Uncertainty quantification of wind turbine fatigue lifetime predictions through binning

Sadeghi, N.; Noppe, N.; Morato, P.G.; Weijtjens, W.; Devriendt, C.

DOI
10.1088/1742-6596/2767/3/032024

Publication date
2024

Document Version
Final published version

Published in
Journal of Physics: Conference Series

Citation (APA)
Sadeghi, N., Noppe, N., Morato, P. G., Weijtjens, W., & Devriendt, C. (2024). Uncertainty quantification of wind turbine fatigue lifetime predictions through binning. *Journal of Physics: Conference Series*, *2767*, Article 032024. https://doi.org/10.1088/1742-6596/2767/3/032024

Uncertainty quantification of wind turbine fatigue lifetime predictions through binning

N. Sadeghi[1], N. Noppe[1,2], P. G. Morato[3], W. Weijtjens[1], C. Devriendt[1]

[1] OWI-Lab, AVRG, Vrije Universiteit Brussel (VUB), Brussels, 1050, Belgium
[2] 24SEA, Brussels, 1000, Belgium
[3] Faculty of Architecture and the Built Environment, Delft University of Technology (TU Delft), Delft, Netherlands

E-mail: negin.sadeghi@vub.be

Abstract. Aging wind energy assets demand the development of methods able to effectively support informed decision-making. These needs have inspired the use of data-driven methodologies, which offer valuable insights to wind turbine owners and/or operators. Many approaches can be found in the literature for extrapolating fatigue damage measurements to estimate the lifetime of wind turbines. In some cases, resampling approaches are proposed to compute the confidence levels associated with the generated projections, yet a standardized framework has not been adopted. Most reported studies identify the relationship between short-term damage and long-term Environmental and Operational Conditions (EOCs) by mainly rendering mean lifetime predictions and their associated confidence levels, whereas additional predicted lifetime statistical information is usually overlooked. In this work, we showcase the importance of properly accounting for the variability in lifetime predictions, describe how to summarize binned damages using statistical estimators and investigate bootstrapping variants for computing the confidence levels in the generated damage estimators.

1. Introduction

Addressing the restrictions imposed by limited available sensor data, existing methods in the literature resort to additional long-term information provided by readily available Supervisory Control and Data Acquisition (SCADA) systems. For instance, [1, 2] identify the relationship between measured damages and concurrently monitored environmental conditions (e.g., wind speed and turbulence intensity) by binning damage measurements into environmental parameters and for only a specific operational condition. Most studies only report the predicted mean value of the asset's lifetime, whereas [3] conservatively suggests relying on the 90th percentile from the estimated lifetime statistics. While quantifying the lifetime's mean value is certainly informative, the often high variability associated with the generated predictions tends to be overlooked. Such an omission ultimately precludes the comprehensive understanding of potential outcomes, specifically with regard to the minimum and maximum projected lifetimes. As shown by [4], fatigue damage measurements that are binned based on specific environmental and operational conditions (EOCs) may exhibit high variability. Explicitly quantifying this variability becomes even more crucial in cases where more sophisticated binning schemes (e.g. high dimensional) cannot be implemented due to, for instance, SCADA parameters' unavailability. With a slightly different application focus, [5] and [6] are among the few studies that explicitly account for damage variability within each bin. Furthermore, some studies [1, 2, 3, 5] rely on resampling methods to quantify confidence levels in the predicted lifetime mean value. By applying bootstrapping on already clustered data, some of these methods deviate from principled bootstrapping methods, which typically involve resampling from a sample representative of the entire population [7].

In this work, we establish the relationship between limited SHM damage data and its corresponding EOCs by binning short-term damage into representative EOC parameters, e.g., wind speed, while restricting the analysis to operational data. More specifically, we introduce the required formulation and quantify the uncertainty associated with fatigue damage predicted through binning by: (i) describing how damage scatter in bins can be encapsulated into estimators, and (ii) investigating bootstrapping variants for computing confidence levels from damage estimators in order to account for statistical uncertainty. Note that the focus of this study is on deriving 10-minute damage statistical information from short-term monitoring campaigns, and contrasting the resulting damage distributions against those from extended measurement periods, without explicitly addressing long-term fatigue damage extrapolation.

2. Predicting wind turbine fatigue damage through binning

To enable the computation of fatigue damage predictions based on information collected over a short-term monitoring campaign, binning fatigue damage into predefined EOC categories is a widely adopted solution within the wind energy community. Once the relationship between EOCs and fatigue damage is established, one can simply rely on recorded or forecasted EOCs for yielding fatigue damage predictions over periods when damage monitoring is no longer available. This relationship can be achieved by binning monitored damage information into EOC bins. However, it is generally impractical to include an extensive set of relevant EOC parameters as this may result in numerous empty bins and/or a limited number of damage measurements available for some EOC bins. In practical scenarios, fatigue damage is normally binned into a reduced subset of EOC parameters, thereby leading to potential fatigue damage variability within each considered bin. This variability is potentially caused by the omission of relevant EOC categories. The majority of existing studies typically assign either the mean or the mode as the measure of fatigue damage for each bin [1, 2, 3, 8], thus disregarding useful information for estimating the uncertainty associated with fatigue damage predictions. We instead propose a probabilistic binning approach in Section 3 that naturally accounts for fatigue damage inherent variability within each bin.

As mentioned previously, fatigue damage can be predicted over a certain period by relying on binned measurements and collected or estimated EOC information. The probability of occurrence of each EOC bin is either estimated from design/theoretical probability distributions or a mass probability function based on long-term collected SCADA information. More specifically, the probability of occurrence of each EOC bin, $p(x_i)$, is computed as the area under the probability density function (PDF) delimited by a specified interval:

$$p(x_i) = \Pr\left(x_i^{(l)} < X < x_i^{(u)}\right) = \int_{x_i^{(l)}}^{x_i^{(u)}} f(x)\, dx\,, \tag{1}$$

where $x_i^{(l)}$ and $x_i^{(u)}$ are the lower and upper bounds of the ith bin, and $f(x)$ stands for the PDF over a defined EOC, X. Alternatively, one can compute the frequency of occurrence associated with a certain bin through a normalized histogram obtained from a discrete number of samples as, $p(x_i) = n_i/n_t$, where n_i stands for the number of samples falling into the ith bin and n_t denotes the total number of samples.

At the prediction stage, the fatigue damage over a certain period can be estimated by combining the clustered fatigue damage, d_i, and the probability of occurrence, $p(x_i)$, for each ith EOC bin, as:

$$d = \sum_{i=1}^{n_b} p(x_i) d_i\,, \tag{2}$$

where n_b stands for the total number of bins. To better understand the predictive capabilities of this fatigue damage model, let us consider a simple scenario where a wind turbine is operational 70% over a reasonably long-term period, whereas the turbine is parked 30% over the same period. If during a short-term monitoring campaign the operational-parked distribution, however, corresponds to 50% for each condition, the long-term EOC's frequency of occurrence, i.e., (70%, 30%), should be considered when computing long-term damage.

3. Quantifying the uncertainty of fatigue damage predictions

As explained in the previous section, the fatigue damage for a given EOC bin, d_i, is often modeled as the mean or mode of clustered measurements. Instead, the fatigue damage variability within each bin can be modeled as a probability distribution or via statistical estimators. If the collected dataset used for binning contains a significant number of measurements, the damage variability within each bin can be categorized as aleatory uncertainty. This means that by collecting additional data, the uncertainty cannot be further reduced. The statistical estimators computed for characterizing the damage distribution for each bin, e.g. approximated mean $\hat{\mu}_{d_i}$ and variance $\hat{\sigma}^2_{d_i}$, might not be very representative if the dataset does not contain sufficient samples. In that case, the estimators themselves are uncertain, yet this uncertainty can be classified as epistemic, particularly statistical in this case. By collecting additional data, the statistical uncertainty can be reduced, thereby resulting in more accurate estimators.

Hereafter, we describe: (i) the necessary formulation for combining binned statistical estimators with EOCs' frequency of occurrence, and (ii) a bootstrapping approach for estimating the statistical uncertainty associated with statistical estimators.

3.1. Combining binned statistical estimators

The binned fatigue damage within each bin is here modeled as a random variable, D_i, whose variability stems from omitted EOC parameters and/or other intrinsic random phenomena. D_i can be described through statistical estimators, e.g. mean, variance, or density estimators such as a histogram, which ultimately represent a subset of the underlying population. For a defined period, collected fatigue damage measurements are clustered into EOC bins, and statistical estimators, $\hat{\theta}_{d_i}$ can then be computed for each respective ith bin. Once the binned estimators are calculated, the combined fatigue damage estimator, $\hat{\theta}_d$, of the combined damage random variable, D, can be straightforwardly computed as:

$$\hat{\theta}_d = \sum_{i=1}^{n_b} p(x_i) \cdot \hat{\theta}_{d_i}, \tag{3}$$

where $p(x_i)$ corresponds to the probability of occurrence of bin i, and $\hat{\theta}_{d_i}$ stands for the ith binned damage estimator. This process can be followed to compute a wide range of statistical estimators, from the combined mean damage, $\hat{\mu}_d$, and discrete mass function values to events, e.g. probability that a certain value is exceeded. The combined central variance, $\hat{\sigma}_d^2$, can also be determined directly through a closed-form solution:

$$\hat{\sigma}_d^2 = \sum_{i=1}^{n_b} p(x_i) \left[\hat{\sigma}_{d_i}^2 + (\hat{\mu}_{d_i} - \hat{\mu}_d)^2\right], \tag{4}$$

where $\hat{\sigma}_{d_i}^2$ stands for the fatigue damage variance corresponding to the ith bin. By following this process, the resulting combined variance is naturally equal to the variance estimated directly from unbinned fatigue damage measurements for a given monitoring period, and, in contrast to other works [5], a zero covariance between bins does not have to be assumed.

3.2. Bootstrapping

As explained previously, the computed estimators are subject to statistical uncertainty arising from a limited number of collected fatigue damage samples. For instance, two short-term monitoring campaigns may yield different damage estimators. By extension, binned damages may also vary for each measurement set. Statistical uncertainty can be reduced by collecting more samples, i.e., over a longer measurement period. Obtaining measurements throughout the turbine's entire lifespan is impractical. Consequently, inference about the overall performance must be drawn from the data collected within a specific timeframe. From a frequentist perspective, the degree of confidence over an estimate for a population parameter can be expressed as a confidence interval. The interpretation of confidence intervals should be carefully understood to avoid potential

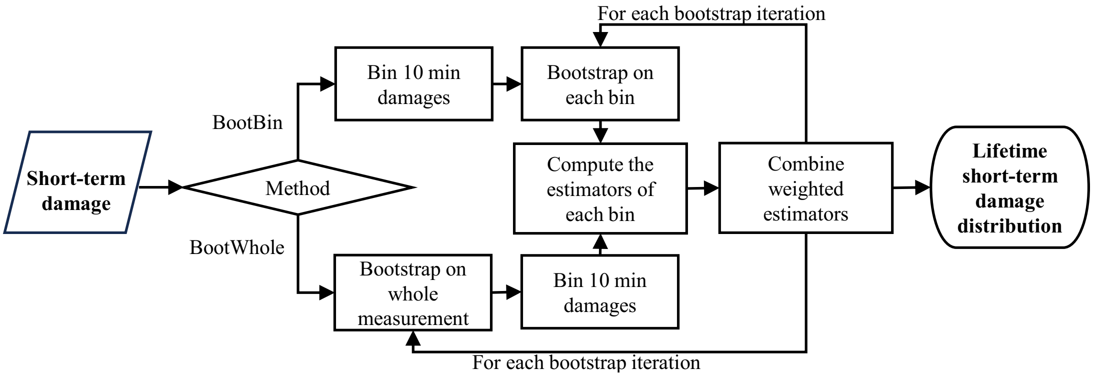

Figure 1: Flowchart representation of the investigated bootstrapping variants.

misconceptions [9]. Following a frequentist view, a 95% Confidence Interval (CI) can be interpreted as the interval that contains the true value in 95 out of 100 repeated experiments. Logically, the width of a CI is influenced by the intrinsic variation within the target population, the size of the sample, and the desired level of confidence.

Sample-based bootstrapping with replacement can be introduced to the measurement-based probabilistic model to estimate the confidence over damage estimators $\hat{\theta}_d$. Bootstrapping consists of sampling with replacement over a defined number of iterations, out of which, certain parameters of interest can be estimated, e.g. mean, variance, confidence intervals, among others. The bounds of the $(1-\alpha)\%$ confidence interval, which are the $(\alpha/2)$-th percentile and the $(1-\alpha/2)$-th percentile, are used to estimate the $(1-\alpha)\%$ CI:

$$(1-\alpha)\%\ \mathrm{CI} = \left[\hat{\theta}(\alpha/2), \hat{\theta}(1-\alpha/2)\right], \tag{5}$$

where $\hat{\theta}(\alpha/2)$ is the $(\alpha/2)$-th percentile, and $\hat{\theta}(1-\alpha/2)$ is the $(1-\alpha/2)$-th percentile. CIs can be computed through bootstrapping for any of the estimators described in Section 3.1. While the calculated CIs may seem narrow, it is crucial to recognize that CIs represent a statistical measure of precision in estimating a parameter, e.g. the mean damage. A narrow CI reflects a high degree of confidence in the estimated mean damage, based on the current sample. With an infinite number of samples, $\hat{\mu}_d$ will asymptotically approach the true combined mean, μ_d.

Furthermore, we clarify the potential applications of two bootstrapping variants, as depicted in Figure 1. In the first case, the damage measurements are firstly categorized into bins, and random sampling with replacement is then directly implemented on clustered data, i.e., bootstrapping within each bin (BootBin). In the second approach, the whole measurement is bootstrapped (BootWhole), and subsequently, each bootstrapping sample is clustered, as shown in Figure 2. The key distinction between the BootWhole and BootBin methods lies in the variation of the number of data points in bins during each bootstrap iteration for BootWhole, whereas BootBin maintains a fixed number. In practice, BootWhole is more representative of repeated experiments, and more importantly, a primary assumption of bootstrapping is that the sample should

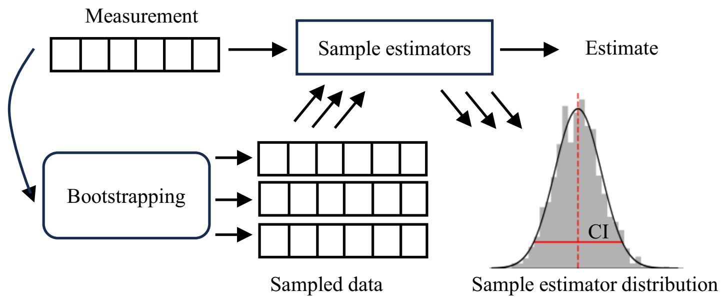

Figure 2: Schematic representation of BootWhole's statistical bootstrapping process, showcasing the resulting sample estimator distribution and its corresponding CI.

be representative of the whole population. Bootstrapping on the whole measurement is generally preferred to bootstrapping on clustered data, especially considering that the latter may result in biased results.

4. Case study

4.1. Monitoring campaign and data processing

The conducted research is based on a six-year measurement campaign, out of which, short-term 10-minute strain measurements and SCADA (Supervisory Control And Data Acquisition) data were collected from a three-megawatt offshore wind turbine installed on a monopile foundation in the North Sea. The turbine is equipped with six longitudinal strain gauges mounted 60° apart and installed at the tower-transition piece interface, 16 m above the LAT (Lowest Astronomical Tide).

From sensor strain measurements, we can calculate stress (S) and bending moments (M) at any heading. Selecting the heading as the average yaw angle within a 10-minute time window results in Fore-Aft (FA) and Side-Side (SS) bending moments for each window [10]. For the sake of simplicity, we only consider short-term damage from FA bending moments retrieved from power-generating conditions. We do not include wind direction and operational conditions as EOCs, with the primary emphasis being placed on wind speed EOC. Focusing only on wind speed and power generating mode constitutes a reasonable assumption since accurate lifetime estimates can be normally obtained if the damage data is divided into wind speed bins, as explained in [11]. For each 10-min measurement, we calculate its short-term (10-min) damage, d, via py-fatigue's Python package [12], adhering to the recommended practice guidelines DNVGL-RP-C203. Note that upon binning, each d_i represents a 10-min damage realization for the ith EOC bin. Specifically, d is computed through Palmgren-Miner's rule, where the stress-life (SN) curve is defined by Basquin's law:

$$d = \sum_{j=1}^{k} \frac{n_{c_j}}{N_j} = \frac{1}{a} \sum_{j=1}^{k} n_{c_j} (\Delta S_j)^m \, ; \text{ with } N(\Delta S)^m = a \, , \tag{6}$$

where (m, a) stand for the SN curve slope and intercept, respectively, k is the number of load blocks, n_{c_j} is the number of cycles in the jth load block, and N_j are the cycles to failure at the load level ΔS_j.

4.2. Case study definition

To comply with the design, the bin size is selected as 2 m/s, and short-term damages are calculated using a linear S-N curve with slope $m = 3$. As explained in Section 3 and shown in [4], binning results in damage variability. In this work, we quantify the damage variability through two estimators, damage mean and variance. The variance accounts for the aleatory uncertainty in short-term damages due to loads' natural variation and/or omitted EOCs. Moreover, each bin is associated with a certain occurrence probability, $p(x_i)$, which can either be determined using measurement data (e.g. several years of SCADA data) or based on design documents or available theoretical relations. Here, we rely on measured wind speed distributions. Note that measured distributions can be either adopted from as-measured probabilities, i.e. normalized histograms, or by fitting a distribution to as-measured wind speeds. We directly retrieve as-measured normalized histograms to avoid any distribution fitting errors.

Following the methodology proposed in Section 3, we investigate two bootstrapping variants: BootWhole and BootBin. Within the analysis, we consider multiple measurement periods: (i) 12 periods of 1 month, (ii) 6 periods of 6 months, and (iii) 6 periods of 1 year. BootBin first bins the selected measurement period into wind speed categories and then bootstraps on each bin. The bootstrapping sample size in each bin equals the number of datapoints in that bin. In contrast, BootWhole first bootstraps the whole studied measurement period and then bins the resampled data. The size of the sample in each bootstrap iteration equals the size of the studied measurement period. In each of the executed 10,000 bootstrap iterations, the combined mean and variance are calculated using Equations 3 and 4, and considering the wind speed probability for each measurement period. From the resulting 10,000 estimators, we subsequently compute bootstrapping statistics, e.g. expected value or CIs.

Since we are relying on the wind speed probability over the measurement period, we can verify our results. It should be, however, noted that for long-term damage predictions, a long-term wind speed probability should instead be used. We compare both bootstrapping variants with respect to the confidence interval (CI) that results from unbinned damage data (NoBin), and we expect that the mean, variance, and CIs computed from both bootstrapping methods are similar to those obtained from NoBin. Specifically, we calculate the 95% confidence intervals, which are delimited by the 2.5th percentile and the 97.5th percentile of the resampled distribution. In our analysis, the ground truth (target) corresponds to the deterministic mean and variance

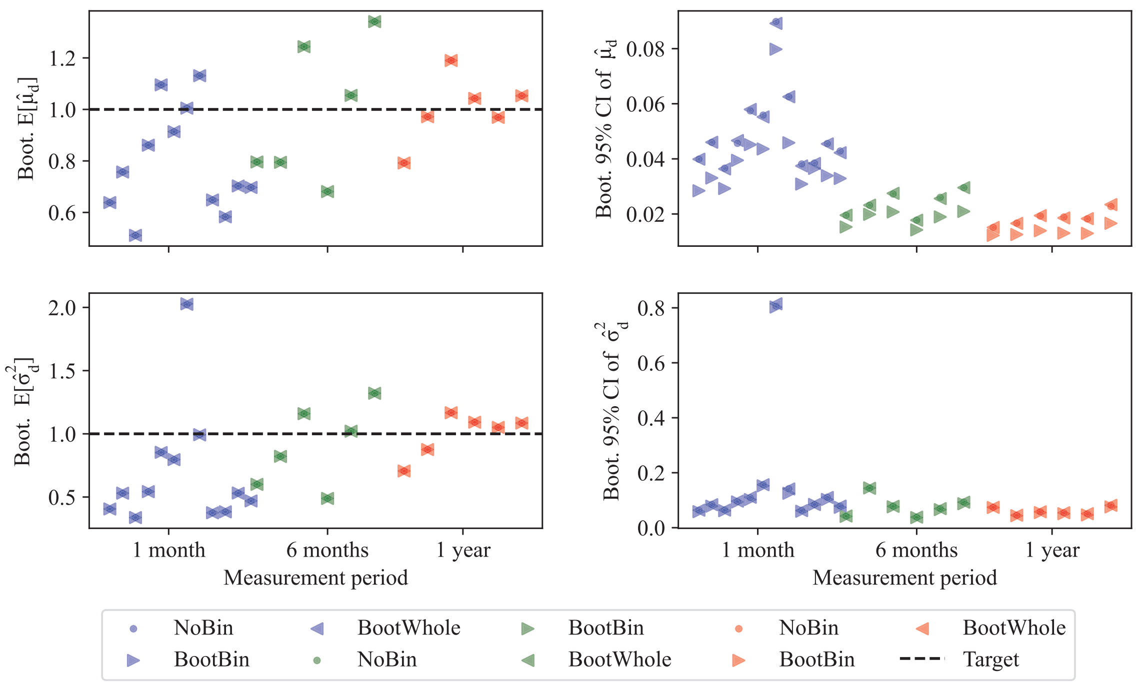

Figure 3: Comparative analysis of bootstrapping variants *BootWhole* and *BootBin* against *NoBin*. It should be noted that values are normalized to the target mean value calculated based on 6 years of data.

computed from 6-year fatigue damage measurements. Note that we intentionally do not fill in empty bins. Regarding BootWhole, this decision stems from the possibility of encountering iterations with empty bins when drawing random samples with replacement. By not filling the empty bins, we maintain a more accurate representation, resulting in smaller final damage values for those specific iterations, as the empty bins have no contribution to the combined damage estimators.

5. Results and discussion

The main results of our study are showcased in Figure 3. The top-left plot illustrates the bootstrapped expected value of the combined damage mean, normalized with respect to the target mean value. The ground-truth damage mean for each studied measurement period closely aligns with the bootstrapped expected value. This is logical since the applied wind speed frequency of occurrence is estimated from the considered measurement period. The bottom-left plot represents the bootstrapped expected value of the combined damage variance, normalized to the target variance value. Again, all bootstrapping variants are in good agreement with the ground-truth damage variance. The top-right plot showcases the 95% CI associated with the bootstrapped combined mean damage, normalized with respect to the target mean value. We observe that BootBin and BootWhole variants exhibit discrepancies for

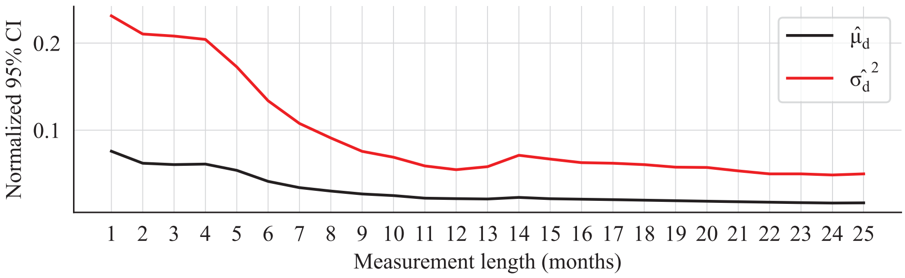

Figure 4: Evolution of normalized damage mean and variance 95% confidence intervals (CIs) over measurement period length.

all measurement periods, with these differences becoming more apparent for shorter measurement periods (1 month). BootWhole's CIs are in good agreement with NoBin, whereas BootBin's CIs are visibly narrower (top-right plot), which could potentially lead to a misleading sense of confidence and consequently biased outcomes. We also observe that 1-month measurements exhibit higher CIs than 1-year measurements, indicating increased confidence with additional available data, yet the CIs for 6 months are very similar to those for 1 year.

The bottom-right plot represents the normalized 95% CI of the bootstrapped combined damage variance, normalized with respect to the target variance. Upon inspection, a negligible discrepancy can be found between BootBin and BootWhole for all measurement periods. More generally, we can see that the combined estimators are more accurate (i.e. closer to the target) for longer measurement periods with more datapoints than those available in short measurement periods. This is, in turn, reflected by smaller CIs obtained for the case of longer measurement periods, which are naturally associated with less statistical uncertainty than for shorter periods.

Besides comparing bootstrapping variants, we also investigate the evolution of estimators' CIs over measurement periods. This process can be followed to define criteria for selecting a long enough measurement period, where the *confidence* over combined estimators can no longer be further reduced. Figure 4 illustrates the evolution of normalized estimators' 95% CIs from 1- to 25-month measurement periods with 1-month steps. Specifically, CIs of $\hat{\mu}_d$ and $\hat{\sigma_d}^2$ are normalized with respect to the target mean and variance, respectively. To exemplify the process, let us consider a 1-month measurement period. We first draw 5,000 bootstrap samples from the entire 1-month measurement dataset, and we then cluster and compute estimators for each bin, i.e. a vector of 5,000 elements is obtained for both damage mean and variance. Finally, binned estimators are combined and the normalized 95% CIs are computed from the resampled points, i.e. 5,000-element vectors.

As explained in Section 3, the width of 95% CIs is influenced by both sample size and variability within the population. Increasing the sample size naturally leads to a reduction in CI with respect to a smaller sample size (e.g. shorter measurement period). Note that CIs represented in Figure 4 mainly show that if we repeat the measurement, 95% of the calculated intervals will include the true value, as explained in Section 3. This does not mean that one could be 95% sure about it. Instead, one can inspect in the represented CI evolution what period is long enough to retrieve an accurate statistical estimator. This can be assessed based on a defined CI width. For example, in this case study, the CIs of estimators reduce at a negligible rate after almost 12 months of measurement. Therefore, we can say that the estimators have low statistical uncertainty using one year of measurement.

6. Conclusion and future work

This work introduces a structured methodology to quantify both aleatory and statistical uncertainty in fatigue damage predictions computed through binning. Our findings reveal inherent variability in fatigue damage estimated from measurements, thereby supporting the need for uncertainty quantification approaches. Moreover, our analysis confirms that bootstrapping directly on binned data may yield biased results, especially in datasets with a limited number of samples. In future research efforts, we envision the application of alternative methodologies for computing the complete statistical description of predicted fatigue estimates, from frequentist statistical approaches to fully Bayesian methods. Ultimately, our aim is to develop a systematic probabilistic framework for computing the lifetime of wind turbines that appropriately takes into account additional uncertainties, e.g., EOC parameters modeled as random variables.

7. Acknowledgments

This research is conducted within the project MAXWind, funded by the Belgian Energy Transition Fund (ETF).

8. References

[1] Marsh G 2016 Fatigue load monitoring of offshore wind turbine support structures *PhD thesis The University of Edinburgh.*

[2] Noppe N, Hübler C, Devriendt C and Weijtjens W 2020 Validated extrapolation of measured damage within an offshore wind farm using instrumented fleet leaders *J. Phys. Conf.* vol 1618 no. 2 p 022005 doi 10.1088/1742-6596/1618/2/022005.

[3] Hübler C 2019 Efficient probabilistic analysis of offshore wind turbines based on time-domain simulations*PhD thesis Leibniz Universität Hannover*

[4] Sadeghi N, D'Antuono P, Robbelein K, Noppe N, Weijtjens W and Devriendt C 2023 Deterministic and probabilistic damage calculation of offshore wind turbines considering the low-frequency fatigue dynamics *Life-Cycle of Structures and Infrastructure Systems Conf. CRC Taylor and Francis Group* p 3570-3577 https://doi.org/10.1201/9781003323020-437.

[5] Mozafari S, Dykes K, Rinker J and Veers P 2023 Effects of finite sampling on fatigue damage estimation of wind turbine components: A statistical study *Wind. Eng.* vol 47 issue 4 p 799-820 doi 10.1177/0309524X231163825.

[6] Mai Q, Weijtjens W, Devriendt C, Morato P G, Rigo P and Sørensen J 2019 Prediction of remaining fatigue life of welded joints in wind turbine support structures considering strain

measurement and a joint distribution of oceanographic data *Mar. Struct.* vol 66 p 307–322 doi 10.1016/j.marstruc.2019.05.002.
[7] Efron B 1992 Bootstrap methods: Another look at the jackknife, in Breakthroughs in statistics *Springer* p 569–593.
[8] Pacheco J, Pimenta F, Pereira S, Cunha A, Magalhaes F 2023 Experimental evaluation of strategies for wind turbine farm-wide fatigue damage estimation *Eng. Struct.* vol 285 115913.
[9] Morey, R D et al. 2016 The fallacy of placing confidence in confidence intervals *Psychonomic bulletin & review* vol 23 p 103-123.
[10] Sadeghi N, D'Antuono P, Noppe N, Robbelein K, Weijtjens W, Devriendt C 2023 Quantifying the effect of low-frequency fatigue dynamics on offshore wind turbine foundations: a comparative study *Wind Energy Sci.*vol 8 issue 12 p 1839-1852 doi https://doi.org/10.5194/wes-8-1839-2023.
[11] Hübler C et al. 2018 Reliability analysis of fatigue damage extrapolations of wind turbines using offshore strain measurements *J. Phys. Conf.* vol 1037 p 32-35.
[12] D'Antuono P, Weijtjens W and Devriendt C 2023 OWI-Lab/py_fatigue*Zenodo* [code] doi https://doi.org/10.5281/zenodo.7656680.