Probabilistic Temporal Extrapolation of Fatigue Damage of Offshore Wind Turbine Substructures Based on Strain Measurements

Clemens Hübler[1] and Raimund Rolfes[1]

[1]Leibniz Universität Hannover, Institute of Structural Analysis, ForWind, Appelstr. 9a, D-30167 Hannover, Germany

Correspondence: Clemens Hübler (c.huebler@isd.uni-hannover.de)

Abstract. Substructures of offshore wind turbines are becoming older and beginning to reach their design lifetimes. Hence, lifetime extensions for offshore wind turbines are becoming not only an interesting research topic, but also a relevant option for industry. To make well-founded decisions on possible lifetime extensions, precise fatigue damage predictions are required. In contrast to the design phase, fatigue damage predictions cannot only be based on aero-elastic simulations but also on strain measurements. Nonetheless, strain measurement-based fatigue damage assessments for lifetime extensions have been rarely conducted so far. Simulation-based approaches are much more common, although current standards explicitly recommend the use of measurement-based approaches as well. For measurement-based approaches, the main challenge is that strain data are limited. This means that measurements are only available for a limited period and only at some specific hot-spot locations. Hence, spatial and temporal extrapolations are required. Available procedures are not yet standardised and in most cases not validated. This work focuses on extrapolations in time. Several methods for the extrapolation of fatigue damage are assessed. The methods are intended to extrapolate fatigue damage calculated for a limited time period using strain measurement data to a longer time period, or another time period, where no such data are available. This could be, for example, a future period, a period prior to the installation of strain gauges or a period after some sensors have failed. The methods are validated using several years of strain measurement data from the German offshore wind farm Alpha Ventus. The performance and user friendliness of the various methods are compared. It is shown that fatigue damage can be predicted accurately and reliably for periods where no strain data are available. Best results are achieved if wind speed correlations are taken into account by applying a binning approach, and if a least some winter months of strain data are available.

1 Introduction

Although offshore wind energy is considered a relatively young industry, the oldest offshore wind turbines (OWTs) are operating for more than 20 years. Some OWTs have even already been decommissioned (Topham and McMillan, 2017). In the upcoming years, many OWTs will reach their predicted design lifetimes. This leads to the question of what to do with these OWTs: continue to run them with a lifetime extension, conduct a repowering – i.e., replace the old turbine with a new one, or just decommission them? Since the substructures of many old OWTs are expected to be overdesigned, lifetime extensions are a viable option for them. Every year of additional operation beyond the expected lifetime can be fairly profitable (Rubert et al.,

25   2019), as all debts, etc., are already paid back.

To enable safe and profitable lifetime extensions, the remaining useful lifetimes of the OWTs need to be determined. For this purpose, a first international standard for lifetime extensions for wind turbines has already been introduced (DNVGL-ST-0262, 2016). It proposes lifetime extensions based on a combination of inspections and renewed fatigue damage simulations using an updated design model. This updated simulation model uses, for example, the measured wind conditions (i.e., SCADA data)

30   during the entire lifetime instead of the wind conditions assumed during the design phase. Hence, wind conditions can be represented more realistically in the simulation. This leads to a better – in most cases less conservative – lifetime estimation. Such simulation-based fatigue reassessments of OWT substructures are investigated, for example, by Ziegler and Muskulus (2016a, b), Bouty et al. (2017), Natarajan and Bergami (2020), Saathoff and Rosemeier (2020), or Nielsen et al. (2021). Nielsen et al. (2021) even include economic effects in their risk-informed, simulation-based fatigue reassessment.

35   However, frequently, not only the wind conditions are known from SCADA data, but additional data are available. For some OWTs, strain gauges at different relevant positions of the substructure measure the real load conditions the OWT is exposed to (Weijtjens et al., 2016). If these load conditions are known from measurements, a strain measurement-based fatigue damage assessment can be conducted. It can enhance or even replace the simulation-based assessment. On the one hand, this has the advantage that the uncertainty of the simulation model does not have to be taken into account. On the other hand, measurement

40   uncertainties become relevant. In contrast to simulation-based remaining lifetime estimations – which have become quite popular in recent years – strain measurement-based remaining lifetime estimations for OWTs are rarely conducted so far, although the standard (DNVGL-ST-0262, 2016) recommends the use of measured load data if available.

Independent of the context of lifetime extensions, the first fatigue damage estimations based on measured strain data were conducted in the 1990s (Seifert, 1995; Söker, 1996). Due to the increasing relevance of lifetime extensions, measurement-

45   based fatigue damage calculations have again become a research focus, as they are a valuable addition to simulation-based analyses. However, an important obstacle for such analyses is the limited strain data (Petrovska et al., 2020). First, strain data have normally not been collected for the entire lifetime of the turbine, but only for a limited period (Louraux and Brühwiler, 2016; Hübler et al., 2018). Strain gauges frequently fail after some time and are not replaced or have not been applied directly during the construction but in a later project phase. That is why temporal extrapolation methods are required to estimate fa-

50   tigue damage for periods where no strain data are available or even for future periods. Second, strain gauges are only placed at a few important positions (e.g., hot spots) on the turbine or substructure. For all other positions, spatial extrapolations are required. This is why strain measurement-based approaches can be divided into three method types: no extrapolation, spatial extrapolation and temporal extrapolation.

Early approaches by Seifert (1995) and Söker (1996) did not apply any extrapolation approaches. They focused on the fa-

55   tigue load determination for a directly measured data set. There are also some more recent approaches (Smith et al., 2014; Marsh, 2016; Petrovska et al., 2020) that do not apply any extrapolations. They concentrate instead on correlations with environmental and operational conditions (EOCs) and the uncertainty of the fatigue damage calculation. Correlations with EOCs are considered by applying binning approaches. The uncertainty is determined using a statistical resampling technique, i.e., bootstrapping.

60   Spatial extrapolations can either be extrapolations to other positions on the same turbine or even to other turbines in the same wind farm. The former is done, for example, by Ziegler et al. (2017, 2019), Maes et al. (2016), Iliopoulos et al. (2017), or Henkel et al. (2020). Ziegler et al. (2017, 2019) use aero-hydro-elastic simulations and a k-nearest-neighbour regression algorithm to extrapolate to other positions on the same turbine. Maes et al. (2016), Iliopoulos et al. (2017) and Henkel et al. (2020) conduct their extrapolations via modal expansion. Extrapolations to other turbines within a wind farm are conducted by
65   Weijtjens et al. (2016), Noppe et al. (2020) and Santos et al. (2021) by applying so-called "fleet leader concepts".

Finally, extrapolations in time – being the focus of this work – are analysed by Louraux and Brühwiler (2016) and Hübler et al. (2018). Louraux and Brühwiler (2016) calculated a fatigue lifetime by applying a straightforward linear extrapolation in time. Moreover, for future work, they proposed the assessment of a binning approach for two EOCs wind speed and wind direction – similar to the approach of Marsh (2016) – to increase the accuracy of the extrapolation. Hübler et al. (2018) took up the
70   recommendation of Louraux and Brühwiler (2016). They assessed a strain measurement-based fatigue damage extrapolation approach that makes use of correlations between fatigue damage and EOCs. Moreover, the uncertainty of the extrapolation is estimated by applying bootstrapping similarly to the approaches of Marsh (2016) or Petrovska et al. (2020).

Hence, there are first approaches for temporal extrapolations of strain measurement-based fatigue estimations. However, up to now, the available procedures are neither comprehensively validated nor is there a consensus regarding the most suitable meth-
75   ods. The approach of Hübler et al. (2018), which correlates fatigue damage with EOCs, shows reasonable results. However, a comprehensive comparison of it with other available methods (e.g., machine-learning approaches known from simulation-based design (Dimitrov et al., 2018; Müller et al., 2021), lifetime extensions (Dimitrov and Natarajan, 2019; Natarajan and Bergami, 2020) and spatial extrapolations (Santos et al., 2021)) is missing. Moreover, many assumptions and decisions made by Hübler et al. (2018) are only based on expert knowledge. For example, so far, it has not been conclusively analysed which
80   EOCs should be included in the correlation approach (Hübler et al., 2018; Marsh, 2016; Louraux and Brühwiler, 2016; Petrovska et al., 2020). Therefore, in this work, several approaches for strain measurement-based fatigue damage extrapolations in time are analysed in detail. They are compared to each other and validated using real offshore measurement data from the German wind farm "Alpha Ventus". The three method types investigated are simple extrapolations, binning approaches for EOCs and machine-learning approaches. It will be shown that binning approaches yield the best results, especially for non-expert
85   users and limited data. Therefore, for binning approaches, further investigations regarding the most suitable bin types and sizes are conducted. Finally, the required amount of data is analysed.

Extrapolation approaches always feature some uncertainty. Therefore, for all methods considered, not only a deterministic extrapolation is conducted but also a probabilistic one. This enables an estimation of their uncertainty.

The rest of this work is structured as follows. In the next section, the underlying measurements are explained. This includes
90   a description of the measurement setup as well as the presentation of some raw data. Moreover, the applied data processing is illustrated. In Section 3, all methods for the fatigue damage extrapolation are explained. This includes the standardised short-term damage calculation as well as the extrapolation in time – being the focus of this work. In Section 4, results of the comparison and validation are presented. Finally, in Section 5, benefits and limitations of the current work are summarised and in Section 6, an outlook is given.

2 Measurements

2.1 Measurement setup

In this work, offshore data from a measurement campaign in the German "Alpha Ventus" wind farm are utilised. The raw data are freely available for research purposes after signing an agreement concerning the data usage (https://www.rave-offshore.de/en/data.html). Alpha Ventus consists of twelve 5 MW turbines: six Senvion 5M turbines mounted on jackets and six Adwen 5-116 turbines mounted on tripods (see Fig. 1). The wind farm is located about 45 km north of the German island Borkum (see Fig. 2). It has water depths of about 30 m.
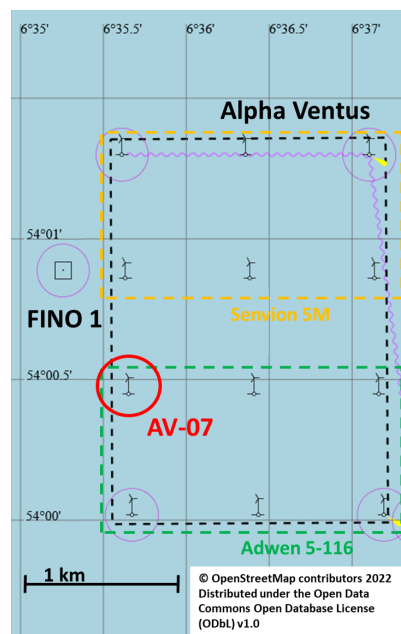
Figure 1. Farm layout of Alpha Ventus with considered AV-07 turbine marked (adapted from OpenStreetMap – www.openstreetmap.org/copyright).

Alpha Ventus was commissioned in April 2010. The measurement campaign started in 2011. Since then, not only SCADA data are being collected, but environmental conditions, strains, accelerations, etc., are measured as well. Further environmental data are available from the metmast FINO1 (https://www.fino1.de/en/). FINO1 is located next to the Alpha Ventus wind farm (cf. Fig. 1). This work focuses on the AV-07 turbine (see Table 1). It is marked in Fig. 1. This turbine is equipped with more than 100 sensors on the rotor-nacelle assembly, the tower and the substructure above and below sea level. Data concerning environmental conditions are available as statistical values – i.e., mean values, standard deviations, maxima and minima – of ten-minute intervals. Strain data are provided as high resolution (50 Hz) time series for several locations (see Fig. 3). As an example, this work uses the strain data from one location on the tower, as marked in Fig. 3. At this location, four strain gauges are placed around the circumference of the tower (see Fig. 4).
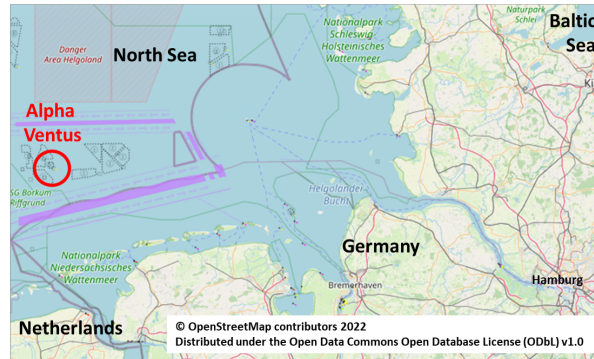
Figure 2. Location of Alpha Ventus and the metmast FINO1 (adapted from OpenStreetMap – www.openstreetmap.org/copyright).

Table 1. Properties of the investigated AV-07 turbine (Bartsch, 2020).

Type	Adwen 5-116 turbine
Substructure	Tripod
Rotor diameter	116 m
Hub height	90 m
Water depth	approx. 30 m
Rated power	5 MW
Rotor speed	$5.9-14.8$ rpm
Rated wind speed	12.5 ms^{-1}
Cut-in wind speed	3.5 ms^{-1}
Cut-out wind speed	25 ms^{-1}

Although measurement data are, in general, available for time periods since 2011, for many periods, the data quality is not sufficient for fatigue damage extrapolations. Many sensors have experienced defects, leading to missing or erroneous data. For fatigue extrapolations, it is important that data are recorded with a high availability for a continuous period of at least one year to cover seasonal effects properly. Since data of higher quality are available for the sensors on the tower compared to sensors
115 on the substructure, this work only considers data from the previously mentioned strain gauges on the tower. Moreover, only the data from three specific years have a sufficient quality to be taken into account: 1st January 2011 to 31st December 2011 and 1st October 2015 to 30th September 2017. For these three years, raw data are post-processed as described in the next section before calculating fatigue damage.

2.2 Raw data and data processing

120 For this work, three types of data are required: strain data, data regarding environmental conditions and data concerning operational conditions.
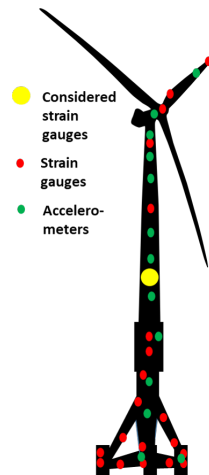
Figure 3. Illustration of the AV-07 turbine (not to scale) and some of the installed sensors according to the sensor documentation (IWES, 2019).
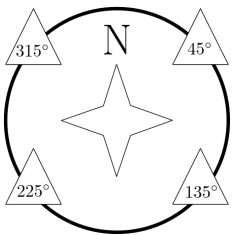
Figure 4. Positions of the strain gauges around the circumference of the tower.
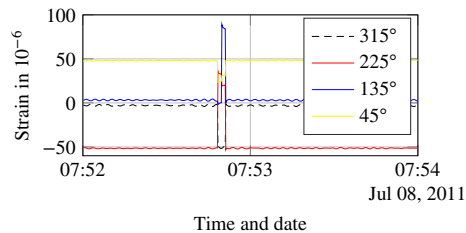
Figure 5. Example strain data before post-processing, which include clearly erroneous measurements, i.e., physically unrealistic peak without any oscillation afterwards.
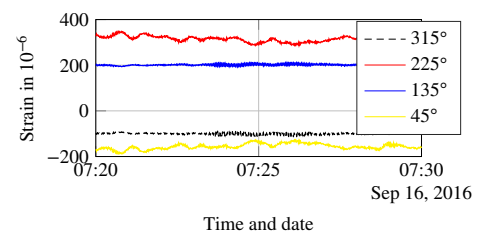
Figure 6. Example strain data. A wind direction of approximately 235° leads to wind-induced oscillations for the sensors at 225° (tension) and 45° (compression) and oscillations in its eigenfrequencies in side-to-side direction (sensors at 135° and 315°).

Strains are measured on the tower of the AV-07 turbine (see Fig. 3). Four temperature-compensated strain gauges are positioned around the circumference of the tower (see Fig. 4). The raw data was post-processed using semi-automatic methods to exclude, for example, erroneous data as shown in Fig. 5. Profound post-processing and full sensor calibrations were not performed. The reasons for this are: first, in an industry context, time-consuming manual post-processing is prohibitive. Second, due to the long measurement period of approximately 7 years, the identification of all sensor drifts is not straightforward. And, lastly and most importantly, the mean value of the signals is not relevant for the fatigue damage calculation (cf. Section 3.1). Some example data after the post-processing can be seen in Fig. 6.

Operational conditions are taken from SCADA data from the AV-07 turbine. Environmental conditions are, in most cases,

130   taken from the FINO1 met mast. Only if no data are available from FINO1, the wind conditions included in the SCADA data
from the AV-07 turbine are taken into account. The reason for this approach is that the environmental condition data measured
at FINO1 are of higher quality. There are no or at least less disturbance effects, e.g., no increased turbulence caused by the
rotor. For this work, six environmental conditions, namely wind speed, wind direction, turbulence intensity, significant wave
height, wave peak period and wave direction, are considered. In addition, the turbine status – recorded by the SCADA system,

135   e.g., normal operation, start-up, emergency stop, etc. – is taken into account. For all EOCs, only statistical values, e.g., mean
values of ten-minute intervals are available. At FINO1, wind conditions are measured at 90 m above mean sea level. The wind
speed is measured using cup anemometers. These are positioned on jibs in secondary wind directions to reduce shadow effects.
A buoy in the immediate vicinity of FINO1 (about 150 m away) measures the wave conditions. The available EOC data are
post-processed using semi-automatic methods to remove outliers. As stated before, missing or erroneous data are replaced

140   by SCADA measurements to increase the number of ten-minute intervals for which strain and EOC data are available. Some
example data after the post-processing can be seen in Fig. 7. Since EOC data are mainly used for the binning (see Section 3.2),
the slightly biased wind data, due to rotor disturbance when using SCADA data, are less relevant compared to the increase in
the overall amount of data. Or in other words: in this context, the amount of EOC data is more relevant than the EOC data
quality.
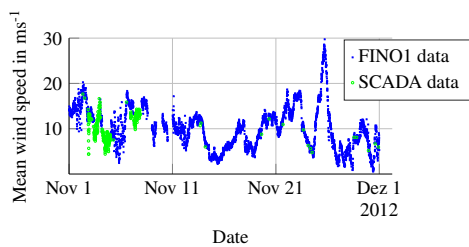
Figure 7. Example wind speed data after the post-processing.
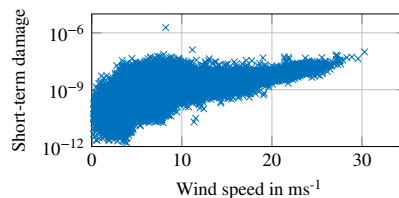
Figure 8. Correlation of wind speeds and short-term damage values determined using the strain gauge at $315°$. All ten-minute intervals in 2016.
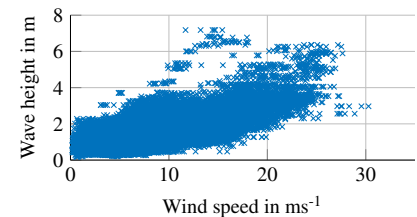
Figure 9. Correlation of wind speeds and wave heights. Mean values of all ten-minute intervals in 2016.

145

As stated in the introduction, most extrapolation approaches are based somehow on correlations between fatigue damage and
EOCs. This correlation is shown as an example in Fig. 8 for the wind speed and a short-term damage to the tower. There is
a pronounced correlation for all wind speeds. For high wind speeds, it is more visible, since the amount of data is lower for
high wind speeds. As EOCs are also correlated with each other – as shown in Fig. 9 – it is not straightforward to determine the

150   relevant EOCs that need to be considered for the extrapolation in time. Before discussing different approaches for extrapolation
in time in Section 3.2, in the next section, some more information on the short-term fatigue damage calculation are given.
In the following, a ten-minute interval is only considered if strain and complete EOC data are available.

3 Fatigue damage calculation

Assuming a linear damage accumulation according to the Palmgren-Miner rule, the overall damage, e.g., the lifetime damage

155 of a structure, can be calculated as the sum of many short-term damage values. For all steel components of a turbine, the assumption of linear damage accumulation is justified. Hence, for each ten-minute interval, the (short-term) fatigue damage is calculated separately. The calculation procedure for the short-term damage based on strain measurements is fairly standardised and is briefly presented in the following section. It should be noted that this work focuses on the damage at a single location (cf. Fig. 3). Most results are given for the strain gauge at 315°. This is reasonable for the current purpose, i.e., to assess

160 and validate methods for extrapolations in time. If fatigue damage values at other locations are required, for example, for an industry application, either data from additional strain gauges must be used or spatial extrapolations (Maes et al., 2016; Henkel et al., 2020) are needed.

3.1 Short-term damage

Assuming linear damage accumulation and a fixed location at which high resolution strain data ($\epsilon(t)$) are available, the fatigue

165 damage sustained in a given time period can be calculated as follows. First, stress time series ($\sigma(t)$) are determined by applying Hooke's law

$$\sigma(t) = E \, \epsilon(t), \tag{1}$$

where E is the Young's modulus. Since it is assumed that strain data are already available for the required location, e.g., height and position around the circumference of the tower, a rainflow cycle counting of stress ranges ($\Delta\sigma_i$) can be applied directly to

170 the stress time series. Here, $\Delta\sigma_i$ is the stress range of the ith band (also called block or bin) in the factored stress spectrum (cf. Annex A of Eurocode 3 (2010)). The number of required stress bands (n_σ) is chosen to be 500 bands – logarithmically spaced between 10 kPa and 1 GPa (Hübler et al., 2019). Two example cycle counts for a ten-minute interval are shown in Fig. 10 and 11. Figure 10 shows the cycle count for a strain gauge in fore-aft direction and Fig. 11 for one in side-to-side direction. The latter clearly features many cycles corresponding to the eigenfrequency of the structure (peak just above 10^6 Pa).
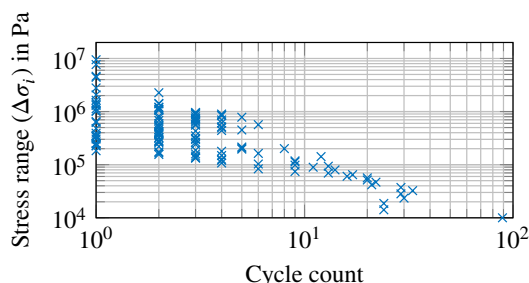
Figure 10. Results of the rainflow cycle count for a ten-minute interval (18 March 2016; 00:10:00). Strain gauge at 315°, which corresponds to fore-aft direction for this interval.
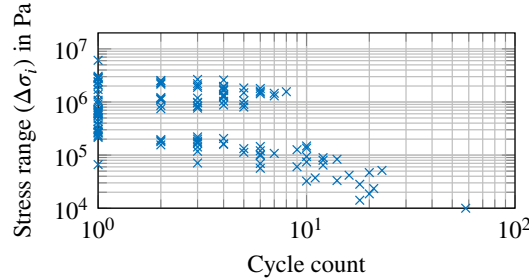
Figure 11. Results of the rainflow cycle count for a ten-minute interval (18 March 2016; 00:10:00). Strain gauge at $225°$, which corresponds to side-to-side direction for this interval.

For nominal stresses at the position of interest (here: the measurement position), an overall safety factor (SF) is applied. It consists of several sub-factors. Using the safety factor, a representative value for the concentrated stresses at the structural detail is achieved. First, a stress concentration factor for the specific detail is used (here: $SCF = 1.0$ according to a recommended practice of DNVGL-ST-0262 (2016)). Second, a correction for large wall thicknesses – the so-called size effect (SE) correction – is applied (DNVGL-ST-0262, 2016). Third, a material safety factor (here: $MSF = 1.0$ due to regular inspections (DNV GL, 2016)) is used. All these factors might be chosen differently and/or can be regarded as uncertain. However, to assess the extrapolation in time, precise values for the safety factor are not required. The corrected stress ranges can be calculated as follows:

$$\Delta\sigma_{cor,i} = \Delta\sigma_i \times SF = \Delta\sigma_i \times SCF \times SE \times MSF. \tag{2}$$

The last step in calculating the damage sustained in a given interval is the application of the Palmgren-Miner rule, i.e., linear damage accumulation, and the application of S-N curves according to the standards (DNVGL-ST-0262, 2016). Here, DNV S-N curve D in air is applied. The fatigue damage for a given time series (D_j) can be calculated as follows:

$$D_j = \sum_{i=1}^{n_\sigma} \frac{n_{ij}}{N_i}, \tag{3}$$

where i and j are indices for the stress band and the time series, respectively, and n_{ij} is the number of cycles associated with the stress range $\Delta\sigma_{cor,ij}$. The endurance (N_i; maximum number of cycles) for the same stress range is obtained from the corresponding S-N curve.

3.2 Damage extrapolation in time

If strain data were available for the entire lifetime of the wind turbine, it would be possible to determine its fatigue lifetime by using the previously described approach of calculating short-term damage values. In this case, the lifetime damage (D_{total}) would be:

$$D_{\text{total}} = \sum_{j=1}^{N_{\text{LT}}} D_j, \tag{4}$$

where N_LT is the number of (short-term) intervals in the entire lifetime, e.g., $N_\text{LT} = 6 \times 24 \times 365.25 \times 20$ for ten-minute intervals and a lifetime of 20 years.

However, normally, strain data are not available for the entire lifetime. Therefore, some kind of extrapolation procedure in time is necessary. In the following, three different approaches are presented: a simple linear extrapolation, an extrapolation based on bins of EOCs and an extrapolation based on machine-learning techniques. The latter two make use of the correlation of short-term damage and EOCs (cf. Fig. 8).

3.2.1 Simple extrapolation

The simplest extrapolation approach is a linear extrapolation. It assumes that the fatigue damage only depends on the elapsed time (Louraux and Brühwiler, 2016). This means that the fatigue damage sustained in any predicted period can be calculated as follows:

$$D_\text{pred} \approx \frac{N_\text{n}}{N_\text{m}} \sum_{j=1}^{N_\text{m}} D_j, \qquad (5)$$

where N_m and N_n are the number of (short-term) intervals in the measurement and the predicted period respectively. If the predicted period is the entire lifetime, it follows that:

$$D_\text{total} \approx \frac{N_\text{LT}}{N_\text{m}} \sum_{j=1}^{N_\text{m}} D_j. \qquad (6)$$

For very long measurement periods ($N_\text{m} \approx N_\text{LT}$), this approach yields accurate results. However, if the measurement period is less than one year, seasonal effects are neglected. For example, storms during the winter lead to increased damage. Even if the measurement period covers more than a year, yearly and/or long-term effects are not taken into account, e.g., varying damage due to climate change (Hübler and Rolfes, 2021).

3.2.2 Extrapolation based on bins of EOCs

A more advanced approach, which makes use of the correlation between fatigue damage and EOCs (cf. Fig. 8), is a so-called binning approach (Marsh, 2016; Hübler et al., 2018). This binning approach is still very simple to apply, and therefore, quite user friendly. The binning approach is based on the idea that most variations in fatigue damage are due to changing environmental conditions. Hence, it is not necessary to know fatigue damage for the entire lifetime. Having determined the correlation between EOCs and fatigue damage, it is sufficient to know the EOCs for the entire lifetime. Since many EOCs are part of the SCADA data, EOCs are frequently known for the entire lifetime. Hence, the only challenge is determining of the correlation between fatigue damage and EOCs. This correlation can be determined by statistical and/or machine-learning techniques (Dimitrov and Natarajan, 2019; Santos et al., 2021) that yield a functional relationship between (short-term) damage and EOCs:

$$D_j = f(\mathbf{x_j}) + \epsilon, \qquad (7)$$

225 where f is the functional relationship, $\mathbf{x_j}$ is the vector of all EOCs considered in the analysed interval j and ϵ is an error term, which cannot be explained by changes in the EOCs considered. Such functional relationships are discussed in the next section. The binning approach presumes that due to the uncertainty in the measured EOCs (e.g., disturbance of the wind conditions by the rotor) and some unexplained variations in the fatigue damage (ϵ), a precise definition of a function is not expedient. Therefore, the (short-term) damage values are clustered according to the EOCs. For each cluster or bin, the mean

230 damage is determined using the available measured strain data. The calculation of mean values is supposed to average out most unexplained effects (ϵ). Subsequently, the damage sustained in the predicted period is:

$$D_{\text{pred}} \approx N_{\text{n}} \sum_{i_1=1}^{M_1} \cdots \sum_{i_d=1}^{M_d} \left(Pr_{\text{pred},i_1,\ldots,i_d} \bar{D}_{i_1,\ldots,i_d} \right), \tag{8}$$

where d is the binning dimension – i.e., the number of EOCs considered, M_1 to M_d are the number of bins for the corresponding EOC and $Pr_{\text{pred},i_1,\ldots,i_d}$ and $\bar{D}_{i_1,\ldots,i_d}$ are the occurrence probability of and the mean damage in bin $i_1,\ldots,i_d$ respectively.

235 Mean damage within the bins can be determined using a limited amount of strain data, e.g., one year (measurement period). To determine the bin probabilities, only data concerning the EOCs are required. Hence, bin probabilities are determined using data of the predicted period. If the predicted period is the entire lifetime, it follows that:

$$D_{\text{total}} \approx N_{\text{LT}} \sum_{i_1=1}^{M_1} \cdots \sum_{i_d=1}^{M_d} \left(Pr_{\text{LT},i_1,\ldots,i_d} \bar{D}_{i_1,\ldots,i_d} \right). \tag{9}$$

The main difference is that bin probabilities ($Pr_{\text{LT},i_1,\ldots,i_d}$) are now determined using EOC data, e.g., SCADA data, covering

240 the entire lifetime. In contrast to the previously presented simple extrapolation, seasonal effects and long-term changes due to changing EOCs are taken into account by the bin probabilities. The main challenge of the binning approach is to apply expedient binning dimensions and bin numbers. Too few bins – i.e., a low number of bins per dimension and/or a low dimensionality – result in highly scattered damage values within each bin. This yields a less accurate approximation of the correlation. In the extreme case of $d = 0$, the binning approach is identical to the simple extrapolation. Too many bins lead to only a few

245 short-term damage values within each bin or even empty bins. These empty bins have to be filled up. This may lead to biased results. In this work, empty bins are filled up relatively conservatively. The largest mean value of the surrounding bins is used. Hence, the mean damage in bin $i_1,\ldots,i_d$ can be calculated as follows:

$$\bar{D}_{i_1,\ldots,i_d} = \begin{cases} \dfrac{1}{N_{i_1,\ldots,i_d}} \displaystyle\sum_{j=1}^{N_{i_1,\ldots,i_d}} D_j & \text{if } N_{i_1,\ldots,i_d} > 0 \\ \max(\bar{D}_{i_1+j_1,\ldots,i_d+j_d}) & \text{otherwise,} \end{cases} \tag{10}$$

where $N_{i_1,\ldots,i_d}$ is the number of short-term damage values in bin $i_1,\ldots,i_d$ and $j_k = \{-1,0,1\}$ if $1 < i_k < M_k$, $j_k = \{0,1\}$ if

250 $i_k = 1$ and $j_k = \{-1,0\}$ if $i_k = M_k$ for $k = 1,\ldots,d$.

3.2.3 Extrapolation based on a functional relationship

As already discussed in the previous section, the correlation between short-term damage and EOCs can also be expressed as a functional relationship, i.e., $D_j = f(\mathbf{x_j}) + \epsilon$. In general, such a functional relationship can be approximated using various

statistical and/or machine-learning techniques, e.g., multiple regression, Gaussian process regression (GPR), artificial neural
255 networks (ANN), etc. To determine the functional relationship, training data are required to train the relation between inputs,
i.e., EOCs, and outputs, i.e., fatigue damage values. Similar to the binning approach, it is not necessary that strain data are
available for the predicted period or the entire lifetime. The strain and EOC data from the measurement period, e.g., one year,
are used as training data. Subsequently, fatigue damage for other time periods can be predicted using EOC data only. EOC data
are normally available for the entire lifetime.

$$260 \quad D_{\text{pred}} \approx \sum_{k=1}^{N_n} f(\mathbf{x_k}) \quad \text{and} \tag{11}$$

$$D_{\text{total}} \approx \sum_{k=1}^{N_{\text{LT}}} f(\mathbf{x_k}). \tag{12}$$

The accuracy of the prediction also depends on the EOCs considered. If too few EOCs are taken into account, important
features might be missed. Too many EOCs might lead to some kind of overfitting.

In this work, GPR and ANN are investigated. Both methods are very powerful machine-learning techniques, which have already
265 been applied successfully in wind engineering (Dimitrov et al., 2018; Santos et al., 2021). On the downside, they are less user
friendly compared to the binning approach. At least some expert knowledge is required to achieve accurate predictions.

All configurations for ANN and GPR used in this work are based on recommendations in literature, e.g., Larose and Larose
(2014), previous work, e.g., Müller et al. (2021), and preliminary studies.

For ANN, inputs and outputs are normalised to values between 0 and 1. Two hidden layers with 10 neurons each are used.
270 Hyperbolic tangent sigmoid transfer functions are applied in all layers to prevent unrealistic negative outputs, i.e., negative
damage values. The network is trained using the Levenberg-Marquardt algorithm. 80 % of the input data are used as training
data and 20 % as validation data. Since the performance of ANN depends strongly on the randomly chosen initial weights for
this problem, an automated control algorithm is implemented. It restarts the learning process using new initial weights if the
relative prediction error is higher than 50% for the training data.

275 For GPR, inputs are normalised to values between 0 and 1. Outputs are standardised to achieve a mean value of 0 and a
standard deviation of 1. A purely quadratic basis function and a Matern kernel with parameter 5/2 and a separate length scale
per input are chosen. Since the amount of training data is extensive, e.g., more than $50\,000$ training data points from one year
of measurements, a random subset selection for the training is conducted. This reduces the time required for training the model
to a practicable level.

280 ### 3.2.4 Uncertainty assessment

Independent of the chosen extrapolation approach, the results will be uncertain, for example, due to unrepresented EOCs.
Nonetheless, the main reason for this uncertainty is the limited amount of strain data, and therefore, of short-term damage
values (D_j). To approximate the uncertainty of the predicted (D_{pred}) or overall fatigue damage (D_{total}), bootstrapping (Efron,
1979) can be applied (Marsh, 2016). Bootstrapping allows the estimation of a distribution of D_{pred} or D_{total} by applying random

285   sampling with replacement. This distribution can be used to judge the uncertainty present. For example, the standard deviation or percentile values of this distribution are relevant measures of uncertainty.

In the present case, short-term damage values (D_j) are limited – e.g., only available from one year of strain measurements. Hence, using another year of measurement data would yield a (slightly) different value for D_{pred} or D_{total}. However, other values for D_j are not available. Therefore, D_{pred} or D_{total} is calculated several times using different samples D_j, which are sampled

290   randomly with replacement from all available samples $\{D_1, \ldots, D_{N_m}\}$. As an illustration: for $N_m = 3$, the standard calculation would be based on the samples $\{D_1, D_2, D_3\}$. Other random realisations with replacement are, for example, $\{D_1, D_1, D_1\}$ or $\{D_2, D_2, D_3\}$.

4   Results

In this section, the three methods for the temporal extrapolation of fatigue damage, which were presented in the previous

295   section, are assessed using measurement data from the Alpha Ventus wind farm (see Section 2). The assessment tries to answer the subsequent questions:

1. How should the parameters of each of the methods, e.g., bin sizes, be chosen to yield the most accurate results?

2. Which method can predict fatigue damage for other time periods most accurately?

3. How high is the uncertainty in the prediction?

300   4. How much training data is required? How long is the minimum measurement period?

5. Do the approaches still yield reasonable results if long-term extrapolations over several years are conducted?

6. Do the approaches still yield reasonable results if extrapolations into the future are conducted, for which no EOC data are available?

Since high-quality strain data are only available for three years, for most steps, an extrapolation of measurement data from

305   one year to another year is conducted. For example, data from October 2015 to September 2016 is extrapolated to October 2016 to September 2017. Since strain and EOC data are available for both periods, the accuracy of the extrapolation can be determined by comparing the predicted damage (D_{pred}) with the damage calculated using the actual measured strain data from the predicted period, i.e., the real damage (D_{real}). Hence, the methods can be validated.

4.1   Parameter selection

310   For the three extrapolation methods considered in this work, different parameters must be chosen, for example, the number of EOCs to be taken into account. To determine the most suitable parameters, the extrapolation from one year to another year is analysed. Since the choice of the parameters might be influenced by the period investigated, several periods should be analysed. However, only three years of data are available and only two of these years are consecutive years. Hence, within

the two consecutive years, the one-year period is shifted, e.g., October 2015 to September 2016 is extrapolated to October

315 2016 to September 2017, November 2015 to October 2016 is extrapolated to October 2015 and November 2016 to September

2017, etc. A visualisation of this shifting procedure is shown in Fig. 12. Using these shifted periods, 13 "different" periods are

available. Therefore, some statistical significance for the determined parameters is given.
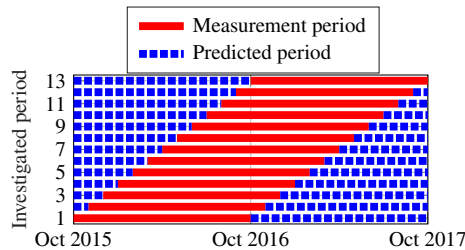
Figure 12. Visualisation of the 13 "different" periods for statistical evaluations.

4.1.1 Parameter selection: simple extrapolation

For the simple extrapolation, no parameters have to be chosen. First results for the simple extrapolation are presented in Fig.

320 13. For all 13 one-year periods, the unsigned percentage errors (PE) of the predicted damage are shown:

$$PE = \frac{|D_{\text{real}} - D_{\text{pred}}|}{D_{\text{real}}}. \tag{13}$$

Moreover, a box plot shows some summary statistics: the median (red centre line), the 25th and 75th percentile (box), the

minimum and maximum values (excluding any outliers) and possible outliers of 13 "different" measurement periods. Hence,

the box plots visualise the variation in the accuracy of the predictions depending on the period considered. All box plots in the

following sections show the same summary statistics.
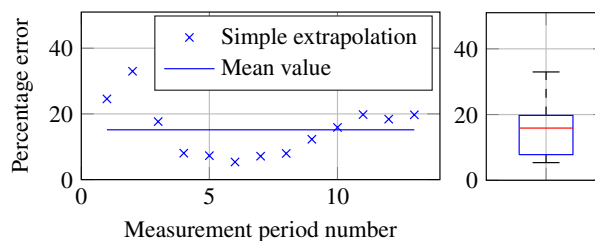
Figure 13. Percentage errors of predicted damage values using a simple extrapolation method compared to real damage values. Prediction

from one year to a second year for 13 "different" years. Box plot shows summary statistics.

325

Clearly, the prediction does not yield precise results. Nonetheless, it is remarkable that even such a simple extrapolation leads

to results with errors of less than 35 %.

4.1.2 Parameter selection: EOC bins

For the extrapolation based on bins of EOCs, the number and type of EOCs to be taken into account and the bin size must

330   be selected. In contrast to previous work by Louraux and Brühwiler (2016) and Hübler et al. (2018), who focused on one to three different wind parameters, in this work, six different environmental conditions (wind speed and direction, turbulence intensity and wave height, period and direction) are analysed in a systematic manner. Bin sizes are chosen in such a way that the overall range of each environmental condition is discretised into about 3 to 60 bins depending on the environmental condition. For example, bins of 0.5 to $6\,\mathrm{m\,s^{-1}}$ are used for the wind speed. Regarding operational conditions, only the turbine status is

335   considered. As bin sizes for the turbine status (a discontinuous variable) have to be defined differently, operational conditions are considered separately in a second step (see Section 4.1.4).

Some example results for the binning approach are presented in Fig. 14. On the one hand, the results clarify that many bin types yield similar results. The detailed performance depends on the time period considered, i.e., the year in this case, as demonstrated by the scatter illustrated by the box plot. On average, it does not make a significant difference whether, for example, wind speed

340   bins or wind speed and turbulence intensity bins are used. Most important is the consideration of the wind speed. This result is in accordance with previous research (Hübler et al., 2018). On the other hand, the results also demonstrate that there are bin types that perform worse. For example, wind direction and turbulence intensity bins without considering the wind speed or six-dimensional bins do not perform very well. If the wind speed is neglected, important effects are missed. If the binning dimension is too high, a large number of bins remain empty. Hence, the empty bins have to be filled up artificially, leading

345   to less accurate extrapolations. Moreover, a high dimensionality leads to increased computing times, which is problematic at least for the uncertainty assessment. The bin size does not have a pronounced effect. Medium sized bins (e.g., 10 bins per environmental condition) are recommended. Again, small bin sizes might result in many empty bins. Excessively large bin sizes lead to higher uncertainties within each bin.
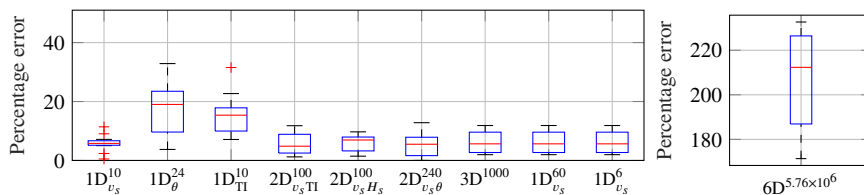
Figure 14. Percentage errors of predicted damage values using a binning method compared to real damage values. Comparison of various bin types and sizes (n is the overall number of bins): wind speed (v_s), wind direction (θ) and turbulence intensity (TI) only ($1\mathrm{D}_{x,n}$); combinations of two environmental conditions out of wave height (H_s), v_s, θ and TI ($2\mathrm{D}_{xy,n}$); v_s, TI and H_s (3D); all six environmental conditions (6D). Please note: for the sake of clarity, the vertical axis is scaled differently for 6D.
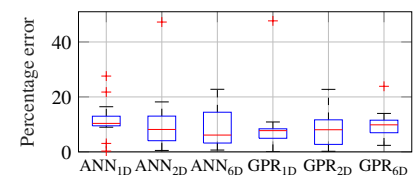
Figure 15. Percentage errors of predicted damage values using a functional relationship compared to real damage values. Comparison of GPR and ANN and different EOCs: v_s (1D); v_s and TI (2D); all environmental conditions (6D).

350 The optimal choice always depends on the turbine considered, the measurement period and the extrapolation period. Hence, an automated selection method would be beneficial. Ideally, different choices would be assessed automatically for the predicted period and the best choice selected. However, strain data from the predicted period are normally not available; otherwise no extrapolation would be required. This is why automated selection must be based on cross-validation, i.e., splitting up the measurement period. One part of the data is used as training data to determine the mean damage in all bins. Another part

355 replaces the prediction period. It is used to evaluate the accuracy of the extrapolation for the chosen settings. This procedure reduces the amount of training data significantly. As a result, predictions become less accurate. In most cases, due to the limited training data, the automated selection yields bin sizes which are too fine, i.e., overfitting. Hence, although automated selection is desirable, it is not suitable for "short" measurement periods. For example, measurement periods of one year or less – as used in this work – are not sufficient.

360 Therefore, it can be summarised that the choice of the bin dimension and size is of minor importance as long as empty bins do not occur at all or only in some rare cases. For most applications, simple wind speed bins with a size of 2 to 3 ms^{-1} are adequate, although they are not optimal. Moreover, fortunately, it is not necessary to take wave conditions into account; these are rarely available, as they are not recorded by the SCADA system.

4.1.3 Parameter selection: Functional relationship

365 For the extrapolation based on a functional relationship, only the number and type of EOCs to be taken into account are relevant. The same six environmental conditions as before are considered. Operational conditions are taken into account separately in a second step. Some example results for the functional relationship are presented in Fig. 15. Again, the detailed performance depends significantly on the measurement period. Moreover, a significant amount of uncertainty is introduced by the random selection of the initial weight using ANN and by the random subset selection using GPR. This is discussed in more detail in

370 Section 4.2.

A slight improvement in the accuracy might be achieved for ANN if additional environmental conditions are taken into account. However, this improvement is not significant, especially when considering the previously mentioned uncertainty due to the random selection of the initial weights. At least for wave conditions, it definitely does not justify the effort needed to measure them.

375 Therefore, in the following, only results using a single environmental condition, i.e., the wind speed, are shown.

4.1.4 Operational conditions

In contrast to the environmental conditions, the turbine status, i.e., the sole considered operational condition, is not defined continuously. Hence, it has to be treated differently from the environmental conditions. Therefore, for all three methods, in a first step, the entire data is split up according to the turbine status. For the simple extrapolation this yields:

380
$$D_{\text{pred}} \approx N_{\text{n}} \sum_{k=1}^{K} \left(\frac{Pr_k}{N_{\text{m,k}}} \sum_{j=1}^{N_{\text{m,k}}} D_j \right),$$
(14)

where K is the number of different turbine statuses considered, $N_{m,k}$ is the number of available short-term damage values for this turbine status and Pr_k is the occurrence probability of this turbine status, determined using data from the predicted period (e.g., the entire lifetime). For the extrapolation based on a functional relationship, the procedure is equivalent. For the binning approach, it means that just another binning dimension is added.

385 Similar to the challenge of determining a suitable bin size, an adequate number of different turbine statuses and the type of statuses must be found. The most simple differentiation is "normal production operation" and "others" (two statuses). This was already proposed by Hübler et al. (2018). Other possible classifications are, for example, "normal production operation", "idling (below cut-in)", "idling (above cut-off)" and "others" (four statuses) or another additional class for "service" (five statuses). Other combinations of these classes are possible as well, but are not investigated in detail in this work. Figures 16

390 to 18 show how the performance of the three methods changes if operational conditions are taken into account. For all three methods, the effects are not very pronounced and relatively uncertain, i.e., depend on the period considered. Nonetheless, percentage errors can be reduced by about 20 to 30 % for the binning approach and ANN, if several different turbines statuses are considered separately. For the simple approach, splitting up the data does not improve the approximation significantly.
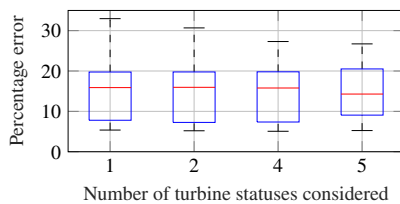
Figure 16. Percentage errors of predicted damage values using simple extrapolation. Data separated according to turbine statuses: no separation (1); "normal production operation" and "others" (2); plus "idling below cut-in" and "above cut-off" (4); plus "service" (5).
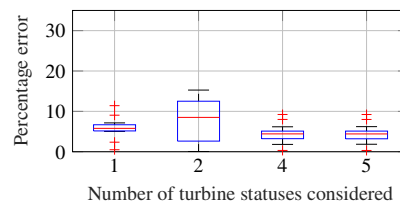
Figure 17. Percentage errors of predicted damage values using a binning method (one-dimensional wind speed bins) compared to real damage values. Data separated according to different turbine statuses (cf. Fig. 16).
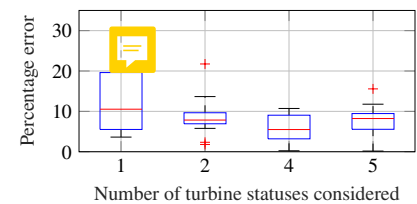
Figure 18. Percentage errors of predicted damage values using an ANN compared to real damage values. Data separated according to different turbine statuses (cf. Fig. 16).

395 To summarise, splitting up the data according to operational conditions can help to improve the extrapolation. However, it is not straightforward to determine the best separation, as it depends on the period considered and probably on the turbine as well. Moreover, if the measurement period is relatively short and/or many environmental conditions are taken into account, e.g., three-dimensional binning, clustering according to operational conditions becomes more challenging. In this case, the amount of data for each turbine status might become insufficient.

400 Hence, as the improvements are not pronounced, for many applications it is sufficient not to cluster the data according to operational conditions. This is why clustering according to operational conditions is not performed in the rest of this work.

4.2 Performance

In the following, the performance of the extrapolation approaches with respect to the accuracy of the prediction, the computing

time, and the required data and knowledge is evaluated. For all methods, no clustering according to operational conditions

405 is applied. For the binning approach, only wind speed bins with a bin size of 3 ms^{-1} are used. Similarly, only wind speed

correlations are taken into account for ANN and GPR. These choices are in accordance with the findings of Section 4.1.

In Fig. 19, the percent errors of all approaches for all 13 years are shown.
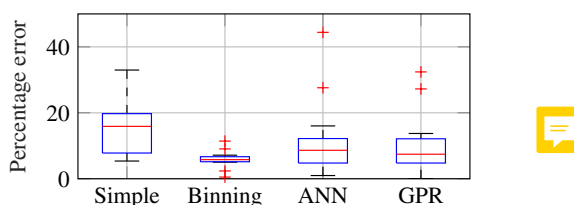
Figure 19. Percentage errors of predicted damage values using all extrapolation methods compared to real damage values. Predictions from
one year to a second year for 13 "different" years.

It becomes apparent that the binning approach reduces the percentage error by about 60 % compared to the simple extrap-

olation. Moreover, the binning approach outperforms ANN and GPR. However, two facts about ANN and GPR should be

410 mentioned. First, the initial weights used by ANN and the subsets used by GPR are chosen randomly. Hence, the performance

of both is not deterministic. This fact becomes clear when comparing the results of ANN and GPR (wind speed only) in Fig.

15, 18 and 19. Therefore, in addition to the previous results for a single run in Fig. 15, 18 and 19, some statistical evaluation

needs to be done. Running ANN and GPR 100 times leads to mean percentage errors of all runs and years of 10.3 % and

8.9 %, for each method respectively. Hence, on average, both are outperformed by the binning approach, which yields a mean

415 percentage error of all years of 5.9 %. Second, it might be possible to improve the accuracy of the machine-learning approaches

by exploiting their full potential, e.g., by using more hidden layers for ANN, etc. However, a comprehensive analysis of the

machine-learning approaches is beyond the scope of this work, as a user-friendly extrapolation approach is being sought. More-

over, Santos et al. (2021), who analysed ANN in more detail in the context of fleet-wide extrapolations, also found out that

predictions – using ten-minute SCADA data only – lead to percentage errors of up to 10 % in damage-equivalent loads. They

420 showed that ANN is suitable for highly accurate predictions, if more or better measurement data (e.g., one-second SCADA

data) are available, but not for user-friendly predictions based on ten-minute SCADA data, which are the focus of this work.

The second performance criterion evaluated in this work is the computing time. For deterministic predictions using wind speed

as the only EOC, all methods are equally suitable. For a prediction from one year to another, the simple approach and the

binning approach require less than 0.1 seconds on a standard desktop computer. ANN requires a few seconds and GPR about

425 30 seconds. For uncertainty assessments, requiring hundreds or thousands of predictions, the computing time becomes rele-

vant. For example, if 1000 predictions are used for the uncertainty assessment, ANN and GPR require more than 30 minutes

and more than 8 hours respectively, whereas the prediction based on bins is finished after about 10 seconds and the simple

prediction within 1 to 2 seconds. Overall, the computing time of GPR is about 15 times higher compared to ANN. For ANN, it is more than 100 times higher compared to the binning approach and another 5 to 10 times higher compared to the simple

430 approach. If additional EOCs are taken into account, the computing time of the binning approach increases quite quickly. For example, for three EOCs, the computing time of the binning approach increases by a factor of 5 compared to the case with a single EOC. For six EOCs, a statistical evaluation already becomes unfeasible on a standard desktop computer for the binning approach. The reason for the increase is the extensive empty bin filling. For GPR, the increase is moderate. The computing time rises by factors of 1.5 to 2 and 4 to 5 for three and six EOCs, respectively. This increase is caused by the more complex

435 training procedure needed for more inputs. For ANN, the computing time does not change significantly.

The last criterion, i.e., the user friendliness and required data, is a more vague criterion. Clearly, the simple extrapolation does not require any additional data (e.g., SCADA data) and is straightforward to apply. The binning approach, especially if only wind speed bins are used, is also quite user friendly and does not rely on detailed data. For the machine-learning approaches, first of all, much more expert knowledge is required to achieve adequate results. Moreover, the two previous criteria demon-

440 strated that machine-learning approaches perform better with respect to accuracy and computing time if additional data (e.g., additional EOCs or one-second SCADA data) are available.

To summarise, the simple extrapolation works relatively well. However, if ten-minute SCADA data are available, the binning approach clearly outperforms the simple extrapolation with respect to accuracy, while computing time and user friendliness are comparable. For expert users and high quality data, ANN and GPR might be alternatives. For the current application, they

445 are less accurate. Moreover, the machine-learning approaches, especially GPR, have significantly higher computing times. The high computing time of GPR makes statistical evaluations nearly unfeasible on a standard desktop computer. This is why GPR is not considered in the rest of this work.

4.3 Uncertainty assessment

The box plots in the previous sections showed that the performance of the various extrapolation approaches depends on the

450 period considered, as there is a significant scatter of the percentage error across the 13 "different" years. In addition, for a single considered period, the uncertainty of the prediction due to the limited available strain data can be estimated. As stated in Section 3.2.4, this uncertainty can be approximated by applying bootstrapping, i.e., resampling with replacement. Therefore, for all three extrapolation approaches, bootstrapping is conducted using $N_{\mathrm{BT}} = 1000$ runs of resampling. For all approaches, the same settings as before are used, e.g., only wind speed bins. The bootstrapping yields N_{BT} values for the extrapolated

455 damage D_{pred}. These values are compared to the real damage values for the second year to calculate an empirical distribution for the percentage error. In this case, signed percentage errors are used – i.e., no absolute values of errors – to analyse the bias of the extrapolation as well. Example resulting empirical distributions are shown in Fig. 20 for an extrapolation from one year (1st October 2015 to 30th September 2016) to another year (1st October 2016 to 30th September 2017). The results demonstrate that there is some uncertainty in the extrapolation, which should not be neglected. Moreover, all approaches lead to slightly

460 biased results. This is not surprising, since the correlation between the environmental conditions and fatigue damage cannot cover all effects. Nonetheless, this bias is not critical, as it is not systematic. For other measurement periods (cf. Fig. 21), biases

can be different.

Overall, even the highest errors are below $\pm\,20\,\%$ for the binning approach. Finally, it should be noted that the variance of the distributions for the binning approach (cf. Fig. 20 and 21) is smaller compared to the other approaches. Hence, the binning

465   approach features the lowest uncertainty. The reason for the higher uncertainty for ANN is the random choice of the initial weights. Compared to the simple extrapolation, the binning approach has a slightly reduced uncertainty, as the scattering of the short-term damage values D_j within each bin is smaller compared to the scattering if no bins are employed.
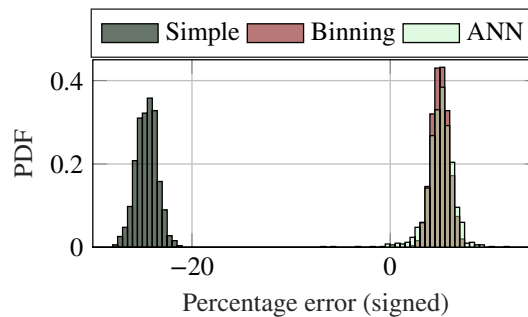
Figure 20. Empirical distribution for the (signed) percentage error of predicted damage using different extrapolation methods. Prediction from one year (1st October 2015 to 30th September 2016) to another year (1st October 2016 to 30th September 2017).
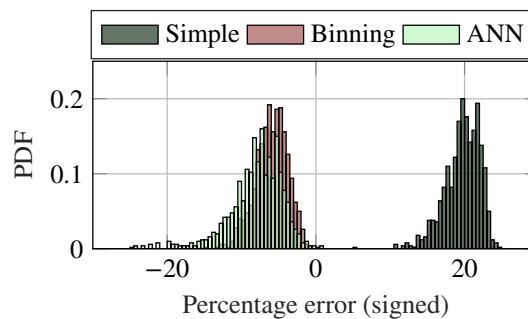
Figure 21. Empirical distribution for the (signed) percentage error of predicted damage using different extrapolation methods. Prediction from one year (1st August 2016 to 31st July 2017) to another year (1st October 2015 to 31st July 2016 and 1st August 2017 to 30th September 2017).

4.4   Minimum measurement length

470   In theory, the three extrapolation methods can be used to extrapolate from any period to another. However, if the measurement length is too short, the extrapolation will be biased (Hübler et al., 2018). A simple example in this context is that an extrapolation based on data from a few summer months with benign environmental conditions will lead to an underestimation of the fatigue

damage for winter months. Therefore, in the following, the convergence of the percentage error of the predicted damage with increasing measurement length is analysed to determine a minimum measurement length. This analysis is conducted for all
475 three extrapolation approaches using the same settings as before.

Measurement lengths of 2 to 12 months are used to predict the fatigue damage expected to occur in a second year. This means that, for example, 1st October 2015 to 30th November 2015, 1st October 2015 to 31st December 2015 and so on are extrapolated to the second year, i.e., 1st October 2016 to 30th September 2017. Again, to enable a statistical interpretation of the results, these predictions using different measurement lengths are repeated using the 13 "different" years that have been used before,
480 e.g., 1st November 2015 to 31st December 2015 and so on are extrapolated to October 2015 and 1st November 2016 to 30th September 2017 (cf. Fig. 12). This yields 13 different values for 11 different measurement lengths. The results for all three approaches are shown in Fig. 22 to 24.
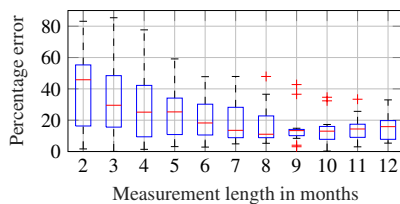
Figure 22. Convergence of the percentage error of predicted damage values using simple extrapolation for increasing measurement lengths. Box plot shows data from 13 "different" measurement periods.
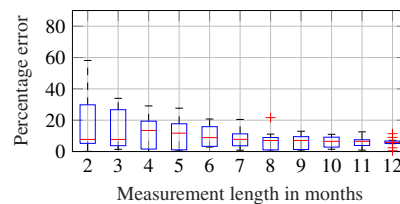
Figure 23. Convergence of the percentage error of predicted damage values using the binning approach for increasing measurement lengths. Box plot shows data from 13 "different" measurement periods.
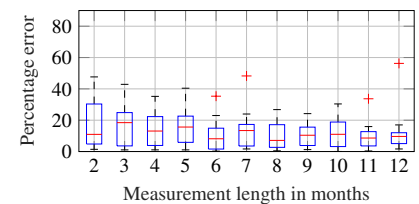
Figure 24. Convergence of the percentage error of predicted damage values using the ANN for increasing measurement lengths. Box plot shows data from 13 "different" measurement periods.

For the simple approach, nearly 12 months are required to achieve convergence. This is not only in accordance with results
485 from Hübler et al. (2018), who recommended a minimum measurement length of 9 months, but also logical. In order to cover seasonal effects, nearly a complete year has to be measured.

For the binning approach, the convergence is similar to that of the simple extrapolation method. For valid results, approximately 8 to 9 months are required. After this period, all bins – especially those for high wind speeds are critical – are filled with enough data for an accurate extrapolation. However, for the binning approach, this time can be reduced if the measurement period starts
490 during the winter. Fig. 25 shows the same convergence plot, but only for starting dates of the measurement period between 1st October and 1st February. For these starting dates, the relevant bins are filled with enough data within a few months. After two or three months, sufficient data might already be available; after a few months, additional data do not improve the prediction. The slight increase in percentage errors for measurement lengths of 10 months and more – visible in Fig. 25 – is probably only due to the limited number of different measurement periods; in this case 5. To sum up, if binning approaches are to be used, it
495 might be expedient to start measurement campaigns in the autumn.
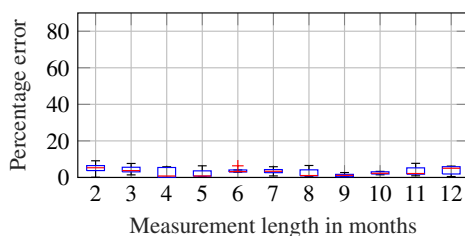
Figure 25. Convergence of the percentage error of predicted damage values using the binning approach for increasing measurement lengths. Box plot shows data from 5 "different" measurement periods starting in the winter.

For ANN, fairly accurate results can be achieved using data from a few months. Here, the advantage of determining a functional relationship becomes obvious if data is scarce. Nonetheless, it should be noted that even for a measurement length of only two months, ANN and the binning approach perform similarly well.

500 ## 4.5 Long-term extrapolations

So far, all extrapolations of fatigue damage have been conducted for two consecutive years, i.e., short-term extrapolations. This has the advantage that long-term changes not only in the environmental conditions, but also in the turbine, can virtually be ruled out. However, for real damage assessments used for lifetime extensions, an extrapolation over several years might be necessary. For example, if strain gauges failed after 5 years of operation and an lifetime extension is planned after 15 years, extrapolations
505 over 10 years, i.e., long-term extrapolations, are required. Such long-term extrapolations might be more challenging, as the "learned" correlation between environmental conditions and fatigue damage might have changed. Therefore, in the following, one year of measurement data is extrapolated to a second year which occurred several years earlier or later. For this purpose, data from the three years 1st January 2011 to 31st December 2011 and 1st October 2015 to 30th September 2017 are used. Similar to before, the starting dates of the two years are shifted month by month in order to realise a higher number of "different" years.
510 This means that the year 1st January 2011 to 31st December 2011 is extrapolated forwards to 1st October 2015 to 30th September 2016, 1st November 2015 to 31th October 2016 and so on. In addition, 1st October 2015 to 30th September 2016, 1st November 2015 to 31th October 2016 and so on are extrapolated backwards to 1st January 2011 to 31st December 2011, i.e. vice versa. This procedure yields 26 "different" years. A visualisation of this shifting procedure is shown in Fig. 26.
The results of the long-term extrapolation for the three extrapolation approaches, using the same settings as before, are shown
515 in Fig. 27 and compared to the previously determined results of the short-term extrapolations.
For all three extrapolation approaches, the resulting percentage errors are in a similar range for both short-term and long-term extrapolations. For the binning approach, the approximation even improves slightly for long-term extrapolations. However, it can be assumed that this improvement is only due to some random effects in the varying environmental conditions across the different years. Nonetheless, it can be concluded that long-term changes in the structural behaviour seem to be less pronounced
520 compared to variations in environmental conditions. Therefore, long-term extrapolations are easily possible, especially if the
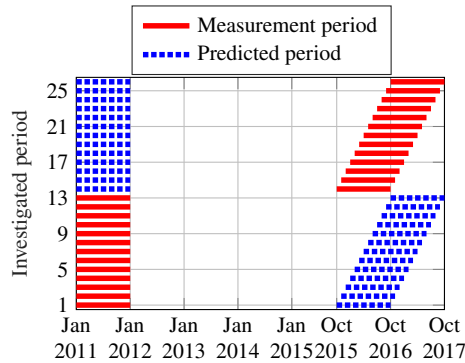
Figure 26. Visualisation of the 26 "different" periods for statistical evaluations in the context of long-term extrapolations.
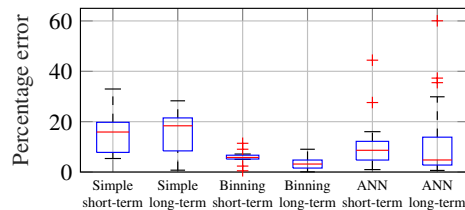
Figure 27. Percentage errors of predicted damage values using all extrapolation methods compared to real damage values. Data concerns both short-term (13 "different" consecutive years) and long-term predictions (one year extrapolated to a non-consecutive second year for 26 "different" years).

binning approach is applied. However, it must surely be asked where the boundaries of these long-term extrapolations lie. Is it still possible to use them if the structure has changed significantly, e.g., rotor blades have been exchanged? This question cannot be answered conclusively by this work, as much more data would be needed. Nonetheless, although the exact changes to the AV-07 turbine may not be mentioned here for reasons of confidentiality, it should be said that the AV-07 turbine has been

525 significantly modified during the period considered. Hence, long-term extrapolations are probably possible, even if moderate to severe modifications to the turbine have been made.

4.6   Extrapolation to the future

In all previous sections, it has been shown that the use of EOC data, i.e., SCADA data, is beneficial compared to a simple extrapolation based on pure strain data. However, it was always assumed that EOC data are available for the predicted period.

530 This is a valid assumption for nearly all predicted periods in the past, since SCADA systems feature a high availability and data quality. However, for extrapolations into the future, this is no longer valid. Extrapolations into the future are especially relevant for lifetime extensions. Hence, in this section, the last question of Section 4, i.e., "Do the approaches still yield reasonable results if extrapolations into the future are conducted, for which no EOC data are available?", is to be answered. For this purpose, the binning probabilities ($Pr_{\text{pred},i_1,...,i_d}$) in Eq. 8 are no longer determined using EOC data from the predicted (future)

period. Instead, past long-term EOC data are used. Long-term EOC data should be available for any lifetime extension, as the wind turbine has already been operating for 15 to 20 years. Moreover, it can be assumed that EOCs in the future period can be better predicted using long-term EOC data compared to using EOC data from the limited measurement period, i.e., the period for which strain data are available. The advantage of long-term EOC data is that random variations are more completely included. Using only EOC data from the measurement period could yield biased results if, for example, the measurement period features relatively harsh conditions compared to the long term. This is why, in this work, long-term probabilities for all wind speed bins ($Pr_{\text{long-term},i_1,...,i_d}$) are determined using EOC data from ten years (2011 to 2020). Using the long-term probabilities, an extrapolation to future periods should be possible, even if no EOC data is available for the predicted period:

$$D_{\text{future}} \approx N_{\text{n}} \sum_{i_1=1}^{M_1} \cdots \sum_{i_d=1}^{M_d} \left(Pr_{\text{long-term},i_1,...,i_d} \bar{D}_{i_1,...,i_d} \right). \tag{15}$$

For approaches using functional relationships, the adaptation for future periods is similar. Here, Eq. 11 remains nearly unchanged. The only difference is $\mathbf{x_k}$. For predictions into the past, $\mathbf{x_k}$ are the EOCs of the predicted period, i.e., $\mathbf{x_k} \in \{\mathbf{x_1},...,\mathbf{x}_{N_n}\}$. For predictions into the future, N_n random realisations (with replacement) of the EOCs of the long-term measurement period are generated, i.e.,

$$D_{\text{future}} \approx \sum_{k=1}^{N_n} f(\mathbf{x_k}) \quad \mathbf{x_k} \in \left\{ \mathbf{x_1},...,\mathbf{x}_{N_{\text{long-term}}} \right\}, \tag{16}$$

where $N_{\text{long-term}}$ is the number of short-term, e.g., ten-minute, intervals in the long-term measurement period.

For the simple extrapolation, there is no difference between extrapolations to periods in the past (with available EOC data) and the future (without EOC data), since the simple extrapolation is based on strain data from the measurement period only.

Results of all three approaches are shown in Fig. 28. In this figure, percentage errors of predicted damage for future periods are compared to previous results for which EOC data are available (short-term predictions of past periods).
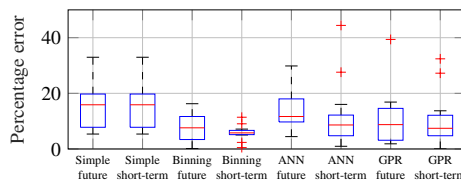
Figure 28. Percentage errors of predicted damage values using all extrapolation methods compared to real damage values. Results of predictions into the future using long-term EOC data are compared to previous results for which EOC data are available (short-term predictions of past periods). Thirteen "different" years are used in both cases.

By definition, for the simple extrapolation, an extrapolation into the future is equally accurate compared to an extrapolation to a period for which EOC data are available. For all other approaches, the results demonstrate that the quality of the prediction decreases slightly for extrapolations into the future. Nonetheless, predictions are still reasonable and yield lower percentage

errors compared to the simple extrapolation. Just like before, the binning approach also leads to the smallest percentage errors for extrapolations into the future.

560 In summary, extrapolations to future periods, for which no EOC data are available, are still possible with a relatively high accuracy provided that past long-term EOC data are used instead. Again, the binning approach is most suitable. Certainly, it must be kept in mind that the accuracy will decrease if long-term changes in the EOCs, e.g., due to climate change, become relevant. Hence, an accurate extrapolation of a few years into the future is possible, but an extrapolation 20 years into the future might be unreasonable.

5 Conclusion, benefits and limitations

To enable well-founded lifetime extensions for OWTs, the remaining useful lifetime has to be determined. Although several simulation-based and strain measurement-based approaches for determining the remaining useful lifetime already exist, especially for strain-measurement concepts, additional research is required. This work addresses the research gap regarding extrapolations of strain measurement-based fatigue damage calculations to other time periods.

570 Regarding the extrapolation in time, several approaches making use of the correlation of EOCs (ten-minute mean values) and short-term fatigue damage values are enhanced, assessed and validated using real offshore measurement data. The approaches are a simple extrapolation, a binning approach and two machine-learning approaches. To summarise the most important results:

1. User-friendly binning approaches yield accurate results.

2. More complex machine-learning approaches do not yield better results for the given data type, i.e., ten-minute EOC data.

575 3. It is sufficient to consider wind speed correlations only. Other environmental conditions do not need to be taken into account.

4. Consideration of different turbine statuses can improve the accuracy of the prediction. However, as it is not straightforward, careful consideration should be given to the question of whether it is beneficial.

5. The uncertainty of the prediction is moderate and no systematic bias occurs.

580 6. It is sufficient to measure strains for only a few months, if these months are winter months.

7. Long-term extrapolations over several years are possible, even if the OWT is moderately to heavily modified in this time period.

8. For extrapolations into the future, the accuracy of the prediction decreases, since EOCs have to be approximated using long-term EOC data. Still, reasonable predictions are possible.

585 Therefore, the results of this work demonstrate that user-friendly binning approaches are a suitable alternative or addition to simulation-based lifetime extensions, even if only limited strain data are available. However, some limitations of this work

should be discussed. First, spatial extrapolations, i.e., extrapolations to other locations on the same turbine and/or to other turbines in the same wind farm, are not addressed. For spatial extrapolations, the reader is referred to current research, e.g., Noppe et al. (2020) or Ziegler et al. (2019). Second, to determine correlations, only ten-minute EOC data are used in this work.

590 This is reasonable, since ten-minute SCADA data is nearly always available. Nonetheless, Santos et al. (2021) already showed that additional data, e.g., one-second SCADA data, is valuable for machine-learning approaches. Third, all present analyses are only conducted for one turbine and one location. Since other studies using different turbines – e.g. Hübler et al. (2018) – yield comparable results, we do not expect significant changes for similar turbines and/or other locations on the substructure and/or tower. For completely different turbines – e.g., floating OWTs – results might no longer be valid. Moreover, for locations

595 on the blades, the assumption of linear damage is at least questionable. Hence, to determine the remaining lifetime of blades accurately, the approaches might need to be adapted slightly. Finally, the analysed turbine is located in the North Sea. There, only small to moderate long-term changes in the environmental conditions occurred during the measurement period. For other locations and/or accelerating climate change, significant long-term changes in environmental conditions might occur. This might reduce the accuracy of the binning approach, even though the bin probabilities should cover most of these changes.

600 6 Outlook

Some of the previously mentioned limitations of this work immediately lead to future work. First, the results of this work should be assessed for other turbine types, for example, in other regions of the world, onshore, etc. This would increase the general validity of the results. Certainly, here, the availability of open-access strain measurement data is a limiting factor. Second, an analysis of extrapolation approaches for turbine blades would be valuable. The presented binning approach is quite

605 user friendly and based on limited data. Hence, it could be interesting for industry applications. If such a "simple" approach were available for blades and other components as well, this could be a useful extension. Third, the value of additional data, e.g., one-second SCADA data, should be analysed. For machine-learning approaches, additional data is valuable. However, it is not known whether it also improves the accuracy of binning approaches, since these are normally based on aggregated EOC data. Finally, a combination of temporal and spatial extrapolation methods would be an interesting addition. In this context,

610 spatial extrapolations cover predictions for other positions on the same turbine and also for other turbines in the same wind farm.

Data availability. After signing an agreement for the data usage, the raw data of the RAVE data archive – operated by the Federal Maritime and Hydrographic Agency (BSH) – is freely available for research purposes (https://www.rave-offshore.de/en/data.html).

Author contributions. Conceptualization, C.H.; Formal analysis, C.H.; Funding acquisition, R.R.; Investigation, C.H.; Methodology, C.H.;
615 Project administration, C.H. and R.R.; Software, C.H.; Validation, C.H.; Visualization, C.H.; Writing—original draft, C.H.; Writing—review & editing, R.R.

Competing interests. R. Rolfes is a member of the editorial board of the journal.

Acknowledgements. We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – *ENERGIZE* – 436547100. Moreover, we would like to thank the RAVE (Reasearch at Alpha Ventus) initiative for making the
620 data available. The RAVE initiative was funded by the German Federal Ministry of Economic Affairs and Energy on the basis of a decision by the German Bundestag and coordinated by Fraunhofer IWES (see www.rave-offshore.de).

References

Bartsch, C.: FACT-SHEET alpha ventus, press release of alpha ventus, https://www.alpha-ventus.de/fileadmin/Dateien/publikationen/av_Factsheet_de_20 20.pdf, accessed on 16.03.2021.

625 Bouty, C., Schafhirt, S., Ziegler, L., Muskulus, M.: Lifetime extension for large offshore wind farms: Is it enough to reassess fatigue for selected design positions? Energy Procedia, 137, 523-530, 2017.

Dimitrov, N., Kelly, M. C., Vignaroli, A., Berg, J.: From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases, Wind Energy Science, 3, 2, 767-790, 2018.

Dimitrov, N., Natarajan, A.: From SCADA to lifetime assessment and performance optimization: how to use models and machine learning
630   to extract useful insights from limited data, Journal of Physics: Conference Series, 1222, 012032, 2019.

Det Norske Veritas. Fatigue Design of Offshore Steel Structures; Recommended Practice DNV-RP-C203; Det Norske Veritas: Hovik, Norway, 2010.

DNV GL AS. Support Structures for Wind Turbines; Standard DNVGL-ST-0126; 4C Offshore: Lowestoft Suffolk, UK, 2016.

DNVGL. Lifetime extension of wind turbines. DNVGL-ST-0262, 2016.

635 Efron B.: Bootstrap methods: another look at the jackknife, The Annals of Statistics, 7, 1-26, 1979.

European Committee for Standardization. Eurocode 3: Design of Steel Structures – Part 1-9: Fatigue; EN 1993-1-9; European Committee for Standardization: Brussels, Belgium, 2010.

Henkel, M., Häfele, J., Weijtjens, W., Devriendt, C., Gebhardt, C. G., Rolfes, R.: Strain estimation for offshore wind turbines with jacket substructures using dual-band modal expansion, Marine Structures, 71, 102731, 2020.

640 Hübler, C., Weijtjens, W., Rolfes, R., Devriendt, C.: Reliability analysis of fatigue damage extrapolations of wind turbines using offshore strain measurements, Journal of Physics: Conference Series, 1037, 032035, 2018.

Hübler, C., Weijtjens, W., Gebhardt C. G., Rolfes, R., Devriendt, C.: Validation of Improved Sampling Concepts for Offshore Wind Turbine Fatigue Design, Energies, 12, 603, 2019.

Hübler, C., Rolfes, R.: Analysis of the influence of climate change on the fatigue lifetime of offshore wind turbines using imprecise proba-
645   bilities, Wind Energy, 24, 3, 275-289, 2021.

Iliopoulos, A., Weijtjens, W., Van Hemelrijck, D., Devriendt, C.: Fatigue assessment of offshore wind turbines on monopile foundations using multi-band modal expansion, Wind Energy, 20, 8, 1463-1479, 2017.

IWES – Fraunhofer Institut für Windenergiesysteme: Liste der Sensoren, technical report, https://www.rave-offshore.de/files/images/Datenarchiv/Datenarchiv_EN/Liste_der _Sensoren.pdf, accessed on 16.03.2021.

650 Larose, D. T., Larose, C. D.: Discovering knowledge in data: an introduction to data mining (Vol. 4), John Wiley & Sons, 2014.

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

Maes, K., Iliopoulos, A., Weijtjens, W., Devriendt, C., Lombaert, G.: Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms, Mechanical Systems & Signal Processing, 2016, 76-77, 592-611,
655   2016.

Mai, Q. A., Weijtjens, W., Devriendt, C., Morato, P. G., Rigo, P., 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, 2019.

Marsh, G.: Fatigue load monitoring of offshore wind turbine support structures, Doctoral dissertation, University of Strathclyde, 2016.

660 Müller, F., Krabbe, P., Hübler, C., Rolfes, R.: Assessment of meta-models to estimate fatigue loads of an offshore wind turbine, Proceedings of the 31st International Ocean and Polar Engineering Conference, 2021.

Natarajan, A., Bergami, L.: Determination of wind farm life consumption in complex terrain using ten-minute SCADA measurements, Journal of Physics: Conference Series, 1618, 022013, 2020.

Nielsen, J. S., Miller-Branovacki, L., Carriveau, R.: Probabilistic and Risk-Informed Life Extension Assessment of Wind Turbine Structural

665 Components, Energies, 14, 821, 2021.

Noppe, N., Hübler, C., Devriendt, C., Weijtjens, W.: Validated extrapolation of measured damage within an offshore wind farm using instrumented fleet leaders, Journal of Physics: Conference Series, 1618, 022005, 2020.

Petrovska, E., Le Dreff, J. B., Oterkus, S., Thies, P., McCarthy, E.: Application of Structural Monitoring Data for Fatigue Life Predictions of Monopile-Supported Offshore Wind Turbines, Proceedings of the 39th International Conference on Ocean, Offshore and Arctic

670 Engineering, 2020.

Rubert, T., Zorzi, G., Fusiek, G., Niewczas, P., McMillan, D., McAlorum, J., Perry, M.: Wind turbine lifetime extension decision-making based on structural health monitoring, Renewable Energy, 143, 611-621, 2019.

Saathoff, M., Rosemeier, M.: Stress-based assessment of the lifetime extension for wind turbines, Journal of Physics: Conference Series, 1618, 052057, 2020.

675 Santos, F., Noppe, N, Weijtjens, W., Devriendt, C.: Data-driven farm-wide fatigue estimation on jacket foundation OWTs for multiple SHM setups, Wind Energy Science Discussions, 1-36, 2021.

Seifert, H.: Monitoring fatigue loads using cycle counting data acquisition systems, DEWI Magazin, 7, 74-79, 1995.

Smith, J. C., Carriveau, R., Ting, D. S.: Inflow Parameter Effects on Wind Turbine Tower Cyclic Loading, Wind Engineering, 38, 5, 477-488, 2014.

680 Söker, H.: Determination of Fatigue Loads on Large Wind Turbines, DEWI Magazin, 8, 45-58, 1996.

Topham, E., McMillan, D.: Sustainable decommissioning of an offshore wind farm, Renewable Energy, 102, 470-480, 2017.

Weijtjens, W., Noppe, N., Verbelen, T., Iliopoulos, A., Devriendt, C.: Offshore wind turbine foundation monitoring, extrapolating fatigue measurements from fleet leaders to the entire wind farm, Journal of Physics: Conference Series, 753, 092018, 2016.

Ziegler, L., Muskulus, M.: Fatigue reassessment for lifetime extension of offshore wind monopile substructures, Journal of Physics: Confer-

685 ence Series, 753, 092010, 2016.

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

Ziegler, L., Smolka, U., Cosack, N., Muskulus, M.: Brief communication: Structural monitoring for lifetime extension of offshore wind monopiles: can strain measurements at one level tell us everything? Wind Energy Science, 2, 2, 469-476, 2017.

690 Ziegler, L., Cosack, N., Kolios, A., Muskulus, M.: Structural monitoring for lifetime extension of offshore wind monopiles: Verification of strain-based load extrapolation algorithm, Marine Structures, 66, 154-163, 2019.