“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:

1) 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.
2) Abstract: It remains unclear what the topic paper is (see comment 1)
    - See comment 1
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.
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?
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
6) L. 121: Where exactly is the met mast situated? Please, show it in Figure 1.
    - Done
7) L. 121: Are shadow effects of the met mast considered, e.g., reduced wind speeds if the anemometer lies behind the met mast.
    - Done
8) L. 121: At which height(s) is the wind speed measured?
    - Done
9) L. 124: Your data is biased, as you only cover periods in the winter/spring. This should at least be discussed. Is this bias relevant for your work?
    - Done
10) L. 131: How much data has been removed?
    - Done

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).
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.
13) L. 159: The site-specific turbulence distribution is not given, but only the reference turbulence intensity.
    - Why are the distributions not given?
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.
15) Table 2: Why are the cut-in, the rated and the cut-out wind speed different compared to the real turbine (Section 2.1)?
    - Done
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.
17) Eq. (3) and (4) are not sufficiently explained, e.g., $di(\theta)$
    - Done
18) Section 2.4.2: Formatting and explanations are not sufficient, e.g., I_y and not Iy, N_s is not explained etc.
    - Formatting is still not completely correct
19) Eq. (8) and (9): At the left side of the equation, the expectation E has to be removed, as $DEL^m_{lifetime} = E(DEL^m_{10min})$ and not $E\left(DEL^m_{lifetime}\right) = E(DEL^m_{10min})$
    - Done
20) Eq. (9): Index i is missing.
    - My mistake.
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.

22) L. 245: You neither show the fitted distribution for the lifetime DEL nor you state what type of distribution it is.
    - Done
23) Eq. (10) where does this equation come from? It does not exactly match with Eq. (12), which is frequently used in literature.
    - Done
24) Eq. (11): This equation is wrong, as it gives negative probabilities, since the CDF is always between 0 and 1.
    - Done
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 $\mathrm{P_f}$ are not explained, it has to be $\mathrm{I_y}$ and not I, the left side of Eq. (16) has to be $\Delta P_f(X, t + \Delta t)$, 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.
26) L. 289: Why do you apply FORM and not MCS? Your limit state function can be evaluated computationally efficiently, so that MCS should not be a problem and MCS is more accurate.
    - I am not convinced, but it is fine for me.
27) L. 308: How do you define "enough data"?
    - Done
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)
29) L. 330: Why do you investigate this type of multi-modal distributions and not others?
    - Done
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.
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)
32) Table 4: How did you determine the sensitivities?
    - My fault, you did mention it
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.

<u>Typos etc.:</u>

34) L. 69: "assess" not "assesses"
    - Done

35) L. 86 and others: “Section” and not “Sect.” or “section”. Same applies to “Eq.”, “Table” etc. Or at least be consistent.
    - Not done everywhere, e.g., line 139 and 505, “table” and not “Table”.
36) L. 133: “in Table 1” not “in 1”
    - Done
37) L. 138: I think it is “D1” and not “D2”. Overall, reference to figures in the appendix are not always correct.
    - Done
38) L. 174: “Rayleigh” not “Reighley”
    - Done
39) L. 241: 365×24 … not $365 * 24$ …
    - Done
40) L. 346: “in Fig. 6” not “in 6”
    - Done
41) Figure 6: I_{ref} not Iref
42) Table 4 (and appendix): Do not use the notation 7.62e-3, but 7.62×10^{-3}
    - Not done in the appendix
43) L. 392: “fatigue” not “Fatigue”
    - Done
44) L. 419: “h and more”?
    - Done
45) L. 446: “In the following sections, we compare the turbulence levels in three scenarios of the study”?
    - Done
46) Caption of Table D1 has to be corrected.
    - Done
47) Caption of Figure D2 has to be corrected.
    - Done

Some other formatting errors and typos are now present due to the revision:

- Missing brackets when citing, e.g., line 19 or 23
- L. 24: “re assess”
- “Aero-elastic” in line 30 but “aeroelastic” in line 33
- L. 209: “wöhler” instead of “Wöhler”
- L. 229 and 230: “Cyclic loading”?
- …