Response letter to referee #1:

We would like to thank referee #1 for the comprehensive review and relevant comments. In the first review, we revised the paper based on given comments leading to significant changes in the structure and presentation of the work. We believe that the major comments are addressed previously, and now we have continued to refine based on the second comment letter. Below, are the initial comments that the reviewer believes is still relevant (in black text), their remaining concern about the comment (black text with bullet point) and our response to them (in blue text). The revised version of the paper is presented with traceable changes.

Note: Below, we have not included the comments which the reviewer approves the application of them in the previous round of review.

"Added value of site load measurements in probabilistic lifetime extension: a Lillgrund case study" (Manuscript number: wes-2024-68 - revision) Thank you for the revision of the manuscript. I still think that fatigue assessments are an important topic in the context of wind turbines. Furthermore, conducting them probabilistically and incorporating real strain gauge data is innovative. Nonetheless, the manuscript is still not sufficiently structured, explanations are still missing and even some mistakes have not been removed when revising it. As some comments have not been addressed sufficiently, I do not think that another major revision will solve all these problems. Hence, I cannot recommend it for a publication in the WES journal.

Comments:

The structure of the paper must be improved to make clear what the main innovation/topic is. Currently, it seems to be a mixture of "probabilistic fatigue assessment", "validation using real data" and "turbulence modelling".

• The structure has been improved. Nonetheless, sometimes it remains unclear. In the introduction, three scenarios are mentioned. However, on the one hand, only two of them are used later on (scenario I and III). Scenario II is only used for validation purposes. On the other hand, three groups of simulations are conducted, which could be linked to the scenarios. However, this is not done. Moreover, frequently measurements are mentioned, but sometimes it remains unclear whether it refers to load measurements or site-specific turbulence measurements.

Thank you for the feedback and the rest of the points. The points are addressed as below:

- The clarification that the study compares two scenarios, and the third one is only used for validation is now added in the introduction.
- There is no one-to-one relation between the simulations and the scenarios and only scenario 2 can be related to group 2 of the simulations. This connection is now elaborated in section 2.3.2.

- Very valid comment. Thank you. All ‘measurement’ words are not clarified through the text (mentioning which measurement we are referring to).

2) Abstract: It remains unclear what the topic paper is (see comment 1)

• See comment 1

Abstract is now modified with use of clearer (and to the point) wording.

3) In my opinion, the title of the paper does not represent in main topic of the work. Perhaps, turbulence modelling can be included in the title.

• Title has been changed but it still does not represent the topic of the work. What is the added value of load measurements? This is not really discussed.

The difference in the overall reliability level (corresponding to extra additional years of lifetime), are the added value of having measurements. This is more clarified now in the abstract to make the title more understandable. As mentioned before, the turbulence modeling uncertainty and extrapolation methodology are two side results of the main work but not the main objective.

4) Introduction: The connection between the assessment using the Frandsen model (simulation-based, l. 24- 44) and the limited data (measurement-based; l. 46-50) is unclear.

• Connection is still not clear. Does the limited data only refer to load measurements or also to turbulence measurements or even to scenario I, where limited simulation data are available?

With application of comment #1 (regarding referring to the type of measurement in the text), we believe this question is now resolved and the reader would understand what we are referring to.

5) The state of the art (L. 52-67) is not sufficient and does not clearly differentiate between simulation-based and measurement-based approaches.

• Has not been addressed

We agree that there is no relevant literature review for this point of novelty as it is a is a new topic and is not continuing of any previous research. However, for sake of the reader’s technical background for the topic we have emphasized on description of scenarios later in methodology instead of referring to unrelated literature. The two challenges which are addressed alongside the main topic (limitations of Frandsen model and extrapolation of loads) have related literature review which is included in the introduction.

11) Table 1: It is not clear for which time period the wind direction bin probabilities are given. Are these the probabilities for the same five years? And are they used somewhere. If yes, please highlight it. If not, you might just remove them.

• It has not been answered for which time period the probabilities (according to Vitulli et al.) apply. Furthermore, it remains unclear which probabilities are used for Eq. (8).

The probabilities are taken from a reference (as also mentioned in the title of table 1). Equation 8 is a general equation and is true for any duration. However, for sake of clarity, we edited the title of the last column in table 1 with emphasizing on the time duration.

12) Section 2.3.2: Your measurements come from an offshore turbine. The simulations seem to be done for an onshore turbine or all details regarding the offshore part are missing. Just simulating an onshore turbine and comparing it to offshore measurements does not seem to be sensible, even if you focus on blade loads.

• Even though there is some validation in the appendix (which should be in the paper itself as it is quite important), I still do not believe that you can directly compare the results of a simulated generic onshore wind turbine with an offshore wind turbine. I think that this is one reason why the reliability indices in Fig. 7 are so different.

The study does not focus on the support structure and only shows the differences regarding fatigue of blades' root (part of rotor nacelle assembly (RNA)). RNA is normally not much affected by the wave loading. However, this limitation of the work is pointed out in the discussion section to suggest future work.

13) L. 159: The site-specific turbulence distribution is not given, but only the reference turbulence intensity.

• Why are the distributions not given?

As distribution of the turbulence has not been used in this research, we do not include this extra information. However, the scatter of measurements is shown in appendix for reader's information.

14) L. 162: How has the exponent of 0.1 been determined using in-situ measurement data?

• Fine, but the values in Table 2 are wrong. 0.1 for group 1 and not for group 3.

Table 2 is correct and shows 0.1 for group 1 and not for group 3. We believe there is a misunderstanding. Please elaborate.

16) L. 174: For groups 1 and 2 you use Rayleigh distributions (covering wind data of full years) whereas the biased measurement data (see comment 9) is used for the strain gauge-based approach. Hence, a direct comparison, as in Figure 7 is not possible.

• I understand that the measurement data is not biased. Nonetheless, in the end, you compare annual reliabilities for group 1 and 2 (where the probabilities are determined using the Rayliegh distribution) and reliabilities for scenario III, where you use the actual data. Hence, the bin probabilities are different. Therefore, you cannot compare these cases directly. This might be another reason for the large differences in reliability indices in Fig. 7.

The intention is comparing these scenarios exactly as they are performed in lifetime extension assessments nowadays. The difference may be coming from different assumptions; however, the difference is exactly what it is when someone switches from one method to the other. This is the main intension of the current paper: to show how the result can be different.

18) Section 2.4.2: Formatting and explanations are not sufficient, e.g., Iy and not Iy, Ns is not explained etc.

• Formatting is still not completely correct

Formatting is corrected.

21) L. 240 and l. 247-264: For me, it is not clear, why we need all this. If I understand it correctly, you fit a distribution to the 10min values (step 1). Then, you sample from this distribution to determine the lifetime value (step 3 and 4). Why do we need the DELs with long return periods? A single DEL with a high return period does not influence the overall lifetime DEL. Hence, they are not relevant and actually not used for the reliability assessment in Section 2.4.4.

• I understand that the high DEL values have an influence on the lifetime DEL. Still, the relevance and even the execution of step 2 is not clear. You could immediately sample (step 3) from the distribution (step 1). However, you somehow form a database (step 2) using the extreme values. Does this mean that you determine a “new” distribution based on Eq. (10) to (12)? In this case, the question is how valid this distribution is, since Eq. (10) is only valid for the tail of the distribution.

The intention is to include the extremes (occurrences with higher magnitude and lower probabilities) in the samples. The effect can be important in high fatigue exponents as the references provided show (kindly see introduction for related literature and the motivation behind).

25) Eq. (14) to (17): Please, revise these equations, as they are not always correct, formatting has to be improved and explanations are missing, e.g., Δt and Pf are not explained, it has to be Iy and not I, the left side of Eq. (16) has to be $\Delta Pf(X, t + \Delta t)$, m not R etc.

• Partly done. You still use K and k; it is not stated that R is the stress ratio; c is defined as the diameter, but in line 221 it is the radius.

All mentions are using capital version ‘K’ to be consistent. R clarified in L. 319 of the revised document. Diameter changed to ‘radius’ in L. 304 of the revised paper and now we are consistently referring to it as radius through the document. Thank you for your comment and attention.

28) L. 313: You state that the Frandsen model and the ICE design underestimate the turbulence for low wind speeds and overestimate it for high wind speeds. I cannot see this in Figure 14, e.g., the Frandsen model is above the 75% quantile for 4m/s and below the same quantile for 20 m/s.

• The discussions about the Fig. 2 are still not correct, e.g., "the Frandsen model estimations are higher than design in high mean wind speeds" → see bin 1, 20m/s: circle (design) lies above square (Frandsen)

Thank you for your attention. The word 'higher' is now changed to 'lower' as it was a mistake in writing. Additionally, the sentence before is edited for sake of correctness.

30) L. 334-344 and Figure 4 and 5: Why do we need this? For Section 3.3, it is not needed.

• I understand that Figure 4 is useful (although I still do not understand step 2 (see comment 21)). Figure 5 just shows the convergence for higher N. This is not really needed. Perhaps, it is useful in the appendix. However, in the paper, you could just use a high N, as you finally did.

We hope that explanations to the comment 21 can help with providing more elaborations. However, we also would like to add that the relevance of the whole section is that (as mentioned in the introduction and abstract):

<u>The study is providing general solutions for using short/mid-term data for long-term fatigue assessment.</u> Although in the current study, the bias in not very large between the resulting DEL in 30 years compared to 5 years, the procedure shown is generally valid and useful.

We believe figure 5 helps with understanding the procedure by illustrating the outcomes of it and since this procedure belongs to the body, we have included the figure in the body of results.

31) L. 336: You state that "the probability of the largest data observed" corresponds to five years. However, this is not correct, since you do not have data of five full years.

• Even if your data does include measurements in all months for five years, this does not lead to a return period of five years. Only if you have measured the highest DEL by chance, it is actually the five-year extreme. In all other cases, it is below. How much below, you do not know. Statistically, it is probably around a two-year return period, as the amount of data sums up to two full years (line 129)

The return period is defined by the mean time between occurrences of a load. Although the possibility of not having the 5-year return period is accepted, we cannot agree to the fact that the unconscious duration of data will define the load level. As the measurements are collected through all months with a constant frequency, the possibility of hitting a load that occurs once every 5 years is not low (we agree it is not 100%). Overall, the extrapolation process which is shown here (which is the main aim of the process) would be the same. <u>The explanation of the limitation is now added to the discussion section.</u>

33) Table D1: How are the parameters of the different distributions defined?

• I asked about a definition, not about the determination of them. This is important, as, for example, Par1 and Par2 in a uniform distribution could be the lower and the upper limit, but also the mean value and the standard deviation, etc.

Apologize for misunderstanding. Definitions are now added to the title of the Table for clarification.

Typos etc.:

All comments are applied and a comprehensive edit is not and edit proofed by NREL.

Response letter to referee #2:

Below suggestions for minor revision of the paper are provided by referee #2 (black text). We would like to very much thank you the referee for the constructive comments and very valuable insights shared in both the comment letters. Our responses are shown in blue. The revised version of the paper is presented with traceable changes.

Line 257
The expression '30-year return loads' can be misunderstood – it is extrapolation to a certain quantile. Delete '30-year return loads'?
The same applies to text about 'maximum loads' which is confusing in relation to fatigue loads.
The extreme value theory is used to cover the extremes of DEL_10min. However, in the referenced instance, the wording is now corrected to avoid misunderstandings.

2.4.1
Frandsen model: In the Dr thesis by Frandsen probabilistic models for model uncertainty is described. Should be included

Thank you for your suggestion. The reference added now in the introduction:

'The Frandsen model involves simplified assumptions, and uncertainties. The uncertainty of the model in estimating the resulting fatigue load in a few examples is presented by Frandsen (2007).'

Eq (14)
It is essential that time t is included

Time added as to be equal to lifetime with additional explanation in the text. Thank you for pointing this out.

Table 3
Are COVs instead of std.dev. values shown?
Is it correct that there is (almost) no uncertainty related to log(DELlifetime)? Generally uncertainty related to estimating the fatigue load is the most important source of uncertainty. Typically the model uncertainty related to fatigue load is the dominating uncertainty, especially for high m values.
The mean value of logK is calibrated to obtain the annual reliability index of 3.7? the mean value and std.dev. of logK from the reference stated should be used.

Yes, the coefficients of variations are mentioned as shown in the title of columns.

Correct, the data in the field showed very low variability which is one of the reasons behind higher reliability compared to the simulation results. This is discussed in the discussion (the second paragraph in the discussion). If referring to the model uncertainty for load (i.e. uncertainty of the aeroelastic model or WTG model), we agree about the large effect on variability and uncertainty of the loads. However, the

CoV shown in the table is taken from strain gauge measurement data. If referring to fatigue model uncertainty (i.e. Miner's rule), that is represented by delta in the present work.

The mean of the DELs is confidential, and we are afraid that with providing the mean of the resistance, mean of DEL can be back calculated. Thus, we have avoided presenting the values. We hope this is understandable.

Section 2.4.3 and Figure 3
Why use extreme value theory to fit fatigue loads? A good fit to the upper tail can be expected to be important for estimating the fatigue life. Add more explanation and show the fit derived into Figure 3.

This sentence added to the text before the formulations of extreme value theory for more clarity and emphasize: '*As we are aiming at adding these low probability high magnitude occurrences, extreme value theory can be a suitable model to use.'*

The reasoning for including the high loads is given with reference in the same part of the text.

As for modeling the DEL_10min the focus has been capturing the extremes and extrapolation of them, we assume that showing the fit on the CDF is more relevant than on the PDF of data. Thus, the fit is shown in the exceedance plot (figure 4) and not on the pdf (figure 3). We hope this is acceptable.

Line 326-328: is it correct that the partial safety factors for blades/composites considered are calibrated to achieve certain reliability level? Please add a reference.

Yes, that is correct. Thank you for your suggestion. We added below reference text is added before presenting table 3:

'The mean of the material resistance is found through calibration. The calibration process entails finding a resistance mean value for which the target reliability is achieved at the end of the design lifetime.'