“A machine learning based approach for better prediction of fatigue life of offshore wind turbine foundations using smaller datasizes” (https://doi.org/10.5194/wes-2025-173)

In this work, a new machine learning-based approach for strain measurement-based fatigue lifetime assessments of offshore wind turbine substructures is presented. The approach is based on a random forest model for the temporal extrapolation of the strain measurements. The approach is compared to state-of-the-art binning approaches and is validated using unique five-year data sets for a 3MW and a 9MW turbine.

The overall topic is very relevant, as many offshore wind turbines start to reach the end of their design lifetime and lifetime extensions based on strain measurements can be very valuable. The idea of using machine learning-based approaches for the temporal extrapolation is not new, although random forest models have not been used so far. However, although no very innovative approach is proposed, the work is definitely of interest to the research community, as the extend of the validation campaign is unique and allows long-term validation (or at least validation using data of five years). The results are interesting and confirm previous results but using a more reliable and broader database.

The paper is well written and nicely structured but would benefit from several minor clarifications and additions.

<u>Comments:</u>

1) In your work, you use SCADA and wave data for the extrapolation. Surely, SCADA should always be available. However, wave data might not always be available or at least not for the precise location (or close to it). Perhaps, you can briefly discuss how far away from the turbine the wave measurements can be conducted to be still useful.
2) In the abstract, you write “However, for longer datasets, greater than 12 months, the performance advantage of RF model over binning methods <u>becomes less pronounced</u>.” Based on the presented results and your discussion, I would conclude that “for longer datasets, greater than 12 months, the performance advantage of RF model over binning methods <u>disappears</u>.”
3) L. 105-113: Perhaps, you can make even more clear, what is actually missing and what has already been done. For example, the 4th bullet point has been investigated before, just as the 1st point (but only for a single turbine)
4) In Fig. 1, it is not clear what the green area/arrow between T_{moe} and T_{curr} is. In Fig. 4, this becomes clear. Nevertheless, perhaps, you can already explain it here.
5) L. 139/140: You write “During the monitoring period $\Delta T_m = T_{moe} - T_{moi}$, extrapolation models are trained using measured strain and corresponding SCADA/ wave data.” This sounds as if the entire data is used for training. However, if I am not mistaken, you also use the data for validation and testing. Hence, perhaps, you can reformulate it.
6) Fig. 2: It is not stated what we see on the axis (I assume that it is the probability). Furthermore, it is not clear why the wind rose of the 9MW turbine is bigger than the one of the 3MW turbine.
7) Section 3.1: You use quite a lot of factors/parameters (e.g., SCF, MSF, m, k, linearly or logarithmic spaced bins etc.). Although the precise values are not so relevant for this study, it would be nice if you could state them or refer to some publications where the reader can look them up.
8) L. 199: You state that using FA is always conservative. It this actually the case? For the 9MW turbine, depending on the site, it might be the case that SS is more damaging due to the reduced aerodynamic damping. Perhaps, you can comment on that for your site+turbine.
9) L. 209: The state “curtailed” mentioned here is not mentioned in Table 1. Furthermore, in l. 211, you state that “abnormal = curtailed” for the 9MW turbine. Perhaps, you can clarify this.

10) L. 252: You state "A filter-based method is applied using the built-in feature importance metric". Could you provide some more information about the feature importance metric or at least refer to a publication etc. where it is explained in more detail.
11) L. 253: You state that you use a 12-month period. Perhaps, you can add which one.
12) L. 259: "A recursive feature elimination with cross-validation is applied": Again, could you provide some more information or refer to a relevant publication?
13) L. 281: Not using optimized bin sizes is totally fine. Still, it would be nice to know what bin sizes you used.
14) L. 323: The SP is selected randomly. Is it allowed that the SP is so late that the measurement period goes beyond the end of the overall period or is the latest possible SP $T_{curr} - \Delta T_m$? In case, you allow very late SP, how do you do this, e.g., continue the measurement period at the beginning, i.e., at T_o? If you do not allow so late SP, you should briefly discuss this, as it adds a slight bias in the data selection (data close to T_o is less likely to be chosen).
15) L. 330/Fig. 5: Is it correct that you generate just a single synthetic 20-year period by resampling from the 5-year period. And then, you reduce it to one year again by taking the mean yearly damage. I do not really understand why you generate the 20-year period. Surely, this is useful if you want to analyze the scatter between different 20-year periods. However, as you only have a single one, you could just take the five-year period and determine the mean yearly damage from that or am I mistaken?
16) L. 354: What do you mean by "all months". I assume, all 20 SP + 10 ΔT_m + all approaches (binning 1D, 2D, ... RF)? At least this is what it looks like in Fig. 6.
17) L. 374: You describe that you generate 100 random years by resampling from the 5-year period. For each year, the fatigue life is calculated. In my opinion, this leads to too high uncertainty, as the fatigue life calculation is based on a single year. In my opinion, you should use 100 random 20-year periods to be consistent with the one 20-year period used for the extrapolation methods.
18) L. 381: You state "As data availability increases, predictions from all models begin to converge, with comparable performance across methods for data sizes beyond 18 months." In my opinion, for large data sizes and this case (3MW, single sensor), RF is outperformed by the binning approaches, as the uncertainty of RF is higher. It is totally fine that RF is outperformed by the binning approach for data sizes beyond 18 months. However, we should not present the ML approach in an overly positive light. For the 9MW turbine (l. 387), I agree with your statement that the difference diminishes.
19) L. 393-399: In my opinion, a discussion of the higher IQR of RF compared to the binning approach for the wind speed bin 21-22 m/s is missing. This higher IQR is probably also the reason why RF is outperformed by the binning approach in Fig. 6 (see comment 18)
20) L. 404: You state: "However, the magnitude of these fluctuations remains relatively low." Fig. 9 has no scale on the axis. Hence, it is hard to judge how high the fluctuations are. Nonetheless, if you look at the 5-95 percentile, it looks as if RF is not well performing for this wind speed bin (Fig. 9b, 25-26m/s).
21) L. 425: Again, I think that you highlight the benefits of RF too much. Not only "RF models maintain consistent and reasonable performance, particularly at smaller data sizes of 6–9 months" but 1D and 2D bins as well.
22) L. 435: "For both FA and single sensor targets, the 10−90th percentile of the prediction errors stabilizes after approximately 9 months of training data. In contrast, predictions for the SS direction exhibit continued variability." Where do I see this difference in stabilisation? I assume that the low value at 12 months (preventing a good stabilisation) is just a statistical artifact/outlier
23) L. 449: The statement "Additionally, the scatter in predictions is significantly lower for RF models than for binning-based extrapolation." is only true for short measurement periods. For longer ones, the scatter for RF models is sometimes even higher. Again, I do not say that RF has no clear advantages, but we should not present it in an overly positive light
24) L. 454: You state that effects of the "starting point of the monitoring period" can by analysed. This is true but I would be careful with the statement, as, in my opinion, such an analysis has not be done in this work.

You analysed the influence of the measurement length but did not check whether a starting point in winter/summer makes a difference.

25) L. 497: You state that "2D binning outperforms 1D binning". This has not been discussed anywhere in the paper and is not clearly visible in the data. I recommend to either discuss it somewhere else using the available figures or to remove this statement in the "conclusions"

Typos etc.:

26) L. 122: Remove space after "Objective section ".
27) L. 140: Remove space after "SCADA/ ".
28) Be consistent when writing "10-minute data" etc. Sometimes you write it with and sometime without hyphen.
29) L. 180: Remove space before (Hubler and Rolfes (2022)).
30) Table 1: I would use "s" and not "sec" and "°" and not "Deg"
31) L. 342: "the model comparison" and not "the models comparison"
32) L. 344: $LFFD_{factor}$ and not only LFFD
33) L. 350: Be consistent when using multiplication signs (· or preferable ×)
34) Table 3: Be consistent with the spelling of "windspeed/wind speed"
35) Table 3: "Turbine" should not be capitalized.
36) L. 431: Be consistent with the capitalisation of "Section" etc. and their abbreviations (e.g., Fig. vs Figure)
37) Table B2 and B3: Remove the * from "Operational State*" or state what it means.