https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Interannual variability in fatigue damage estimation from short-term strain monitoring of offshore wind turbines

Negin Sadeghi[1], Pablo G. Morato[2], Nymfa Noppe[1], Nandar Hlaing[1], Wout Weijtjens[1], and Christof Devriendt[1]

[1]OWI-Lab, AVRG, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[2]Engineering Risk Analysis Group, Technical University of Munich, Munich, Germany

Correspondence: Negin Sadeghi (Negin.Sadeghi@vub.be)

Abstract. Extending the service life of offshore wind turbines demands fatigue assessments that reflect site-specific loading rather than conservative design assumptions. Structural health monitoring can provide strain-based damage estimates, but such monitoring campaigns are typically short. Consequently, long-term damage is often estimated from short-term strain data conditioned on available environmental and operational condition (EOC) records. Despite widespread use of such estimation approaches, the representativeness of a one-year strain window remains insufficiently quantified. In particular, it is unclear whether estimation errors are primarily driven by within-year statistical uncertainty or by interannual variability in the conditional damage–EOC mapping. This distinction is difficult to assess in practice due to the scarcity of long-term paired strain–EOC datasets. In this work, an eight-year strain-EOC dataset from an in-service offshore wind turbine is used to quantify damage estimation error by systematically shifting a one-year strain monitoring window across years. A hierarchy of damage-mapping strategies is evaluated, from unconditional extrapolation to EOC-conditioned estimators based on binned damage models. Unconditional extrapolation yields substantial window-dependent errors, with deviations up to 30% in the estimated long-term mean 10-minute damage, and exhibits pronounced interannual variability. Conditioning on informative EOCs generally reduces errors to around 10% and decreases sensitivity to the considered monitored year. However, these improvements are not monotonic with the dimensionality of the EOC-conditioned model. Bootstrap-based estimates of within-year statistical uncertainty are consistently small (<1%), indicating that estimation error is dominated by interannual variability. Longer strain monitoring periods reduce window-to-window variability, but do not eliminate damage estimation error under non-stationary conditions, where the EOC–damage mapping changes over time. These results show that long-horizon damage estimates remain sensitive to both the timing and duration of the strain-monitoring window. This sensitivity highlights the need to verify that the considered EOCs are representative of long-term conditions and that the damage–EOC mapping remains stable over time.

1 Introduction

Extending the operational lifetime of wind turbines is a critical consideration driven by economic, environmental, and resource-efficiency imperatives. Structural health monitoring (SHM) systems provide site-specific information on loads and structural

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

response, enabling data-driven fatigue assessment that can reduce the conservatism of design-based approaches and support lifetime extension decisions (IEC TS 61400-28, 2023; Hübler et al., 2018; Hübler and Rolfes, 2022). SHM systems can record load histories from installed strain sensors, as well as environmental and operational conditions (EOCs) obtained by supervisory control and data acquisition (SCADA) systems and/or other measurement devices, such as wave radars. However, continuous long-term load monitoring is often impractical, as sensors may drift, break, or be installed in inaccessible locations, and sensor replacement is costly. As a result, strain measurements are typically available only for limited periods, often in the order of months or a few years, while EOC data from SCADA systems are available over much longer time spans. This leads to a central question for data-driven lifetime assessment: how can the uncertainty in fatigue damage estimates over a period of interest be quantified when strain measurements are available only over a limited monitoring window?

A common strategy for long-term fatigue damage estimation is to extrapolate or map short-term fatigue damage from strain measurements to longer periods, often leveraging information about EOCs (Hübler et al., 2018; Mujtaba et al., 2026). In its simplest form, this mapping assumes that the measured damage scales linearly in time, an assumption that can logically lead to significant estimation inaccuracies when short-term strain data are not representative of the longer estimation period. More advanced data-driven approaches condition short-term damage on selected EOC parameters using binning or other conditional models (Hübler et al., 2018; Mujtaba et al., 2026; Hlaing et al., 2024). In such approaches, damage calculated from 10-minute strain measurements is related to the corresponding EOC conditions, and the estimated damage over the available EOC period is obtained by weighting these conditional damages according to the EOC distribution during that period. However, in practice, only a subset of EOCs can usually be considered, as some may not be available, measured, or identified as relevant, whereas exhaustively conditioning fatigue damage on numerous EOCs can result in sparse or empty bins. Consequently, if relevant EOCs are neglected and their distributions or effects vary over time, or if residual variability in damage is non-stationary, the resulting models may not be representative of long-term fatigue damage over the estimation period. These non-stationary effects become more pronounced in highly stochastic offshore environments.

Depending on the intended application, two fundamentally distinct tasks can be defined: prediction and conditional estimation. In predictive settings, short-term strain measurements are used to estimate fatigue damage beyond the period for which strain and EOC data are available, requiring assumptions or models about future or historical EOC variability. In contrast, conditional estimation focuses on estimating fatigue damage over a period for which EOC data are available, while strain measurements are limited to a shorter subperiod. In this case, the target-period EOC distribution is observed and can be used directly, avoiding the need to model unknown future or past EOCs. However, both prediction and conditional estimation share a common source of uncertainty: the inherent variability of fatigue damage under given EOCs and the potential non-stationarity of the conditional damage model due to unobserved or unaccounted factors.

Regarding the duration of the effective short-term strain monitoring period for accurate long-term damage estimation, the literature presents conflicting views (Hübler et al., 2018; Marhadi and Skrimpas, 2015; Marsh, 2016; Loraux and Brühwiler, 2016; Mai et al., 2019). Hübler et al. (2018) and Hübler and Rolfes (2022) suggest that 9-10 months of strain data may already yield representative mean values with long-term damage estimation errors below 10%, whereas Marhadi and Skrimpas (2015) and Marsh (2016) argue that seasonal and interannual variability requires periods of at least 18 months, or even multiple years,

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

to capture rare but influential events, such as storms. The reported conclusions are often supported by analyses that account for statistical uncertainty associated with limited strain monitoring data via bootstrapping or other resampling techniques. However, while statistical uncertainty quantifies variability due to limited data (Morey et al., 2016; Efron, 1992), it does not capture non-stationary uncertainty arising from unconsidered features or non-stationary noise.

Despite significant progress in data-driven fatigue damage assessment, the existing literature relies on several implicit assumptions that remain largely unverified under long-term operating conditions. Although a small number of studies discern between aleatoric and epistemic uncertainty (Mai et al., 2019; Tatsis et al., 2017), these approaches typically assume that the conditional relationship between EOCs and fatigue damage is stationary over time. Uncertainty quantification is thus often restricted to statistical uncertainty arising from finite strain sample sizes, commonly assessed through resampling techniques such as bootstrapping, without explicitly evaluating whether these measures capture stochastic variability beyond the measurement period. Moreover, most empirical validations are based on datasets spanning at most one to three years (Hübler and Rolfes, 2022), limiting the ability to observe interannual variability, rare events, and non-stationary behavior in either the environmental loading or the structural response. As a result, it remains unclear whether the reported uncertainty and accuracy claims remain valid when short-term strain measurements are mapped to substantially longer fatigue damage estimation periods.

In this work, we address these limitations through a comprehensive long-term analysis based on a rarely available eight-year strain and SCADA monitoring dataset from an in-service offshore wind turbine. Rather than targeting predictive lifetime extrapolation (Hübler et al., 2019, 2018; Loraux and Brühwiler, 2016; Marsh, 2016; Pacheco et al., 2022), we deliberately focus on a fatigue damage estimation task, in which damage is estimated over periods with known EOC histories. This choice allows us to isolate and quantify interannual variability in fatigue damage estimates that arises solely from limited strain monitoring, without conflating it with additional uncertainties inherent to prediction tasks, such as unknown future environmental conditions. In our study, we systematically quantify the variability in estimated fatigue damage induced by different short-term strain measurement windows and showcase how this variability persists even when state-of-the-art conditional damage models are employed. We further disentangle the dominant sources of uncertainty by explicitly distinguishing between statistical uncertainty due to finite sample sizes, commonly assessed via bootstrapping, and non-stationary uncertainty arising from unobserved or unaccounted variability in the damage response under given EOCs. By analyzing multiple short-term measurement durations and varying bin sizes and dimensionalities, we show how modeling choices interact with limited data to amplify or attenuate interannual variability. By examining how the inclusion or omission of specific EOCs affects the stability of the conditional damage mapping, we provide a causal interpretation of the observed variability and identify which environmental drivers are responsible for residual non-stationarity. Together, these contributions provide a rigorous validation of existing data-driven fatigue damage estimation practices and reveal fundamental limitations that directly propagate into long-term fatigue damage predictions based on short-term strain measurements. To structure the analysis, our study addresses the following key research questions:

- To what extent does the estimated long-term fatigue damage depend on the specific one-year strain measurement window used, when estimation is performed over a fixed EOC period using state-of-the-art linear and conditional (binning-based) damage models?

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

- What are the relative contributions of statistical uncertainty due to finite strain sample sizes and non-stationary aleatoric variability in the conditional damage response to the observed interannual variability in estimated fatigue damage?
- How does increasing the duration of the strain monitoring period (from one to multiple years) affect the magnitude and dispersion of interannual variability in fatigue damage estimates?
- How should bootstrapping-based uncertainty quantification be correctly interpreted in fatigue damage estimation from limited strain measurements, and why is it fundamentally unable to represent interannual variability arising from non-stationary conditional damage?
- How does the selection and dimensionality of considered EOCs in the conditional damage mapping influence the stability of fatigue damage estimates and the persistence of residual non-stationarity?

In Sec. 2, we review existing data-driven approaches for fatigue damage estimation using strain measurements and identify key unresolved challenges. In Sec. 3, a framework for mapping measured strain-based fatigue damage to longer periods with available EOCs is introduced, including the problem setting, notation, and formulation. In Sec. 4, the available strain measurement data, monitoring setup, and fatigue damage calculation procedure are described, together with the experimental protocol. In Sec. 5, the impact of interannual variability on fatigue damage estimates is systematically analyzed. The effects of monitoring duration, wind-speed discretization, and within-bin variability are presented and discussed. Additionally, an interpretation of narrow bootstrap confidence intervals and their relationship to interannual variability is shown. Finally, in Sec. 6, we summarize the main findings.

2 Related work and unresolved challenges

2.1 Data-driven fatigue damage estimation from limited strain measurements

Data-driven approaches for long-term fatigue damage estimation seek to estimate accumulated damage over extended periods using strain measurements available only over limited monitoring windows, typically complemented by longer records of environmental and operational conditions. Existing methods range from linear temporal extrapolation to conditional damage mapping approaches, including binning-based and machine learning models (Noppe et al., 2020; Sadeghi et al., 2024; Hübler and Rolfes, 2022; Mujtaba et al., 2026). Linear temporal extrapolation (Loraux and Brühwiler, 2016), assumes that both the EOC distribution and the damage response are stationary, such that the mean damage observed during the measurement period can be directly linearly scaled to the estimation period (Hübler et al., 2018; Ziegler, 2018). Conditional approaches partially relax this assumption by estimating fatigue damage as a function of selected EOCs, e.g., wind speed, turbulence intensity, and wave height, and reweighting these conditional estimates using the EOC distribution over the estimation period (Hübler et al., 2018; Noppe et al., 2020; Pacheco et al., 2023). Comparative studies generally report improved performance of low-dimensional conditional models, particularly those conditioned on wind speed, when strain data are scarce (Hübler and Rolfes, 2022). However, these approaches implicitly assume that the selected EOCs capture the dominant drivers of fatigue damage and

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

that the conditional damage–EOC relationship remains stable over time. The robustness of these assumptions under long-term interannual variability is not systematically evaluated due to the limited duration of most analyzed datasets.

2.2 Selection of environmental and operational parameters in conditional models

A fundamental challenge in conditional fatigue damage estimation is the selection of an appropriate subset of EOCs. While increasing the number of conditioning variables can, in principle, reduce unexplained variability in damage estimates, practical implementations are constrained by sparse or empty bins when strain measurements are limited. Consequently, most studies condition damage on a small number of EOCs, typically wind speed, turbulence intensity, wind direction, and operational state (Loraux and Brühwiler, 2016; Hübler et al., 2018; Noppe et al., 2020; Mai et al., 2019; Marsh, 2016). Restricting the conditioning space implicitly assumes that unconsidered EOCs either have negligible influence on fatigue damage or exhibit stationary distributions over the estimation period. Although recent work has explored higher-dimensional conditioning schemes, including wave-related parameters (Hübler and Rolfes, 2022), empirical validation remains limited to short monitoring periods. As a result, it remains unclear whether reported reductions in estimation error reflect improved representativeness of the damage model or an alignment that is specific to the selected measurement window and does not generalize to the estimation period.

2.3 Strain monitoring duration and representativeness

The literature provides conflicting guidance regarding the minimum duration of strain monitoring required for accurate long-term fatigue damage estimation. Such variation is expected, as the minimum measurement duration also depends on factors such as the turbine design and site-specific conditions. Some studies suggest that periods of approximately 9-10 months may yield representative mean fatigue damage estimates with errors below 10% (Hübler and Rolfes, 2022; Hübler et al., 2018), provided that seasonal variability is captured and no major structural or environmental changes occur during the estimation period. Other studies argue that longer monitoring durations, ranging from 12-18 months to multiple years, are necessary to capture influential events such as storms or emergency shutdowns (Marhadi and Skrimpas, 2015; Marsh, 2016; Loraux and Brühwiler, 2016; Mai et al., 2019). These conclusions are typically supported by uncertainty analyses applied to the available strain data. However, due to the limited duration of most datasets, such analyses cannot directly assess interannual variability or evaluate whether short-term measurements are representative of longer estimation periods. Consequently, the extent to which reported accuracy levels generalize beyond the specific monitoring window remains unclear.

2.4 Uncertainty quantification under limited data

Uncertainty quantification in data-driven fatigue damage estimation is most commonly performed using resampling techniques, such as bootstrapping (Morey et al., 2016), to characterize statistical uncertainty arising from finite sample sizes (Hübler and Rolfes, 2022; Pacheco et al., 2023; Noppe et al., 2020; Marsh, 2016; Petrovska et al., 2020). These methods are often interpreted as providing confidence bounds on long-term damage estimates. However, bootstrapping quantifies only statistical uncertainty associated with the measurement window and does not account for variability arising from non-stationary damage behavior or

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

unobserved EOCs. A detailed methodological treatment of bootstrapping strategies, including their assumptions, limitations, and correct interpretation in the context of fatigue damage estimation, is provided in Sadeghi et al. (2024). In Sadeghi et al. (2024), it is further clarified that bootstrap-based bounds quantify data sufficiency within the measurement window, but do not imply that the measurement period is representative of longer estimation periods.

Only a limited number of studies explicitly model the stochastic behavior beyond the measurement period by propagating uncertainty from short-term data to longer estimation horizons. For example, Mai et al. (2019) addressed predictive fatigue damage estimation by distinguishing between the year-to-year variation of loading conditions, captured through the scale parameter of the wind speed distribution updated via Bayesian inference, and the statistical uncertainty in the fitted stress-range distributions arising from limited strain measurement duration. Under the assumption that waves are primarily wind-driven, interannual variability in fatigue loading was considered to be sufficiently represented by the variability in wind speed, such that the wind speed binning captures the year-to-year differences in fatigue damage. The residual uncertainty on the stress-range distribution parameters was, therefore, treated as a sampling effect due to finite data, incorporated by assigning a normally distributed random variable to the scale parameter of the fitted stress-range distribution within each wind speed bin, with a coefficient of variation that decreases as more measurement data become available. A Bayesian framework was then used to update the joint distribution of EOCs and infer long-term fatigue damage. While this approach constitutes an important step toward uncertainty-aware damage prediction, it relies on assumptions regarding the representativeness of selected EOCs and the adequacy of short-term strain measurements to characterize long-term conditional damage behavior. Moreover, such formulations have not been empirically validated against long-duration strain datasets, limiting the ability to assess whether the adopted uncertainty models capture observed interannual variability. Overall, the scarcity of datasets combining multi-year strain measurements with comprehensive EOC records has constrained systematic evaluation of uncertainty quantification methods and the stability of conditional damage mappings over time, leaving key questions regarding representativeness and non-stationarity unresolved.

3 Methods

This section formalizes the fatigue damage estimation task, the family of damage mapping models considered in this study, and the protocols used to quantify (i) *statistical uncertainty* induced by finite strain data and (ii) *interannual variability* induced by the choice of the strain monitoring window.

3.1 Problem setting and notation

We consider 10-minute time indices $t = 1, \ldots, N$, where N denotes the dataset length. For each timestamp t, the strain measurements are converted into a short-term (10-minute) fatigue damage value $d_t \in \mathbb{R}_+$, and the corresponding EOCs are summarized by a feature vector $\mathbf{x}_t \in \mathbb{R}^X$ (e.g., wind speed, wind direction, turbulence intensity, operational state). Two time periods are defined:

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

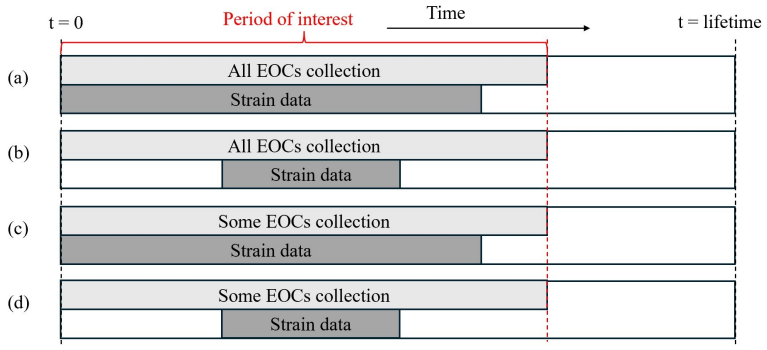

Figure 1. Schematic representation of data availability in data-driven fatigue damage estimation. The red bounding box denotes the period of interest, while the full timeline indicates the system lifetime. Each row illustrates a distinct combination of strain data (damage-relevant measurements) and environmental or operational conditions (EOCs). In (a) and (b), EOCs are fully observed over the period of interest, whereas in (c) and (d), only a subset of representative EOCs is considered, despite covering the full period. In (a) and (c), strain data nearly cover the period of interest, while in (b) and (d), strain data are restricted to a shorter subinterval.

- Measurement window $\mathcal{M}$ (limited strain monitoring): a contiguous set of indices with $|\mathcal{M}| = N_m$, for which both $\{d_t, \mathbf{x}_t\}_{t\in\mathcal{M}}$ are available.
- Period of interest (PI) $\mathcal{P}$ (long-term estimation): a larger set of indices than the measurement window with $|\mathcal{P}| = N_{\mathrm{PI}}$, for which EOCs $\{\mathbf{x}_t\}_{t\in\mathcal{P}}$ are available, and for which we estimate accumulated fatigue damage.

The target quantity is the mean 10-minute damage over the PI:

$$\mu_d = \frac{1}{N_{\mathrm{PI}}} \sum_{t\in\mathcal{P}} d_t, \tag{1}$$

and the corresponding accumulated damage over $\mathcal{P}$:

$$D = N_{\mathrm{PI}} \cdot \mu_d \,. \tag{2}$$

In this work, we focus on a fatigue damage estimation task: $\mathcal{P}$ corresponds to a time span for which EOCs are known (e.g., SCADA history), while strain-derived damages are available only for a shorter subperiod $\mathcal{M} \subset \mathcal{P}$. This isolates the uncertainty induced by limited strain monitoring from additional uncertainty sources specific to prediction beyond the available EOC record. Even in a task where the EOC history over $\mathcal{P}$ is known (e.g., from SCADA), two practical limitations remain: (i) the fatigue load response (strain-derived damages) is only observed in a limited window $\mathcal{M}$, and (ii) some relevant EOCs may be missing, unavailable, or unaccounted for in $\mathbf{x}_t$. Figure 1 illustrates representative cases ranging from an ideal setting (representative $\mathcal{M}$ and complete EOCs) to more common situations where $\mathcal{M}$ is unrepresentative of $\mathcal{P}$ and/or the conditioning EOCs are incomplete, which can induce systematic estimation bias even when conditional damage models are used.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

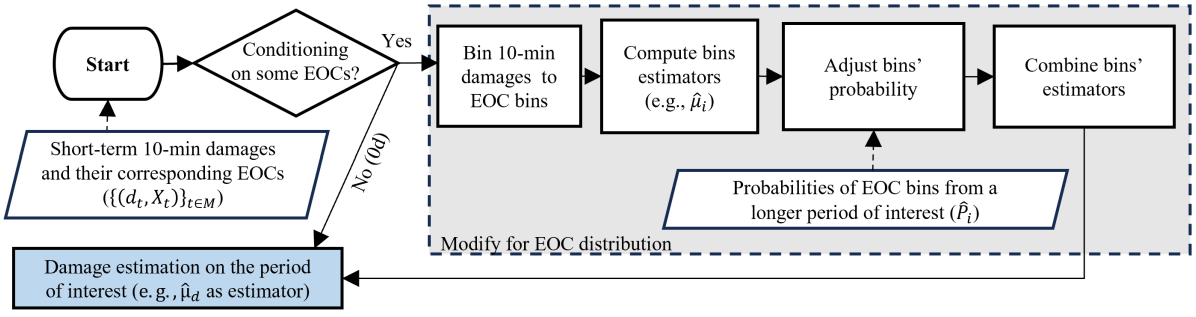

Figure 2. Flowchart for estimating fatigue damage over a defined period of interest from short-term (10-minute) damage measurements. The procedure considers two estimation pathways: linear damage extrapolation (0D) and a binning approach conditioned on environmental or operational conditions (EOCs). Starting from 10-minute damage–EOC pairs, it is first assessed whether conditioning on EOCs is performed. If not, damage is extrapolated directly over the period of interest using linear scaling (0D). If conditioning is applied, short-term damages are assigned to EOC bins, bin-wise damage estimators are computed, and their contributions are weighted by the probability of occurrence of each EOC bin over the period of interest. The final estimate is obtained by combining the probability-adjusted bin contributions.

3.2 Damage mapping models

We study a hierarchy of mapping models that estimate $\hat{\mu}_d$ (and thus $\hat{D}$) from the measurement window $\mathcal{M}$ and the EOC history from $\mathcal{P}$. Models differ in the degree to which they condition on EOCs. Figure 2 provides a workflow view of the estimation task. Given paired observations $\{(d_t, \mathbf{x}_t)\}_{t\in\mathcal{M}}$ and an EOC history $\{\mathbf{x}_t\}_{t\in\mathcal{P}}$, we either: (i) use the measurement-window mean directly (0D, i.e., no EOC conditioning), or (ii) estimate a conditional damage mapping over EOCs and reweight it using the empirical EOC distribution over $\mathcal{P}$, yielding an EOC-adjusted estimator (Sadeghi et al., 2024). This work focuses on binning-based conditional mapping, but the same workflow applies to other conditional estimators that map EOCs to damage and are aggregated over $\mathcal{P}$.

Unconditional linear extrapolation

The simplest approach assumes that the measurement window, $\mathcal{M}$, is statistically representative of the entire period of interest, PI (Amiri et al., 2019; Ziegler and Muskulus, 2016a). Consequently, the PI mean damage is estimated directly as the sample mean of measured damages (Hübler and Rolfes, 2022):

$$\widehat{\mu}_{0\mathrm{D}} = \frac{1}{N_m} \sum_{t\in\mathcal{M}} d_t \,, \tag{3}$$

yielding an accumulated damage estimate:

$$\widehat{D}_{0\mathrm{D}} = N_{\mathrm{PI}} \cdot \widehat{\mu}_{0\mathrm{D}}. \tag{4}$$

This model inherently neglects any distributional shifts in EOCs between the measurement window $\mathcal{M}$ and the target period $\mathcal{P}$, potentially leading to estimation errors (Mozafari, 2023).

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Conditional binning models

To account for EOC variability over $\mathcal{P}$, we consider conditional damage mapping via binning. Let $\mathbf{z}_t \in \mathbb{R}^Z$ denote a selected subset of EOCs ($0 < Z \leq X$), and let $\mathcal{B} = \{B_1, \ldots, B_{N_b}\}$ be a partition of the Z-dimensional EOC space into N_b discrete bins. Each timestamp t is assigned a bin index $b(t) \in \{1, \ldots, N_b\}$ based on the observed $\mathbf{z}_t$. The process consists of a conditional estimation phase using the measurement window, followed by an estimation phase over the PI. For each bin i, we estimate the conditional mean damage from the measurement window $\mathcal{M}$:

$$\widehat{\mu}_i = \frac{1}{N_{m,i}} \sum_{t \in \mathcal{M}\ b(t)=i} d_t\,, \qquad N_{m,i} = |\{t \in \mathcal{M} : b(t) = i\}|\,. \tag{5}$$

In the absence of observations for bin i within the measurement window (i.e., $N_{m,i} = 0$), the estimator $\widehat{\mu}_i$ is undefined. Let $z \in \{0, 1, \ldots, Z\}$ denote the conditioning level, where level z corresponds to conditioning on the first z selected EOC variables, and level 0 corresponds to the unconditional model. To establish a fully specified damage mapping, we recursively define a robust estimator $\tilde{\mu}_i^{(z)}$ at hierarchy level z. This estimator prioritizes local data where available, while imputing missing values using the parent level $z - 1$ when necessary:

$$\tilde{\mu}_i^{(z)} = \begin{cases} \widehat{\mu}_i, & \text{if } N_{m,i} > 0 \quad \text{(observed)}, \\ \tilde{\mu}_i^{(z-1)}, & \text{if } N_{m,i} = 0 \quad \text{(imputed)}, \end{cases} \tag{6}$$

where the base case $\tilde{\mu}^{(0)}$ defaults to the unconditional mean $\widehat{\mu}_{0\mathrm{D}}$. Note that for all populated bins, this robust estimator is identical to the standard sample mean ($\tilde{\mu}_i^{(z)} \equiv \widehat{\mu}_i$); the recursive imputation applies only to sparse regions of the EOC space. The resulting set $\{\tilde{\mu}_i^{(Z)}\}_{i=1}^{N_b}$ constitutes the fully specified damage mapping derived from $\mathcal{M}$. For notational conciseness, we denote these final estimates simply as $\widehat{\mu}_i$.

To estimate the damage over the full PI, we characterize the target EOC distribution by computing the empirical probability of each bin using the long-term record from $\mathcal{P}$:

$$\widehat{P}_i = \frac{1}{N_{\mathrm{PI}}} |\{t \in \mathcal{P} : b(t) = i\}|\,, \qquad \sum_{i=1}^{N_b} \widehat{P}_i = 1. \tag{7}$$

The estimated PI mean damage is then obtained by reweighting the bin-level conditional means, $\widehat{\mu}_i$, according to the PI probabilities, $\widehat{P}_i$, (Sadeghi et al., 2024):

$$\widehat{\mu}_{ZD} = \sum_{i=1}^{N_b} \widehat{P}_i \cdot \widehat{\mu}_i, \tag{8}$$

and the total accumulated damage is estimated as:

$$\widehat{D}_{ZD} = N_{\mathrm{PI}} \cdot \widehat{\mu}_{ZD}. \tag{9}$$

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

3.3 Estimation error and uncertainty decomposition

To evaluate model performance, we define the *estimation error* of the estimated PI mean as:

$$e = \mu_d - \widehat{\mu}_d, \tag{10}$$

where μ_d is the ground-truth PI mean computed from the full record $\mathcal{P}$ and $\widehat{\mu}_d$ is the estimate derived from the measurement window $\mathcal{M}$. In this paper, we report the *normalized error* as a percentage:

$$e_{\mathrm{norm}} = \frac{e}{\mu_d} \cdot 100\,. \tag{11}$$

The error e arises from two distinct mechanisms: (i) *statistical uncertainty* due to the finite number of samples available in $\mathcal{M}$, and (ii) *interannual variability* (or systematic bias) caused by limited representativeness, where the distribution of unconsidered EOCs or structural response in $\mathcal{M}$ differs from $\mathcal{P}$. We quantify these components separately as follows.

3.3.1 Statistical uncertainty (bootstrap analysis)

Even if the measurement window $\mathcal{M}$ is fixed, estimators of fatigue damage derived from finite strain samples exhibit variability due to sampling effects. We quantify this sampling variability conditional on $\mathcal{M}$ using the *BootWhole* ("resample-then-bin") bootstrap strategy, in which resampling is performed on the full paired set in $\mathcal{M}$ before recomputing bin statistics (Sadeghi et al., 2024):

- Resampling: Construct N_{boot} bootstrap replicates. For each replicate j, sample N_m paired observations $\{(d_t, \mathbf{z}_t)\}_{t\in\mathcal{M}}$ with replacement from the measurement window.
- Re-estimation: For each replicate, recompute the conditional bin statistics (e.g., bin-wise mean damage) using Eq. (5), and evaluate a damage estimator $\widehat{\theta}(\cdot)$ based on the resampled data.
- Aggregation: For estimators based on bin reweighting, the bootstrap replicate is obtained as

$$\widehat{\theta}^{(j)} = \sum_{i=1}^{N_b} \widehat{P}_i \cdot \widehat{\theta}_i^{(j)}, \tag{12}$$

  where $\widehat{\theta}_i^{(j)}$ denotes the conditional statistic in bin i computed from the bootstrap sample, and $\widehat{P}_i$ are the fixed target probabilities derived from $\mathcal{P}$ via Eq. (7).

The ensemble $\{\widehat{\theta}^{(j)}\}_{j=1}^{N_{\mathrm{boot}}}$ provides an empirical approximation of the sampling distribution of the estimator conditional on the observed measurement window. An empirical $(1-\alpha)$ confidence interval (CI) can then be derived from the quantiles of this ensemble as

$$\mathrm{CI}_{1-\alpha} = \left[Q_{\alpha/2}(\widehat{\theta}^{(j)}),\, Q_{1-\alpha/2}(\widehat{\theta}^{(j)})\right]. \tag{13}$$

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Importantly, these bootstrap quantile bounds reflect data sufficiency and sampling variability within the fixed window $\mathcal{M}$. They do not imply that $\mathcal{M}$ is representative of $\mathcal{P}$, and therefore cannot capture window-to-window variability or non-stationary conditional damage behavior across years. In the specific case of estimating the PI mean fatigue damage using conditional binning, the estimator reduces to:

$$\widehat{\theta}^{(j)} \equiv \widehat{\mu}^{(j)} = \sum_{i=1}^{N_b} \widehat{P}_i \cdot \widehat{\mu}_i^{(j)}, \tag{14}$$

which corresponds to the reweighted conditional mean defined in Eq. (8) evaluated on the bootstrap sample. This interval reflects statistical uncertainty associated with finite strain samples within $\mathcal{M}$.

3.3.2 Interannual variability (sliding window analysis)

To quantify the sensitivity of the estimated damage to the choice of monitoring period (temporal representativeness), we perform a sliding window analysis. We use time-consecutive windows rather than random resampling because random sampling tends to collapse temporal structure toward the dataset mean and can obscure year-by-year variability (Hübler et al., 2018). Specifically, we generate K time-consecutive measurement windows $\mathcal{M}^{(k)}$ of fixed duration L (e.g., $L = 1$ year) using a circular sliding scheme with step Δ (e.g., 90 days), to use the measurement period evenly. Figure 3 illustrates the resulting set of potentially overlapping windows and the wrap-around at the dataset boundary. For each window $\mathcal{M}^{(k)}$, we:

- Compute the conditional mapping $\widehat{\mu}_i^{(k)}$ using data exclusively from $\mathcal{M}^{(k)}$ (Eq. (5));
- Estimate the global mean $\widehat{\mu}^{(k)}$ over $\mathcal{P}$ using the fixed weights $\widehat{P}_i$ (Eq. (8));
- Evaluate the normalized error $e_{\mathrm{norm}}^{(k)}$ relative to the ground truth (Eq. (11)).

The dispersion of the resulting error distribution $\{e_{\mathrm{norm}}^{(k)}\}_{k=1}^{K}$ quantifies the interannual variability of the considered strategy. To assess the impact of monitoring duration, this protocol is repeated for varying window lengths L.

3.4 Bin-wise error attribution

To diagnose the sources of interannual variability, we decompose the normalized estimation error into contributions from individual EOC bins. Let μ_i denote the *reference* (ground-truth) mean 10-minute damage within bin i computed over the full period of interest $\mathcal{P}$, and let $\widehat{\mu}_i^{(k)}$ denote the corresponding bin-wise estimate obtained from monitoring window $\mathcal{M}^{(k)}$ (Eq. (5)). We weight bin-level discrepancies using the empirical bin probabilities over $\mathcal{P}$, $\widehat{P}_i$ (Eq. (7)), yielding a bin-wise contribution to error (in percent) defined as:

$$C_i^{(k)} = \frac{\widehat{P}_i \left(\mu_i - \widehat{\mu}_i^{(k)}\right)}{\mu_d} \cdot 100\,, \tag{15}$$

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

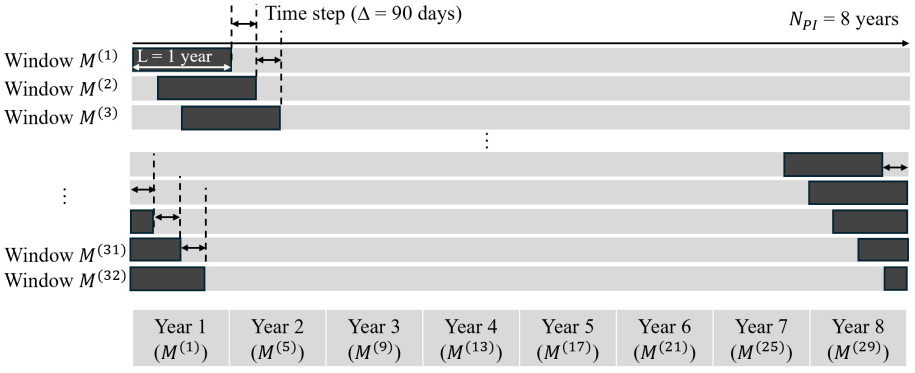

Figure 3. Schematic representation of the considered time windows for fatigue damage estimation. The K=32 windows $M^{(k)}$ of duration L = 1 year (black) are constructed using a sliding scheme with time step $\Delta = 90$ days, resulting in overlapping segments across the 8-year period of interest, N_{PI}. The panel at the bottom shows an alternative scheme based on 8 consecutive non-overlapping yearly windows (gray).

where μ_d is the ground-truth PI mean damage (Eq. (1)). By construction, the sum of bin contributions recovers the normalized error for window k:

$$e^{(k)}_{\mathrm{norm}} = \sum_{i=1}^{N_b} C_i^{(k)} = \frac{\sum_{i=1}^{N_b} \widehat{P}_i \left(\mu_i - \widehat{\mu}_i^{(k)}\right)}{\mu_d} \cdot 100\,. \tag{16}$$

To identify bins that consistently drive the error across windows, we define an importance score, I_i, as the average absolute contribution over K windows:

$$I_i = \frac{1}{K} \sum_{k=1}^{K} \left| C_i^{(k)} \right|. \tag{17}$$

Bins are ranked by I_i, and the highest-ranked bins can be identified as the dominant contributors to interannual variability. To characterize within-bin error independently of the global reweighting, the normalized bin-level error can be calculated:

$$e^{(k)}_{\mathrm{norm},i} = \frac{\mu_i - \widehat{\mu}_i^{(k)}}{\mu_i} \cdot 100\,. \tag{18}$$

Unlike $C_i^{(k)}$, which weights discrepancies by the empirical bin probability $\widehat{P}_i$ over $\mathcal{P}$, $e^{(k)}_{\mathrm{norm},i}$ measures the relative error conditional on bin i.

4 Eight-year strain and SCADA monitoring dataset: empirical evaluation

Our monitoring campaign provides an extensive, long-duration record to assess how short-term strain measurements can support fatigue damage estimation over a much longer period with known environmental and operational conditions. Specifically, the eight-year joint strain–SCADA dataset enables (i) a controlled setting in which the period of interest is fixed, (ii) systematic variation of the strain measurement window, and (iii) a direct quantification of interannual variability that cannot be observed in typical one-to-three-year datasets.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

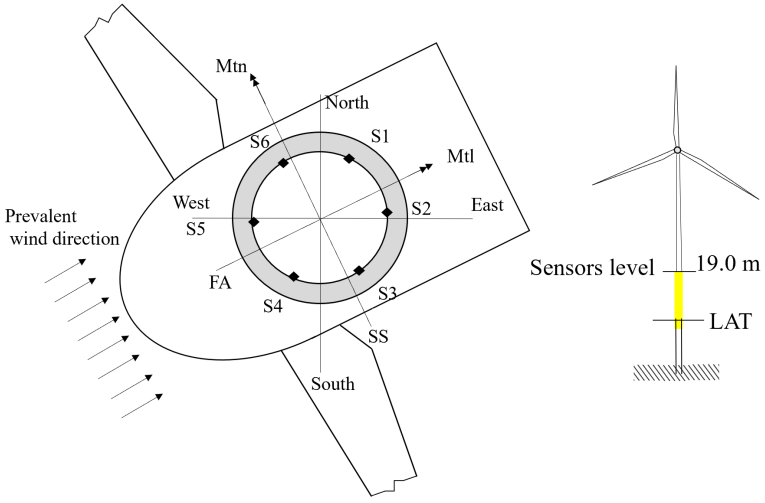

Figure 4. Sensor layout on the turbine tower and associated loading directions (adapted from Sadeghi et al. (2023)). Six strain sensors (S1–S6) are mounted in a circumferential arrangement at a height of 19 m above the lowest astronomical tide (LAT). The coordinate system and principal directions are indicated, together with the prevalent wind direction. M_{tn} and M_{tl} denote bending moments in the normal (along-wind, fore-aft) and lateral (cross-wind, side-side) directions, respectively.

4.1 Data availability and setup

We analyze an eight-year monitoring dataset from a 3 MW, monopile-supported offshore wind turbine located in the North Sea. The dataset includes 10-minute strain time series and concurrent 10-minute SCADA records. As shown in Figure 4, the turbine is instrumented with six strain gauges at the tower-transition piece interface. Figure 5 summarizes daily strain availability over the eight years, computed as the fraction of available 10-minute strain records per day. All years exceed 95% availability, except years 3, 4, and 6 (72%, 85%, and 89%, respectively).

4.2 Fatigue damage calculation from strain measurements

While a comprehensive explanation of the transformation from raw strain measurements to 10-minute fatigue damage in a specific heading is provided by Sadeghi et al. (2022), the procedure is summarized in Figure 6. For each 10-minute interval, we convert measured strain time series to bending moments in the fore-aft (FA) and side-side (SS) directions using the nacelle yaw angle from SCADA data. We then compute the damage following the DNV-RP-C203 guidelines (DNV, 2016). Specifically, we apply rainflow counting to the reconstructed stress time series and calculate Palmgren-Miner's damage for each sensor's heading, FA, and SA directions, adopting the linear SN curve DNV-D-W with a slope $m = 3$ (Socie, 1992; Palmgren, 1924).

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

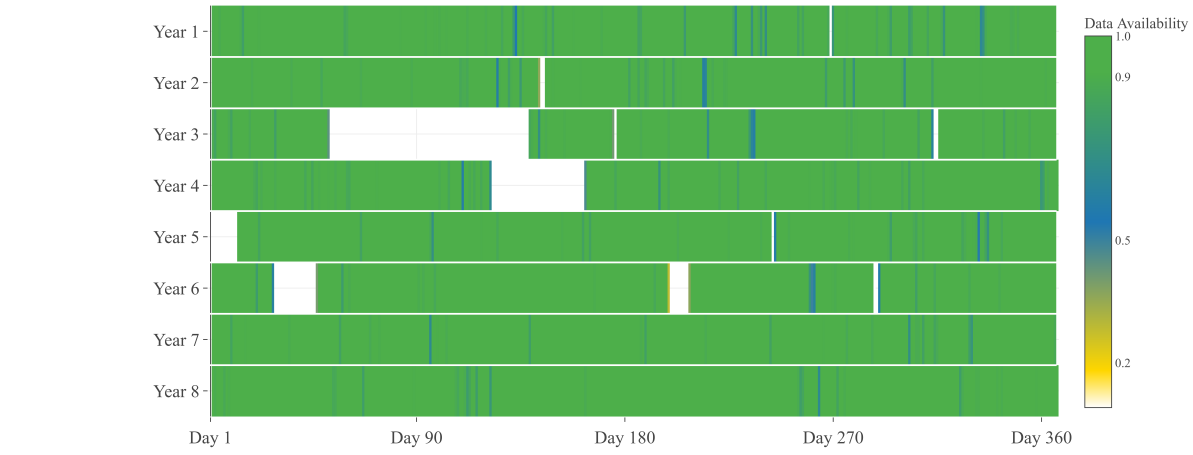

Figure 5. Daily strain data availability over eight monitoring years. Each row corresponds to one year, with color indicating the fraction of available data per day. Values close to unity (green) denote near-complete availability, while lower values (yellow to blue) and white gaps indicate reduced or missing data.

.

The resulting 10-minute damage d_t is computed as:

$$d_t = \sum_{\ell=1}^{N_\sigma} \frac{n_{t,\ell}}{N_f(\Delta\sigma_{t,\ell})} = \sum_{\ell=1}^{N_\sigma} \frac{n_{t,\ell}}{\bar{a}\,(\Delta\sigma_{t,\ell})^{-m}}\,, \tag{19}$$

where $\Delta\sigma_{t,\ell}$ and $n_{t,\ell}$ denote the stress range and associated cycle count in stress-range bin ℓ obtained from rainflow counting within the 10-minute interval t, N_σ is the number of stress-range bins, and $N_f(\Delta\sigma) = \bar{a}(\Delta\sigma)^{-m}$ is the S-N curve expressed as the number of cycles to failure at stress range $\Delta\sigma$ with slope m and intercept $\bar{a}$. All calculations are performed using the Py-fatigue Python package (D'Antuono et al., 2022).

4.3 Experimental protocol

In this section, we describe the experimental protocol followed in this study. Consistent with the notation in Sec. 3, the PI is the full eight-year record $\mathcal{P} = \{1, \ldots, N_{\mathrm{PI}}\}$ for which the EOC history $\{\mathbf{x}_t\}_{t\in\mathcal{P}}$ is available, while measurement windows $\mathcal{M} \subset \mathcal{P}$ are contiguous subsets of $\mathcal{P}$ for which paired observations $\{(d_t, \mathbf{x}_t)\}_{t\in\mathcal{M}}$ are available. For each experiment, we fit a damage mapping using data in $\mathcal{M}$ only, then estimate the PI mean damage by reweighting with the fixed PI EOC distribution, and finally evaluate performance using the normalized error metrics. All analyses are reported for three damage signals derived from the strain gauges: fore-aft (FA), side-side (SS), and the sensor heading S5. For each direction, we apply the same estimation workflow and the same definitions of ground-truth PI mean and error metrics, enabling direct comparisons across damage directions.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

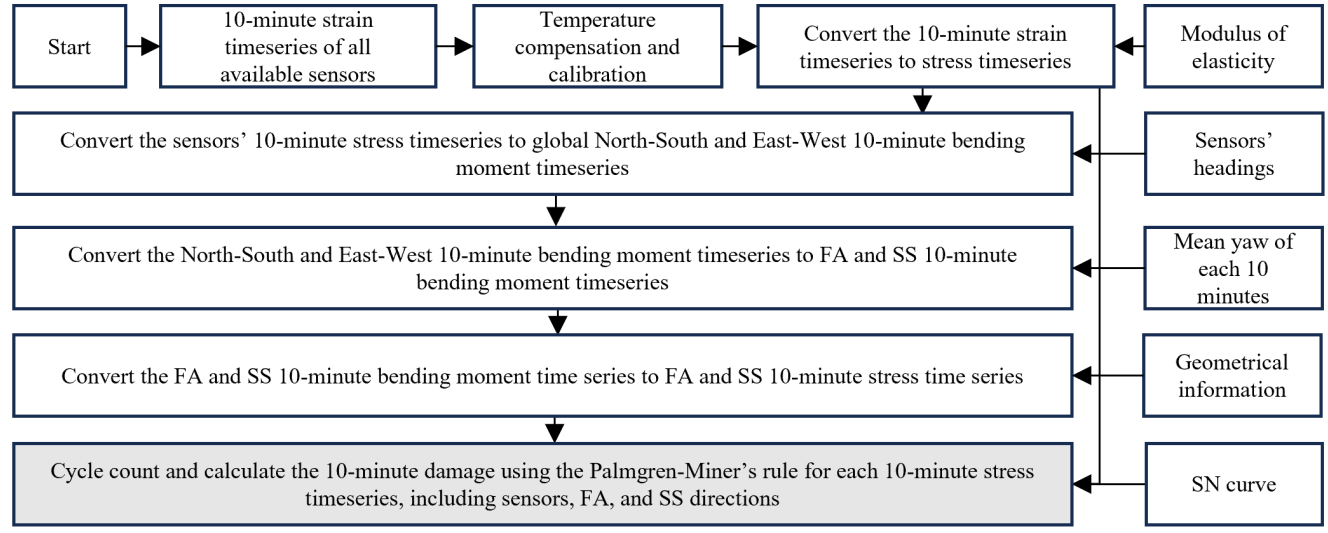

Figure 6. Flowchart for the computation of 10-minute fatigue damage in the fore-aft (FA) and side-side (SS) directions from strain sensor measurements. Starting from 10-minute strain time series of all available sensors, the procedure includes temperature compensation and calibration, followed by conversion to stress. The stress signals are then combined to obtain global North–South and East–West bending moment time series, which are subsequently transformed into FA and SS components using sensor headings and the mean yaw angle. These bending moments are converted back to stress time series based on geometrical properties, after which cycle counting is performed and damage is computed using the Palmgren–Miner rule and the corresponding S–N curve.

We consider the unconditional 0D extrapolation approach and the hierarchy of conditional binning models introduced in Sec. 3. Conditioning variables are selected from wind speed (WS), operational case (Case), wind direction (WD), and turbulence intensity (TI), yielding the following model set: 0D; 1D-WS; 1D-Case; 2D-Case,WS; 3D-Case,WS,WD; 3D-Case,WS,TI; and 4D-Case,WS,WD,TI. Unless stated otherwise, bins are defined using widths of 2 m/s for WS, 30° for WD, and 0.1 for TI. Operational cases include operating (positive power production), non-operating (zero or negative power production), and rare or transitional events. Empty bins within a measurement window are handled using the hierarchical imputation rule defined in Eq. 6. Whenever a conditional bin has no observations in $\mathcal{M}$ ($N_{m,i} = 0$), its statistic is imputed from its parent bin at the next coarser conditioning level.

To quantify interannual variability and the sensitivity to the specific monitoring period, we first perform a sliding-window analysis over the eight-year dataset. We construct $K = 32$ time-consecutive measurement windows $\{\mathcal{M}^{(k)}\}_{k=1}^{K}$ of fixed duration $L = 1$ year, using a circular sliding-window scheme with a step size of $\Delta = 90$ days (Figure 3). For each window $\mathcal{M}^{(k)}$ and each model configuration, we estimate the conditional damage mapping using data in $\mathcal{M}^{(k)}$ only, estimate the PI mean damage over the common $\mathcal{P}$, and compute the corresponding normalized error. This protocol yields, for each model and damage direction, an empirical distribution of errors across the 32 monitoring windows. As a complementary evaluation that avoids overlap between monitoring periods, we repeat the same procedure using eight consecutive annual blocks as measurement win-

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

dows, denoted $\mathcal{M}^{(k)}$ for $k \in \{1,5,9,13,17,21,25,29\}$. Each $\mathcal{M}^{(k)}$ contains all available 10-minute damage observations and EOCs within that year. For each year, we fit the mapping on $\mathcal{M}^{(k)}$, estimate the PI mean over the common $\mathcal{P}$, and compute the corresponding normalized error.

To examine how monitoring duration affects damage estimation, we extend the sliding-window analysis to window lengths $L \in \{1,2,3\}$ years. For each duration, we again construct $K = 32$ windows using the same circular sliding scheme with step $\Delta = 90$ days, to preserve a constant number of windows across durations. For each window, model, and damage direction, we estimate the PI mean over $\mathcal{P}$ and compute the normalized error, enabling a consistent comparison of error distributions as a function of L. We additionally isolate the effect of discretization in the 1D-WS model by varying the wind-speed bin width while keeping all other settings unchanged. Specifically, we test WS bin widths in $\{0.1, 0.5, 1, 2, 4, 8\}$ m/s. This sensitivity analysis is performed using the eight consecutive annual measurement windows $\mathcal{M}^{(k)}$ and focuses on the S5 damage direction. For each bin width and each year, we fit the 1D-WS mapping on $\mathcal{M}^{(k)}$, estimate the PI mean over $\mathcal{P}$, and compute the normalized error. To quantify statistical uncertainty induced by finite strain data within a fixed monitoring window, we apply the *BootWhole* bootstrap strategy with $N_{\text{boot}} = 1000$ replicates. We perform this analysis for the 0D and 1D-WS models using (i) each of the eight consecutive annual windows $\mathcal{M}^{(k)}$ and (ii) a set of eight two-year windows constructed with a one-year step size. In all cases, bootstrap resampling is performed within the chosen $\mathcal{M}$, the estimator is recomputed for each replicate, and empirical 95% confidence intervals are obtained from the corresponding bootstrap quantiles (Eq. 13).

Finally, to investigate which regions of the EOC space drive the observed interannual variability, we perform a hierarchical, bin-wise attribution analysis on the S5 damage signal using the eight consecutive annual measurement windows $M^{(k)}$, $k \in \{1,5,9,13,17,21,25,29\}$. We begin with the 1D-WS mapping and rank wind speed bins by their importance score (Eq. 17), identifying those that contribute most strongly to the total normalized error e_{norm}. For the highest-ranked wind speed bins, we refine the partition by conditioning on operational cases, yielding 2D-(Case,WS) bins and repeating the same ranking procedure. We further subdivide the most influential 2D-(Case,WS) bins by turbulence intensity to obtain 3D-(Case,WS,TI) bins and quantify within-bin discrepancies. Across all hierarchy levels, we compute (i) the bin importance via Eq. 17 and (ii) the normalized bin-level error via Eq. 18. Empty bins within each measurement window are handled using the hierarchical imputation rule (Eq. 6) before evaluating importance scores or bin-level errors.

5 Results and discussion

This section reports the empirical accuracy and variability of fatigue damage estimation methods when only a limited strain measurement window is available. Using the eight-year dataset as the period of interest (PI), we (i) quantify interannual variability of the estimated PI mean damage across multiple measurement windows, (ii) analyze the dependence of this variability on the chosen conditioning variables and damage direction, and (iii) distinguish the window-to-window variability from the statistical uncertainty typically estimated via bootstrapping.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

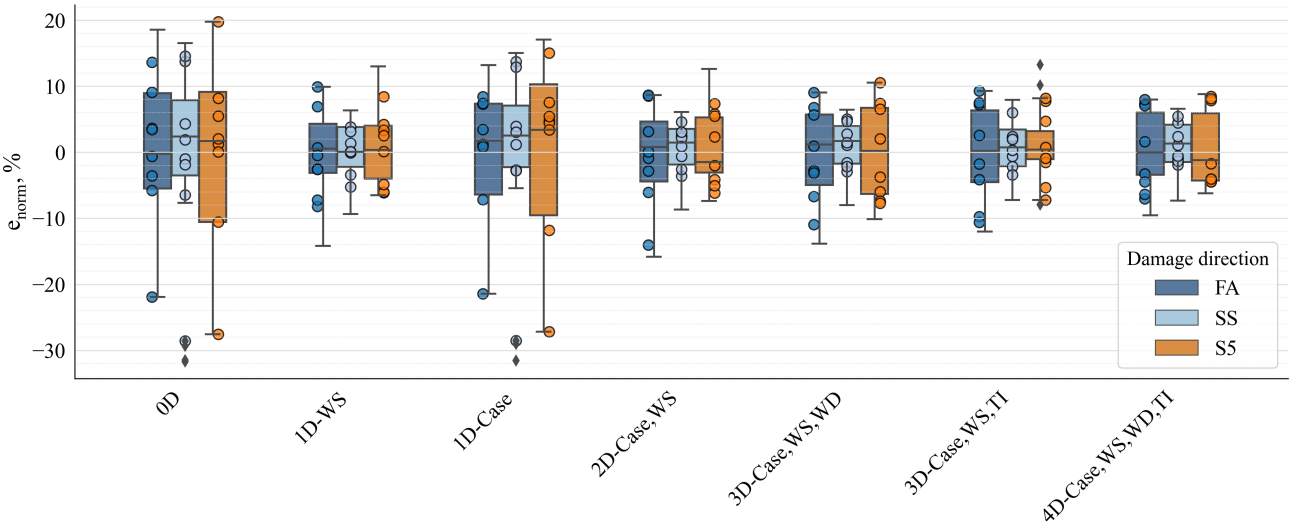

Figure 7. Normalized error of the mean fatigue damage in the fore-aft (FA), side-side (SS), and S5 directions for the considered binning strategies. Box plots show the error distribution across the 32 yearly windows. For each strategy and direction, the box represents the interquartile range (IQR, defined as the 25th–75th percentiles), the central line indicates the median, and the whiskers extend to values within $1.5 \times$ IQR. Individual markers denote the eight consecutive, non-overlapping yearly windows.

5.1 Interannual variability

5.1.1 Sliding window analysis

Figure 7 illustrates the distribution of normalized errors, e_{norm}, obtained when the conditional damage model is fitted on different one-year measurement windows. For each binning strategy, the box plots represent the variability across the sampled windows, and the circle markers indicate the errors obtained from the eight consecutive calendar-year windows. The spread of the error distribution (i.e., interannual variability) depends strongly on the selected mapping model. For the unconditional linear extrapolation model (0D), the estimated mean can deviate more than $\pm$ 20%, depending on the considered annual window, showing that simple extrapolation is sensitive to interannual changes in operating and environmental conditions (Ziegler and Muskulus, 2016b). Conditioning on wind speed (1D-WS) reduces this spread but does not eliminate it. Even after correcting for wind-speed distribution shifts, some windows still yield deviations exceeding $\pm 10\%$. This indicates that wind-speed conditioning alone does not capture the full variability observed across measurement windows. Similar observations are mirrored in recent literature. Hübler and Rolfes (2022) reported errors of up to 35% when extrapolating to a different year using 13 overlapping annual blocks, while Sadeghi et al. (2024) identified discrepancies of up to 40% between the mean damage of six individual years, even when restricting the analysis to nominal operations. Higher-dimensional conditioning can further reduce interannual variability for some configurations (e.g., 4D-Case,WS,WD,TI), but improvements are not monotonic with increas-

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

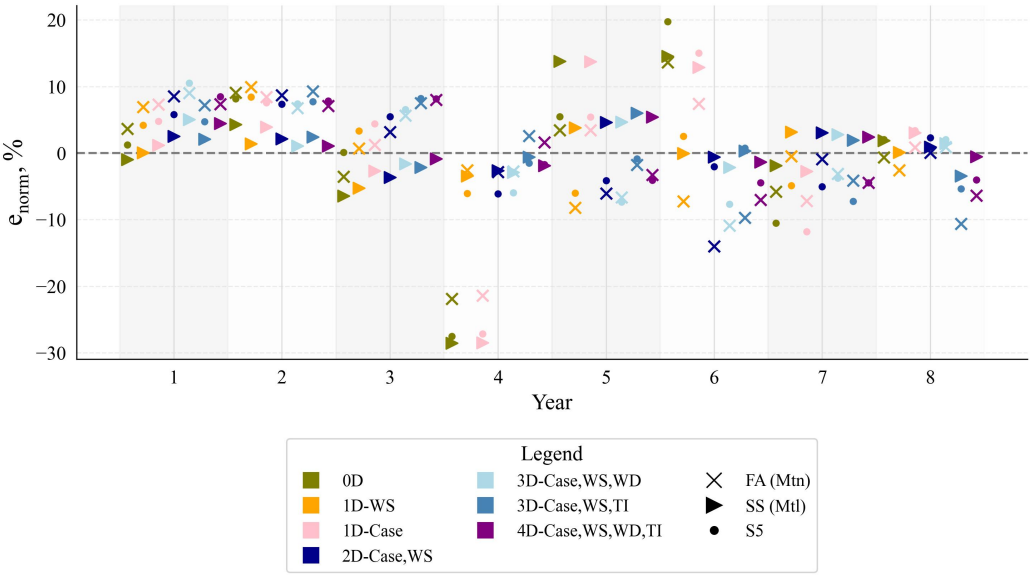

Figure 8. Normalized error of the mean fatigue damage for the considered binning strategies and directions, evaluated separately for each of the eight consecutive, non-overlapping yearly windows. Each point corresponds to one year–strategy–direction combination. Colors indicate the binning strategy (dimensionality and the selected features), while marker shapes indicates the damage direction (FA, SS, and S5). The horizontal dashed line marks zero error.

ing dimension. In particular, the benefit of additional conditioning variables depends on whether they capture the dominant drivers of variability for the considered turbine and direction, and on whether sparsity effects counteract the potential gains.

Taken together, the residual spread observed for 1D-WS indicates that the interannual variability of e_{norm} cannot be attributed to wind-speed distribution shifts alone. In the decomposition introduced in Sec. 3, this residual reflects the combined influence of (i) uncontrolled EOCs, which modulate loads even at comparable wind speeds (e.g., direction-, turbulence-, or operational-driven effects, among others), and (ii) limited sampling of influential, rare high-damage episodes within a finite measurement window. Consequently, approaches that adjust only for wind speed (Hübler and Rolfes, 2022; Noppe et al., 2020) can underestimate variability in fatigue damage: two annual windows with similar wind-speed histograms may still yield markedly different estimated PI means if their joint distribution over other load-driving EOCs differs, or if one window undersamples rare but consequential conditions. This also clarifies why increasing the binning dimension does not guarantee a monotonic reduction in interannual variability. Adding conditioning variables can reduce the contribution of uncontrolled EOCs to e_{norm} when the variables are informative for the considered direction, but it simultaneously increases sparsity and the reliance on the bin-filling (empty-bin imputation) mechanism. When many bins are sparsely populated, variance and imputation-induced bias can offset the gains from additional conditioning, and model performance becomes sensitive to the specific choice of EOCs and discretization settings rather than the nominal dimension alone.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 1. Normalized error of the mean fatigue damage across the eight consecutive, non-overlapping yearly windows. For each binning strategy and damage direction, the mean $\bar{e}_{\text{norm}}$ and standard deviation $\text{std}(e_{\text{norm}})$ are computed from the eight year-wise errors.

Binning strategy	$\bar{e}_{\text{norm}}$ (FA)	$\bar{e}_{\text{norm}}$ (SS)	$\bar{e}_{\text{norm}}$ (S5)	$\text{std}(e_{\text{norm}})$ (FA)	$\text{std}(e_{\text{norm}})$ (SS)	$\text{std}(e_{\text{norm}})$ (S5)
0D	-0.3	-0.4	-0.2	10.1	12.7	13.0
1D-WS	-0.4	0.0	0.2	5.9	2.9	5.0
1D-Case	0.0	0.1	0.2	9.4	12.3	12.5
2D-Case,WS	-0.4	0.8	0.4	7.1	2.7	5.0
3D-Case,WS,WD	-0.1	1.0	0.2	6.5	2.9	6.9
3D-Case,WS,TI	0.0	0.8	0.8	7.3	2.8	5.3
4D-Case,WS,WD,TI	0.4	1.1	0.7	6.0	2.6	5.8

5.1.2 Calendar-year window analysis

Figure 8 compares year-wise errors across the three damage directions (FA, SS, S5) using the eight calendar-year windows. Errors vary substantially with both the selected year and the damage direction. For example, for FA under the 1D-WS binning model, the error ranges from approximately +10% in one year (i.e, year 2) to -10% in another (i.e., year 5). Differences between directions are similarly not negligible and can exceed 10% for the same year and model, indicating that directional effects can influence both the magnitude and the variability of damage estimation errors. We observe that with an appropriate binning method, such as 1D-WS, the error across all directions generally remains within 5–10%. Across models, several approaches maintain year-wise errors within approximately 5-10% for many calendar years and directions (e.g., 1D-WS), consistent with earlier reports that wind-speed conditioning can reduce damage estimation errors (Hübler and Rolfes, 2022). However, unfavorable combinations of measurement year and model choice yield errors exceeding 10%, underscoring that a single measurement window (e.g., 2D-Case,WS; FA; year 6) is not sufficient to characterize accurate damage estimation.

Table 1 summarizes the mean error metrics and variability of the year-wise errors in Figure 8, reporting normalized mean and standard deviation errors across the eight calendar years and for each damage direction. The variability differs by damage direction, with SS generally exhibiting smaller spreads than FA among the better-performing binning models, while S5 is intermediate. Conditioning on the operational case alone (1D-Case) does not improve performance relative to 0D and exhibits large variability, suggesting that the operational case is too coarse to explain the dominant load variability. In general, higher-dimensional binning strategies (3D or 4D) often reduce the standard deviation of year-wise errors relative to 0D and 1D-Case, and they are broadly comparable to 1D-WS. These results indicate that additional conditioning variables can reduce interannual variability in some settings, but that the benefit is limited by sparsity and by the relevance of the selected variables to the dominant variability mechanisms.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

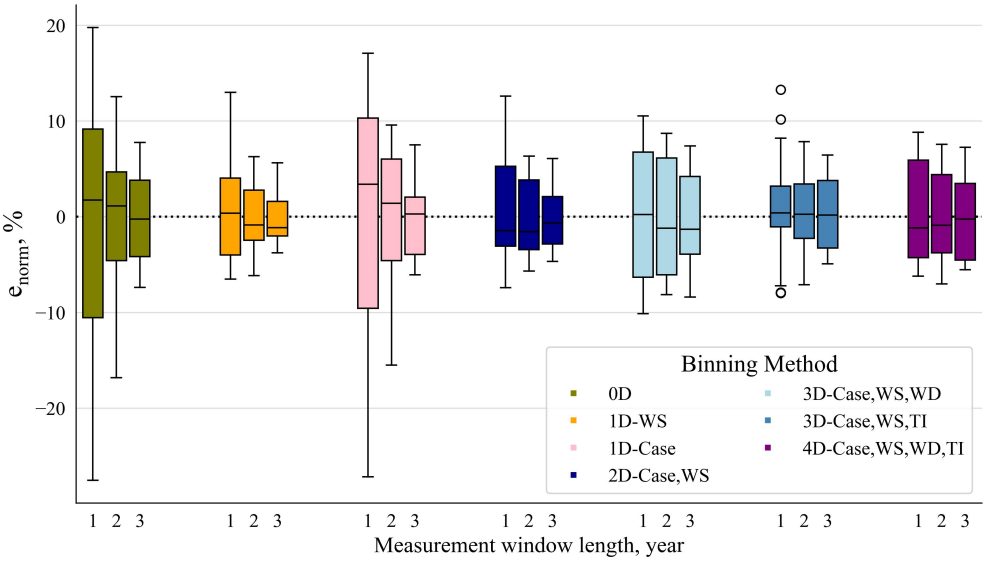

Figure 9. Effect of measurement window length on the normalized mean error of fatigue damage in the S5 direction. For each binning strategy, box plots summarize the error distribution for window lengths of 1, 2, and 3 years based on 32 windows.

5.1.3 Effect of monitoring duration

Figure 9 illustrates the effect of measurement duration, $L \in \{1, 2, 3\}$, on the distribution of normalized mean errors for S5. For low-dimensional mapping strategies (i.e., 0D–2D), increasing the strain monitoring duration is associated with a clear reduction in error dispersion. This reflects a reduction in sampling variability within coarse bins, as longer monitoring periods provide more observations per bin and improve the stability of bin-wise mean damage estimates. In contrast, for higher-dimensional binning strategies (i.e, 3D–4D), extending L only results in modest changes in the error spread. Overall, these results indicate that longer monitoring can reduce window-to-window variability for coarse mappings, whereas the benefits are less pronounced for more finely discretized mappings in this setting. This aligns with the conclusion in Mujtaba et al. (2026), which showed that higher dimensionality is favorable at smaller datasets, but has limited return when the dataset grows.

As the measurement duration increases, a wider spectrum of EOCs is collected, which typically reduces the contribution of finite-data noise and narrows the interquartile range of errors. However, longer monitoring does not inherently guarantee representativeness of the period of interest (PI); even multi-year windows can deviate from the ground-truth PI EOC distribution, leaving a residual interannual variability. This nuance challenges the conventional assumption found in several studies, which suggests that a period of several months or a few years may be sufficient for reliable fatigue assessment (Hübler and Rolfes, 2022; Hübler et al., 2018; Marhadi and Skrimpas, 2015; Marsh, 2016). We show that multi-year datasets may still yield

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

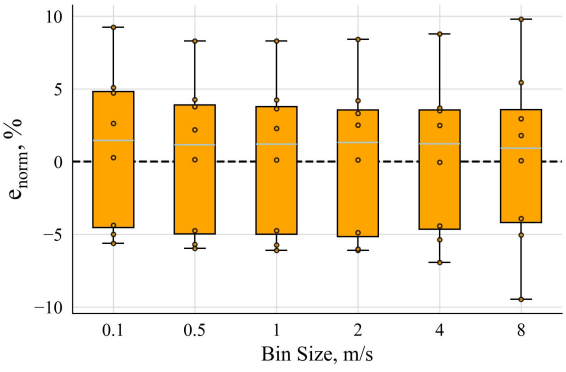

Figure 10. Effect of wind speed bin size on the normalized error of fatigue damage for the 1D-WS strategy in the S5 direction. For each bin size, box plots summarize the distribution of errors across the eight consecutive, non-overlapping yearly windows, with individual markers indicating the corresponding yearly values. The boxes represent the interquartile range (IQR, defined as the 25th–75th percentiles), the grey line denotes the mean error, and the whiskers extend to values within $1.5 \times$ IQR.

inaccurate estimates for mapping methods when window-to-window variability (limited representativeness) remains dominant, even as finite-sample variability decreases with longer monitoring.

5.1.4 Sensitivity to wind-speed discretization.

Figure 10 isolates the effect of wind-speed discretization in the 1D-WS model for S5. Varying the wind-speed bin width from 0.1 to 8 m/s has only a minor effect on the error distribution across the 8 calendar-year windows. Extremely small bins increase sparsity, while very large bins reduce the ability to correct for wind-speed distribution shifts. For this case study, intermediate bin widths like 1 or 2 m/s provide a reasonable trade-off. This limited sensitivity suggests that, once bin widths are within a reasonable range, residual interannual variability is dominated less by the discretization itself and more by (i) variability driven by EOCs not captured by wind speed alone and (ii) window-to-window differences in the occurrence of rare, high-damage conditions. Practically, discretization should therefore be chosen to maintain adequate occupancy in the bins that carry most probability mass under the PI distribution, and it should be reported alongside the empty-bin handling rule, as these two choices jointly determine the effective bias-variance trade-off of the binned estimator.

5.1.5 Variability in conditional fatigue damage estimates

Figure 11 shows the normalized relative bin-wise error ($e_{norm,i}$) for the S5 damage direction across wind speed bins as scatter points, together with the corresponding bin importance (I_i) shown as gray bars. The importance metric quantifies the average contribution of each bin to the total normalized error (e_{norm}). Empty bins are imputed using the 0D mean, as defined in Eq. 6. Although high wind speed bins exhibit larger within-bin variability, their contribution to the total error is negligible due to their low probability of occurrence. From this analysis, the bins centered at 8 and 16 m/s are identified as having the highest

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

importance. For instance, the 8 m/s bin contributes on average approximately 1.6 % to the total error of the 1D-WS model, for which the overall error is on the order of $\pm 10\%$ (Figure 7).

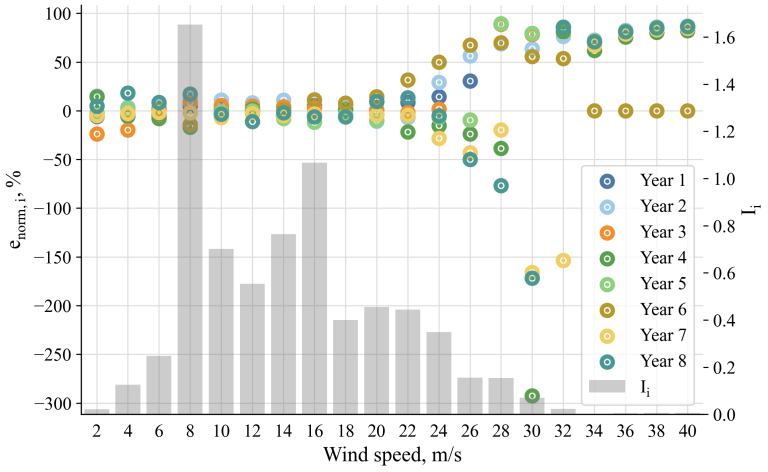

Figure 11. Local normalized error of fatigue damage as a function of wind speed bin, together with the corresponding bin importance. Colored markers indicate the error contribution in each wind speed bin for the eight consecutive yearly windows, while gray bars represent the relative importance I_i of each bin. The left axis shows the normalized error per bin, and the right axis reports the associated bin importance.

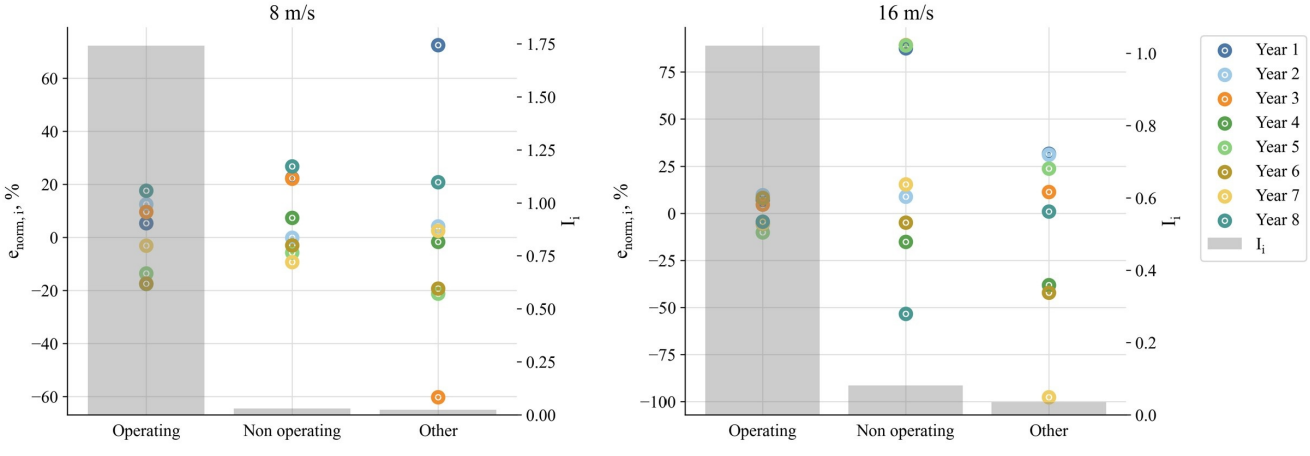

Figure 12. Local normalized error of fatigue damage across operational case bins and corresponding bin importance for wind speeds of 8 m/s (left) and 16 m/s (right). Colored markers represent the error in each operational case bin for the eight consecutive yearly windows, while gray bars indicate the relative importance I_i of each bin. The left axis shows the normalized error per bin, and the right axis reports the associated bin importance.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Figure 12 shows that the wind speed bins centered at 8 and 16 m/s, which exhibit the highest importance, also display substantial local error variability ($e_{norm,i}$) when an additional conditioning dimension based on operational case (2D-Case,WS) is introduced. Among the three operational cases, the operating case has the highest importance, whereas the non-operating and other cases contribute negligibly. Focusing on the operating case, the local error variability remains on the order of $\pm 20\%$. Extending the conditioning to include turbulence intensity (3D-Case,WS,TI), and restricting the analysis to the most important bins (WS=8 or 16 m/s under operating conditions), Figure 13 shows the corresponding variability across turbulence intensity bins. The bins corresponding to turbulence intensities of 0.1 or 0.2 exhibit the highest importance and continue to show a wide range of local errors, again on the order of $\pm 20\%$. These results indicate that increasing the number of conditioning variables does not eliminate the observed variability. Despite the inclusion of additional EOCs, a substantial residual variability remains (see comparison of 1D-WS, 2D-Case,WS, and 3D-Case,WS,TI in Figure 7). This residual variability suggests that, within the most important bins, unobserved or uncontrolled conditions vary across years and contribute significantly to the error (e_{norm}).

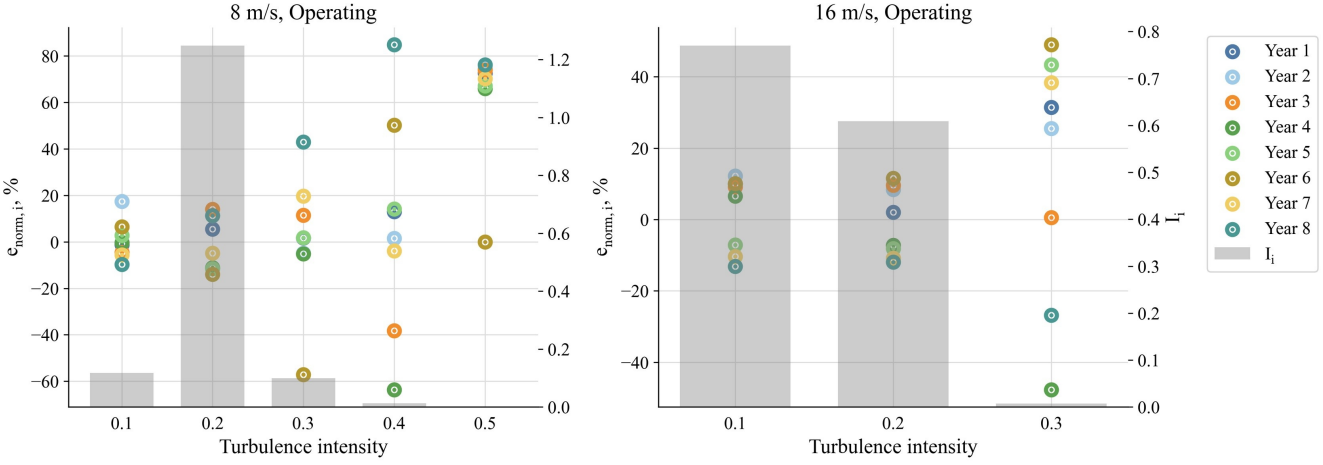

Figure 13. Local normalized error of fatigue damage across turbulence intensity bins under operating conditions, and corresponding bin importance, for wind speeds of 8 m/s (left) and 16 m/s (right). Colored markers represent the error in each turbulence intensity bin for the eight consecutive yearly windows, while gray bars indicate the relative importance I_i of each bin. The left axis shows the normalized error per bin, and the right axis reports the associated bin importance.

5.2 Do narrow bootstrap confidence intervals imply low interannual variability?

Bootstrap confidence intervals are commonly used to quantify uncertainty arising from finite monitoring data. However, bootstrap resampling is conditional on the observed measurement window and therefore primarily reflects sampling variability within that window. To assess whether such intervals also reflect window-to-window variability in the estimated PI mean, we compare bootstrap-based uncertainty estimates with the dispersion of errors obtained across multiple measurement windows. Table 2 reports the standard deviation of the normalized error in the estimated mean damage from bootstrap resampling within

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 2. Comparison of normalized error standard deviations from bootstrapped samples for each individual year, together with the standard deviation of the normalized mean errors across the eight consecutive windows for window lengths of 1 and 2 years. Results are reported for the 0D and 1D-WS binning strategies in the S5 direction. Columns Btrp. (w1) to Btrp. (w8) denote the standard deviation of bootstrapped normalized errors for each yearly window, while std(e_{norm}) reports the standard deviation of the eight normalized mean errors across windows. Values are expressed in percentage.

Strategy	Window length	Btrp. (w1)	Btrp. (w2)	Btrp. (w3)	Btrp. (w4)	Btrp. (w5)	Btrp. (w6)	Btrp. (w7)	Btrp. (w8)	std(e_{norm}) (Table 1)
0D	1 year	0.6	0.5	0.7	0.8	0.7	0.6	0.7	0.6	13
1D-WS	1 year	0.4	0.4	0.4	0.5	0.5	0.5	0.4	0.4	5
0D	2 years	0.4	0.4	0.5	0.5	0.4	0.4	0.4	0.4	8
1D-WS	2 years	0.3	0.3	0.3	0.3	0.4	0.3	0.3	0.3	4

each measurement window, together with the corresponding standard deviation computed across multiple windows without resampling. The latter is based on eight consecutive individual years and eight 2-year sliding windows with a 1-year step. Results are reported for the S5 damage direction and expressed as percentages for both the 0D and 1D-WS models. The results show that the standard deviations obtained from bootstrap resampling are negligible relative to the overall variability observed across measurement windows. This indicates that, when long-term mean damage is estimated from limited-duration measurements, interannual variability constitutes the dominant source of uncertainty, whereas sampling variability due to finite data within a given window plays a comparatively minor role.

Figure 14 reports normalized errors for the 0D and 1D-WS models based on (left) eight non-overlapping calendar-year windows and (right) eight two-year windows. For each window, 95% bootstrap confidence intervals are computed using $N_{\text{boot}} = 1000$ replicates under the BootWhole scheme (Sadeghi et al., 2024). The bootstrap intervals are consistently narrow across windows, indicating limited sampling variability within each window. However, the corresponding point estimates of the PI mean exhibit substantial variation across windows, particularly for the 0D model. In several cases, windows associated with large errors yield narrow intervals that do not include the PI mean. This behavior reflects that bootstrap resampling quantifies estimator variability conditional on the observed data, but does not capture the dominant window-to-window variability in e_{norm}. Increasing the monitoring duration (e.g., from one to two years) further reduces the width of bootstrap intervals, confirming that the window-conditional sampling variability decreases with additional data. However, this reduction does not imply a corresponding decrease in window-to-window variability. Bootstrap confidence intervals are appropriate for assessing sampling uncertainty within $\mathcal{M}$, but they should not be interpreted as reflecting interannual variability when extrapolating limited monitoring data to the PI.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

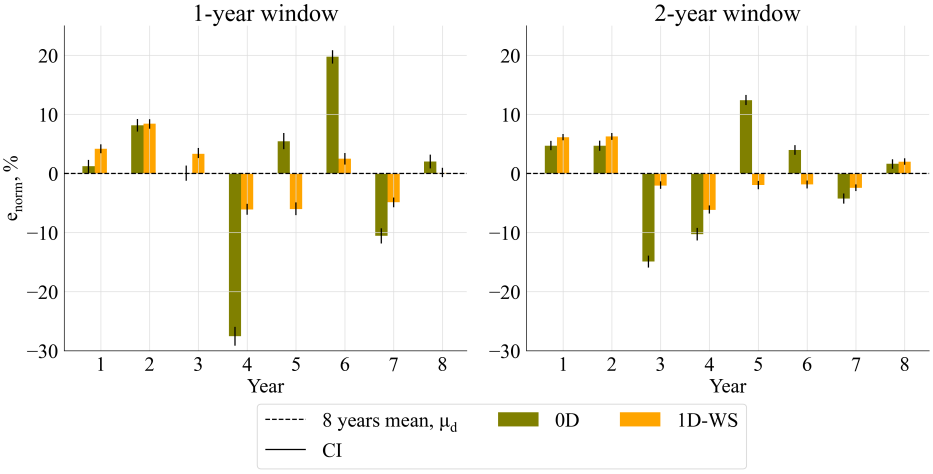

Figure 14. Variability of normalized total error of the estimated mean fatigue damage for two windowing schemes: eight consecutive individual years (left) and eight 2-year windows with a 1-year step (right), over an 8-year estimation period. Results are shown for the 0D and 1D-WS binning strategies. For each window, bars indicate the normalized error relative to the 8-year reference mean μ_d, and vertical lines denote the corresponding 95% confidence intervals obtained via bootstrapping.

6 Conclusions

In this work, we examine whether fatigue damage estimated from a short strain monitoring campaign can be accurately mapped to a longer period of interest (PI) for which only SCADA-derived environmental and operational conditions (EOCs) are available, and identify the modeling conditions under which such mapping is justified. Using an eight-year measurement campaign, we quantify the accuracy of PI mean damage estimates obtained from limited measurement windows and compare commonly used damage mapping strategies based on discretized (binned) conditional models. Beyond reporting average errors, we characterize how damage estimation accuracy depends on the considered monitoring window, damage direction, selected conditioning variables, and discretization settings.

A central result is that short monitoring windows can induce substantial interannual variability in the estimated PI mean fatigue damage. This variability arises from year-to-year differences in the conditional damage-EOC mapping over yearly windows, rather than from finite-sample effects within a fixed window. In particular, limited sampling of rare, high-damage events and the influence of unobserved or uncontrolled drivers lead to window-dependent estimates of the conditional mean damage. In contrast, bootstrap-based statistical uncertainty estimates are conditional on the selected measurement window and therefore capture only the sampling variability associated with the finite data available in that window. Consequently, narrow bootstrap confidence intervals may coexist with pronounced differences between monitoring windows, and do not provide evidence that a given year provides a representative estimate of the PI mean damage. Increasing the monitoring duration generally reduces within-window sampling variability and mitigates sensitivity to individual years by improving the coverage

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

of damaging conditions. However, the results show that multi-year windows do not guarantee small estimation errors when the conditional damage-EOC is not stationary over time.

The goal of damage mapping is to estimate a conditional relationship between EOCs and short-term damage for which the mapping derived from the measurement window provides a valid estimate when applied over the PI. In our study, unconditional extrapolation (0D) yields large errors for certain yearly windows, with deviations reaching up to 30% in the estimated PI mean of 10-minute damage. Low-dimensional conditioning reduces these errors but may leave substantial variability unresolved. Conditioning on informative EOCs, especially wind speed and its combinations with wind direction and turbulence intensity, generally reduces sensitivity to the choice of monitoring year, with errors often on the order of 10% for the better-performing configurations. However, these improvements are not monotonic with increasing model dimensionality. High-dimensional binning remains susceptible to variability across years and becomes sensitive to discretization due to sparsity and empty-bin imputation. These results indicate that a fraction of interannual variability persists even after conditioning on commonly used EOCs, either because relevant drivers are not captured or because the conditional damage-EOC relationship is not stationary over time.

Although this study focuses on a specific family of discretized estimators, the limitations associated with short monitoring windows are not specific to binning-based approaches. Any data-driven mapping derived from a limited measurement window remains sensitive to the representativeness of that window, as well as to residual variability arising from unobserved or insufficiently characterized EOCs and rare high-damage events. Future work may investigate the transferability of damage mapping across turbines, farms, and operating regimes to distinguish site-specific effects from consistent patterns. In addition, Bayesian updating frameworks that assimilate monitoring data sequentially may provide a systematic approach to quantify how uncertainty in PI estimates evolves with increasing monitoring duration and coverage, while incorporating prior information and model discrepancy.

Data availability. The data used in this paper are proprietary to the industrial partners of this project and cannot be made publicly available.

Author contributions. NS: Formal analysis, Software, Investigation, Methodology, Visualization, Writing (original draft preparation). PM: Conceptualization, Validation, Writing (review and editing). NN: Supervision. NH: Writing (review and editing). WW: Validation, Data curation, Writing (review). CD: Supervision, Data curation, project administration.

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

Acknowledgements. The authors would like to express their acknowledgment for the support of the Smartlife project funded by Energy Transition Fund of the Belgian FPS Economy.

We acknowledge Parkwind for their permission to use the monitoring data.

During the preparation of this work, the author used ChatGPT5 to improve the text. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

References

Amiri, A. K., Kazacoks, R., McMillan, D., Feuchtwang, J., and Leithead, W.: Farm-wide assessment of wind turbine lifetime extension using detailed tower model and actual operational history, in: Journal of Physics: Conference Series, vol. 1222, p. 012034, IOP Publishing, 2019.

D'Antuono, P., Weijtjens, W., and Devriendt, C.: Py-Fatigue, https://owi-lab.github.io/py_fatigue, 2022.

DNV: DNV-RP-C203: Fatigue design of offshore steel structures, DNV-Recommended Practice, 2016.

Efron, B.: Bootstrap methods: Another look at the jackknife, in: Breakthroughs in statistics: Methodology and distribution, pp. 569–593, Springer, 1992.

Hlaing, N., Morato, P. G., Santos, F. d. N., Weijtjens, W., Devriendt, C., and Rigo, P.: Farm-wide virtual load monitoring for offshore wind structures via Bayesian neural networks, Structural Health Monitoring, 23, 1641–1663, 2024.

Hübler, C. and Rolfes, R.: Probabilistic temporal extrapolation of fatigue damage of offshore wind turbine substructures based on strain measurements, Wind Energy Science Discussions, 2022, 1–29, https://doi.org/10.5194/wes-7-1919-2022, 2022.

Hübler, C., Weijtjens, W., Rolfes, R., and Devriendt, C.: Reliability analysis of fatigue damage extrapolations of wind turbines using offshore strain measurements, in: Journal of Physics: Conference Series, vol. 1037, p. 032035, IOP Publishing, https://doi.org/10.1088/1742-6596/1037/3/032035, 2018.

Hübler, C., Weijtjens, W., Gebhardt, C. G., Rolfes, R., and Devriendt, C.: Validation of improved sampling concepts for offshore wind turbine fatigue design, Energies, 12, 603, https://doi.org/10.3390/en12040603, 2019.

IEC TS 61400-28: Through life management and life extension of wind power assets, , 2023.

Loraux, C. and Brühwiler, E.: The use of long term monitoring data for the extension of the service duration of existing wind turbine support structures, in: Journal of Physics: Conference Series, vol. 753, p. 072023, IOP Publishing, 2016.

Mai, Q. A., Weijtjens, W., Devriendt, C., Morato, P. G., Rigo, P., and Sørensen, J. D.: Prediction of remaining fatigue life of welded joints in wind turbine support structures considering strain measurement and a joint distribution of oceanographic data, Marine Structures, 66, 307–322, https://doi.org/10.1016/j.marstruc.2019.05.002, 2019.

Marhadi, K. S. and Skrimpas, G. A.: Automatic threshold setting and its uncertainty quantification in wind turbine condition monitoring system, International Journal of Prognostics and Health Management, 6, https://doi.org/10.36001/ijphm.2015.v6i4.2291, 2015.

Marsh, G.: Fatigue load monitoring of offshore wind turbine support structures, Ph.D. thesis, The University of Edinburgh, 2016.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., and Wagenmakers, E.-J.: The fallacy of placing confidence in confidence intervals, Psychonomic bulletin & review, 23, 103–123, https://doi.org/10.3758/s13423-015-0947-8, 2016.

Mozafari, S.: Probabilistic fatigue reliability assessment of the wind turbine's structural components, Ph.D. thesis, Technical University of Denmark, 2023.

Mujtaba, A., Weijtjens, W., Sadeghi, N., and Devriendt, C.: A machine-learning-based approach for better prediction of fatigue life of offshore wind turbine foundations using smaller data sizes, Wind Energy Science, 11, 443–467, 2026.

Noppe, N., Hübler, C., Devriendt, C., and Weijtjens, W.: Validated extrapolation of measured damage within an offshore wind farm using instrumented fleet leaders, Journal of Physics: Conference Series, 1618, 022 005, https://doi.org/10.1088/1742-6596/1618/2/022005, 2020.

Pacheco, J., Pimenta, F., Pereira, S., Cunha, Á., and Magalhães, F.: Fatigue assessment of wind turbine towers: Review of processing strategies with illustrative case study, Energies, 15, 4782, https://doi.org/10.3390/en15134782, 2022.

Pacheco, J., Pimenta, F., Pereira, S., Cunha, A., and Magalhães, F.: Experimental evaluation of strategies for wind turbine farm-wide fatigue damage estimation, Engineering Structures, 285, 115 913, https://doi.org/10.1016/j.engstruct.2023.115913, 2023.

https://doi.org/10.5194/wes-2026-65
Preprint. Discussion started: 20 April 2026

Palmgren, A.: Die Lev/bensdauer von kugellagern, VDI. Z., 68, 339–341, 1924.

Petrovska, E., Le Dreff, J.-B., Oterkus, S., Thies, P., and McCarthy, E.: Application of structural monitoring data for fatigue life predictions of monopile-supported offshore wind turbines, in: International Conference on Offshore Mechanics and Arctic Engineering, vol. 84416, p. V009T09A058, American Society of Mechanical Engineers, 2020.

Sadeghi, N., Robbelein, K., D'Antuono, P., Noppe, N., Weijtjens, W., and Devriendt, C.: Fatigue damage calculation of offshore wind turbines' long-term data considering the low-frequency fatigue dynamics, in: Journal of Physics: Conference Series, vol. 2265, p. 032063, IOP Publishing, 2022.

Sadeghi, N., D'Antuono, P., Noppe, N., Robbelein, K., Weijtjens, W., and Devriendt, C.: Quantifying the effect of low-frequency fatigue dynamics on offshore wind turbine foundations: a comparative study, Wind Energy Science, 8, 1839–1852, https://doi.org/10.5194/wes-8-1839-2023, 2023.

Sadeghi, N., Noppe, N., Morato, P. G., Weijtjens, W., and Devriendt, C.: Uncertainty quantification of wind turbine fatigue lifetime predictions through binning, Journal of Physics: Conference Series, 2767, 032 024, https://doi.org/10.1088/1742-6596/2767/3/032024, 2024.

Socie, D.: Rainflow cycle counting: A historical perspective, The Rainflow Method in Fatigue, pp. 3–10, 1992.

Tatsis, K., Chatzi, E., and Lourens, E.-M.: Reliability prediction of fatigue damage accumulation on wind turbine support structures, in: Proceedings of the 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering, pp. 76–89, National Technical University of Athens (NTUA), 2017.

Ziegler, L.: Assessment of monopiles for lifetime extension of offshore wind turbines, Ph.D. thesis, Norwegian University of Science and Technology (NTNU), 2018.

Ziegler, L. and Muskulus, M.: Fatigue reassessment for lifetime extension of offshore wind monopile substructures, in: Journal of physics: Conference series, vol. 753, p. 092010, IOP Publishing, 2016a.

Ziegler, L. and Muskulus, M.: Lifetime extension of offshore wind monopiles: Assessment process and relevance of fatigue crack inspection, in: Proceedings of the 12th EAWE PhD Seminar on Wind Energy in Europe, pp. 25–27, 2016b.