

Response to RC3 on Data-driven farm-wide fatigue estimation on jacket foundation OWTs for multiple SHM setups

General comments:

There are some typing and/or grammatical errors that should be revised. For example, 'within the an offshore wind farm' in Line 114-115.

The manuscript was further proof-read.

One of the unique aspects of this work is stated to be the use of ANN estimated thrust load as an input for the fatigue estimator (Line 72). It would then be sensible to compare the effect of including or excluding the thrust load on the accuracy of the fatigue estimation. Only one such comparison is possible in Section 3.2.2 between scenario D and scenario F. The cross-validation for these scenarios however show that the inclusion of the thrust load increases the MAE which may result due to overfitting. This is briefly mentioned in Line 441-444, but for the farm wide evaluation (Section 4.3) the thrust load is included even though it has been shown to result in overfitting. It is recommended that the authors evaluate and discuss whether the use of estimated thrust load is beneficial to a model that can be used for farm-wide estimates or whether scenario D would not better address the goals of this work.

In addition, the impact of including/excluding thrust load is said to be studied in future work (Line 581). If the inclusion of the thrust load is indeed deemed to be unique to this work, it falls within the scope of this paper and should be addressed thoroughly.

This comment is much appreciated and has indeed made a very fair point which led to re-writing the entirety of section 3.4. The authors invite the reviewer to revisit this section. But shortly, it can be said that both scenarios (D and F) were included, as one presents better values for validation and the other for cross-validation. The overall trends are the same for both plots, apart from one outlying behaviour – which is further explained.

Comments:

Line 118: Rotational speed is also referred to as RPM which is confusing as the unit in which rotation is measured is cpm and not RPM. Recommended use of only rotational speed.

Correct to just present rotational speed

Line 119: Unit of ambient temperature is given as ° which does not distinguish between °C and °F.

Corrected to °C

Line 368: This is the first time that reference is made to the X direction in the text. At that stage it was unclear what the X direction is. It was found that Figure 2 also refers to X and Y, but it might be clearer if the text mentions that X and Y is the measurement directions of the accelerometers (perhaps in Line 121-122). Possibly also show in a figure like Figure 16 how these directions are defined.

Added X and Y as the sensor directions in line 122: “Apart from the accelerations in the two sensor directions, X and Y...”.

Line 388: Y vibration is said to be unnecessary. Is this due to yaw angles close to 0° or 180° for most data used during feature selection? If the yaw angle is close to 90° or 270°, the FA direction would be the same as the Y direction. Would this case not result in the Y SHM measurements to be better correlated to the fatigue? Perhaps the exclusion of Y measurements would not generalize for all conditions of the studied turbines or generalize well for all wind farms.

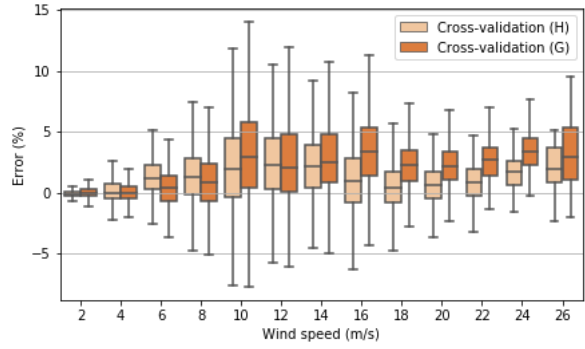
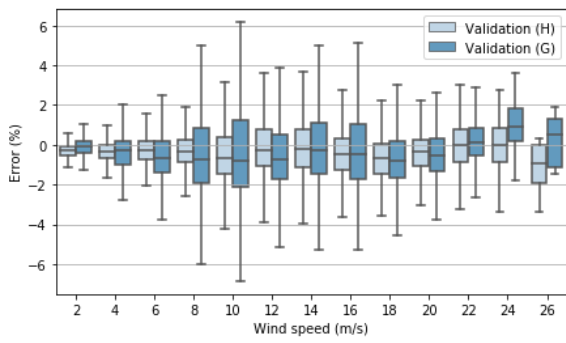
This is indeed correct, but feature selection should always be redone for different sites, and conclusions related with selected features are relevant for the current site. The sentence: “Apart from the accelerations in the two sensor directions, X and Y (with X in the dominant wind direction, often aligning with the fore-aft and Y 90° from it)” was added at lines 126-128. The sentence: “SS and Y SHM accelerations appear to be unnecessary for this site” was added at lines 400-401. The sentence: “The authors would like to point that the conclusions related to which features are to be selected is connected with the site at question and that it is a good practice to redo feature selection for a different site” was added to lines 421-423.

Line 424: The final scenario is that of the final instrumentation setup with only the selected features used as inputs for the ANN. It is recommended that an additional scenario is included in this section which uses the same instrumentation but no feature selection. This would be beneficial to see the impact of the feature selection on the errors for validation and cross-validation. This would then show the increased performance due to feature selection which is referred to in Line 229.

This suggestion is taken wholeheartedly, but the author is of the opinion that it isn't fully necessary. The main reason for the feature selection isn't just connected with improving the models' performance (which is, admittedly, stated in section 2.2.2.), but mainly with helping identify key features/input parameters and performing dimensionality reduction. Indeed, with over 430 engineered features, the training of models and processing of such an amount data would result in a very time-consuming process.

Line 451: The use of the word 'clearly' seems to contradict the initial understanding of Figure 10. This is evidenced by the need of the author to explain the lack of generalization in Line 452. If one were to choose a case that is generally 'best' based on Figure 10, it seems that scenario G performs better. Please explain how these results are used to determine the superiority of scenario H.

The use of the word 'clearly' is indeed abusive and contradicts the spirit of the text. The issue with Figure 10, and scenarios G and H, is that, albeit G presents a slightly better MAE cross-validation value, H has a better result for the validation turbine. Additionally, we can observe that the spread for both the box and whiskers of H is lower than G. If we compare the error (%) values for models H and G, we can see that for both cases the boxplots for model H have a lower spread for most wind speeds (including for cross-validation). The MAE might be marginally lower for G because it's general offset might be lower than for H. In sum, MAE is a valuable metric, but, as a single value, it can't show the entire picture. Altered sentence to: “Finally, the results for the model based on the RFE DTC selected features are the best for the turbine (H), as it presents the lower spread in its boxplots.”



Line 471: Is the 8000 datapoints each selected randomly or is a period of 8000 consecutive datapoints selected randomly?

8000 randomly selected datapoints (non-consecutive). Added 'non-consecutive' at the sentence referenced.