unknown states into known states, we assume that if the state was available previously, the turbine remains available, and similarly, if the turbine was failed, it remains failed. This assumption allows us to construct a time series for the wind turbine that contains no unknown states, which we can then finally turn into a series of failure and repair times, as shown in Table 6.

This analytical technique and visual representation allows for a detailed examination of the wind turbine performance and state, facilitating a comprehensive understanding of the system dynamics and behavior. The low wind speed instances will be used to extract time series data of operating and non-operating states of the wind turbines. This data will then be used to calculate times to failure and times to repair of the wind turbine. These times to failure, and times to repair will then be used as inputs for the parameter estimation of a probability density function, initially using the MLE, as shown above in the methodology.

Journal of Physics: Conference Series **2767** (2024) 062002 doi:10.1088/1742-6596/2767/6/062002

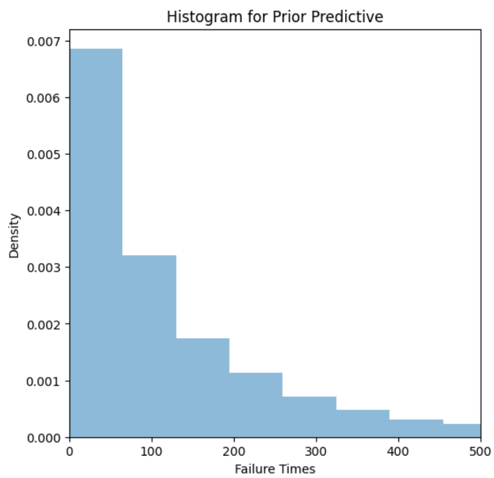

Figure 3. Simulated prior predictive data with an exponential likelihood function

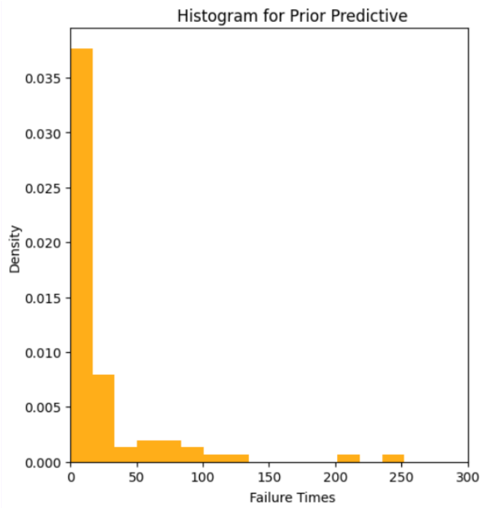

Figure 4. Simulated prior predictive data with a Weibull likelihood function

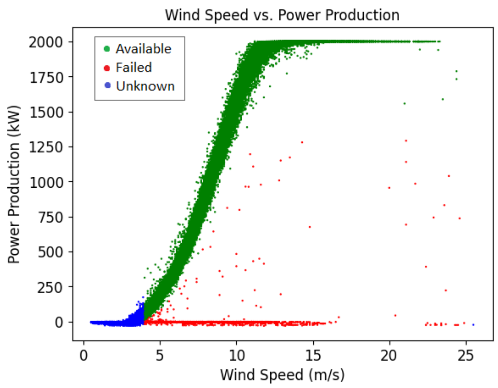

Figure 5. Power Curve for Turbine 07 for 2017

Figure 6. Determining unknown states for the given data

Time	State	Final State
0:00:00:00	1	1
0:00:00:10	1	1
0:00:00:20	?	1
0:00:00:30	?	1
0:00:00:40	1	1
...	...	...
5:10:00:20	0	0
5:10:00:30	?	0
5:10:00:40	?	0
5:10:00:50	1	1

When the shape parameter is set to 1, the value of scale parameter is estimated to be 21.7 by using Bayesian estimation (Figure 8) and 21.03 by using MLE, which is an exponential curve (Figure 7). If k is not fixed, the value of k is obtained to be 0.872 and the value of λ is 19.71 using MLE (Figure 7) whereas the values of k is 0.84 and λ

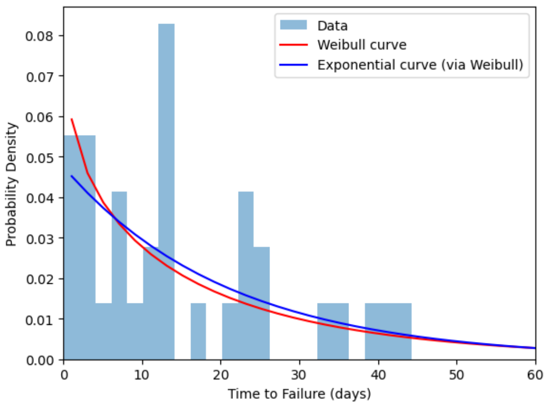

Figure 7. Exponential and two parameter Weibull distribution parameter estimation using Maximum Likelihood Estimation

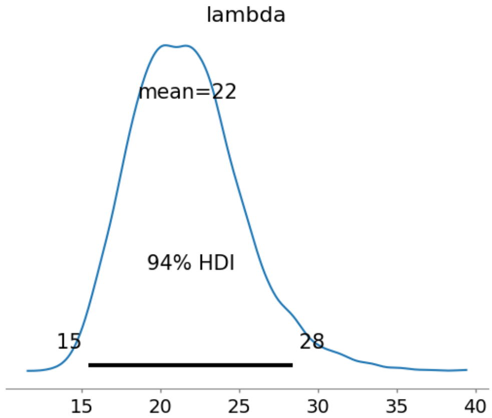

Figure 8. Parameter estimation for θ in case of an exponential curve

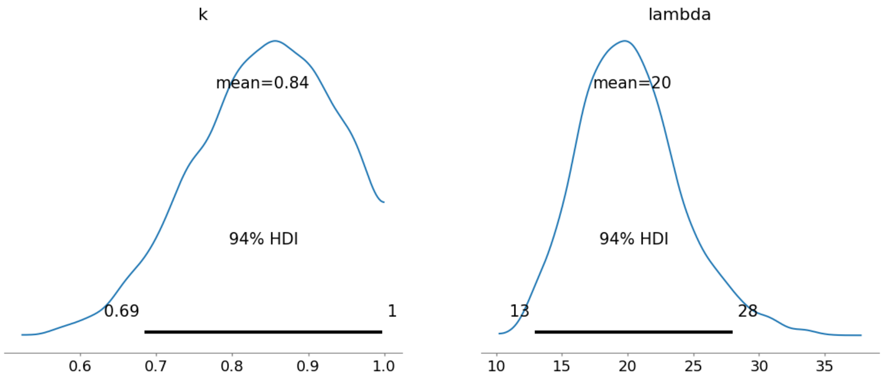

Figure 9. Parameter estimation for k and λ

is 20.31 by conducting Bayesian analysis (Figure 9). Here, the value of k in case of a two parameter Weibull function is close to 1, hence, both the Weibull and exponential distributions are close to each other.

6. Conclusion

This research focuses on determining a more generalised statistical model pertaining to times to failure (and repair) in offshore wind turbines, which in turn could lead to better estimations of impacts of failure/repairs on measures such as Expected Energy Not Served or Levelised Cost of Energy for offshore wind farms. Moreover, these models are used to predict the failures and repairs of wind turbine components and to determine any correlations between the failures and repairs and their wider environmental conditions. This newly developed model can be used to simulate the operation of exemplar wind farms.

The results indicate that both Bayesian inference and MLE provide estimates for the parameters of the statistical models pertaining to failure and repair times of components of the turbine. The estimated parameters differed slightly between the two methods, with Bayesian analysis offering a more robust estimation considering parameter uncertainty.

In this paper, we have only focused on uncertainty quantification of the model parameters, and not on prediction of future failures and repairs. In future work, we plan to explore other numerical methods for efficient prediction and inference based on the Bayesian models developed in this paper, such as for instance variational Bayesian methods, and approximate Bayesian computation, potentially in combination with sequential Monte Carlo.

References

[1] gov.uk. *The Ten Point Plan for a Green Industrial Revolution*. URL: `https://www.gov.uk/government/publications/the-ten-point-plan-for-a-green-industrial-revolution/title`.

[2] T. Badouard et al. *Study on energy costs, taxes and the impact of government interventions on investments in the energy*. Trinomics, July 2020.

[3] Ernst G. Frankel. *Systems reliability and risk analysis*. 2nd, rev. ed. Engineering applications of systems reliability and risk analysis. Dordrecht ; Boston: Kluwer Academic Publishers, 1988. 429 pp. ISBN: 9789024736652.

[4] Robert Kissell and Jim Poserina. *Chapter 4 - Advanced Math and Statistics*. Available online 14 April 2017, Version of Record 14 April 2017. 2017.

[5] R. Billinton and R N Allan. *Reliability evaluation of engineering systems: concepts and techniques*. 2nd ed. New York: Plenum Press, 1992. 453 pp. ISBN: 9780306440632.

[6] J. Peter Poletto. "An Alternative to the Exponential and Weibull Reliability Models". In: *IEEE Access* 10 (Nov. 2022), pp. 118759–118778. DOI: `10.1109/ACCESS.2022.3219426`.

[7] Cuong Dao, Behzad Kazemtabrizi, and Christopher Crabtree. "Wind turbine reliability data review and impacts on levelised cost of energy". In: *Wind Energy* 22 (12 Dec. 2019), pp. 1848–1871. ISSN: 1095-4244. DOI: `10.1002/we.2404`.

[8] M Xie and CD Lai. "Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function". In: *Reliability Engineering and System Safety* 52 (1 Apr. 1996), pp. 87–93. DOI: `https://doi.org/10.1016/0951-8320(95)00149-2`.

[9] *Data*. en. `https://opendata.edp.com/open-data/en/data.html`. Accessed: 2023-6-2.

[10] James Carroll, Alasdair McDonald, and David McMillan. "Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines". In: *Wind Energy* 19.6 (2015), pp. 1107–1119. DOI: `10.1002/we.1887`.

[11] M. H DeGroot and M. J Schervish. *Probability and statistics*. Pearson Education, 2012.