Response to 2nd round on Data-driven farm-wide fatigue estimation on jacket foundation OWTs for multiple SHM setups

The authors would like to express their acknowledgement to the reviewer and editor and thank them for their work.

Editor comments:

The authors meaningfully engaged with the reviewer comments by adding clarifying statements to the text. There are no apparent changes to the scientific content of the document.

A rewriting of section 3.4 aimed at clarifying the apparent lack of comparison between cases that include and exclude the estimated thrust load. This comment stemmed from the authors identifying the inclusion of the estimated thrust as the distinguishing component of the research while not seeming to thoroughly address whether this inclusion adds to the accuracy of predictions.

The rewriting of section 3.4 contains the two best options (one which uses the estimated thrust and one that does not) for farm-wide DEL estimation. Even though reference is made to both options, the authors offer no discussion comparing the superior of these two options. This leads the reader to use the comments from section 3.2.2 that the inclusion of thrust results in overfitting of the model which degrades the generalisation of the model to other turbines (Line 459-460). This apparently shows that the inclusion of thrust is not a good idea if a fleet leader concept is to be used.

The aim of this paper is stated to "assess the feasibility of this strategy, study the added value of various sensors and statistics and provide insight in how the most suitable parameters can be selected" (Line 76-77). This aim is especially relevant for the ANN estimated thrust load which is unique to this work (Line 72).

The authors however do not refer to whether this "unique" aspect was indeed beneficial to the creation of a farm-wide DEL estimator. From the presented results, it appears that the inclusion of the thrust load decreases the accuracy of the model when using it for farm-wide applications. The only reference to the importance of the thrust is not substantiated with external references or the results of this work ("It should, however, be said that, in itself, the thrust load is a highly relevant parameter, be it standalone, or if intended for further use in training.

It is recommended that the authors critically evaluate what the impact is of including the thrust load in the accuracy of a farm-wide model from the presented results and remove references to future work that will evaluate this impact.

It is the opinion of the reviewer that this would increase the clarity of this work. This comment was made during the first round of revisions after which the authors made changes to the document which did not completely clarify all objections of the reviewer. It is requested that the editor evaluate the validity of this comment and communicate with the authors if further changes is deemed to be necessary.

This is indeed a fair comment. And we acknowledge that the answer on whether to include the thrust load or not is not fully given in this work. However, in our opinion this was also not the key learning objective as we also look into the added value of e.g. 1s SCADA in general and a (high quality) accelerometer, a comparison of different feature selection strategies and a study on the feasibility of ANN for fatigue estimations of offshore wind turbines considering different restrictions in available data.

While initial working version(s) of the manuscript started by considering the thrust load model as a default part of the methodology (and perhaps this was still somewhat reflected in some passages), we now consider it as just another possible source of data. The text has been reworked (in particular the abstract) to reflect the reduced role of the quasi-static thrust load and avoid any remaining confusion.

To answer the question whether the thrust load model should be included we're currently restricted by the considered data. The limiting factor in this matter is that at this farm we only have two turbines instrumented that allow us to the validate the models. In the current research we opted to consider one turbine as our training turbine and one turbine for cross-validation. Or, in other words, we imposed ourselves the restriction that the training data would originate from just one turbine.
From the validation data we learn that the inclusion of the quasi-static load model does offer a clear improvement over all other models. However, from the cross-validation we now learn that this might translate in a slightly worse transferability of the model and we openly discuss so.
Arguably we could train our thrust load model on both turbines and likely this would result in a better performance towards cross-validation at a reduced validation cost. But in absence of more instrumented turbines we have no way to validate this model outside the training turbines. It is for this reason we opted to keep the discussion as is, concluding that the inclusion of the thrust model offers the best validation score, however we must be aware of the lower cross-validation score.
We've revised the text to assure this conclusion is consistent throughout the manuscript. We also put this more explicit in the manuscript's conclusion. We also make no claims that the thrust model is essential to the success of the algorithm.

Reviewer comments:

Figure 16: The colour scale for the side-by-side figures differs. This reduces the ability to compare the results generated from scenario D and scenario F. A consistent colour scale would be beneficial to the interpretation of the figure.

This comment is indeed valid. As such, figures 16(a) and 16 (b) have been altered to have a common scale and the colour map changed to better allow to identify outliers within each plot. It ought to be signalled that, albeit upon first glance the plots seem radically different in their values, this is due to the narrowness of the scale. Taking the example of OWT12: for the first plot it has a mean DEL of about 0.18, whereas for the second plot this value is 0.24. One could interpret this as being a difference of over 30% (0.24-0.18/0.18). However, as the mean DEL is already expressed as a ratio of the maximal training DEL value (in this case, and due to the considered period being summer months, the ratio will be rather small) the difference between both models would be simply 0.24-0.18=0.06 (or 6%). This is also reflected in Figure 10, wherein the difference of the mean between models D and F is about 4%.

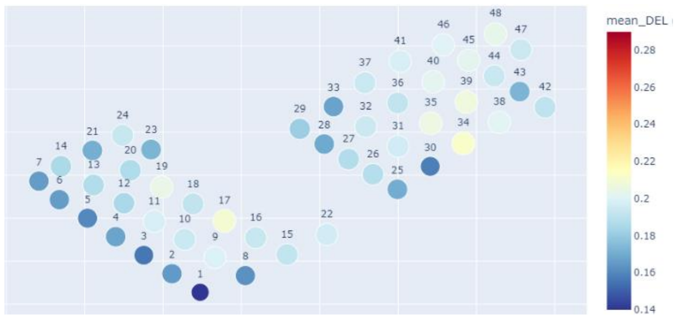

Figure 16(a)

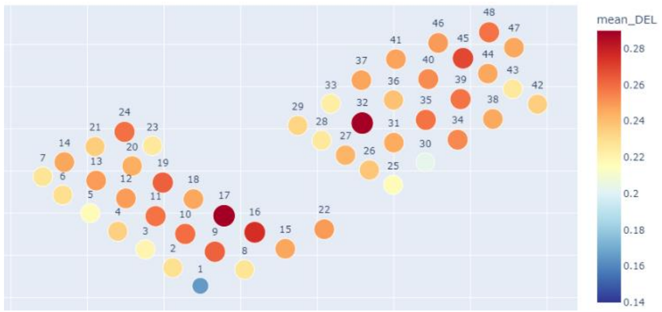

Figure 16(b)

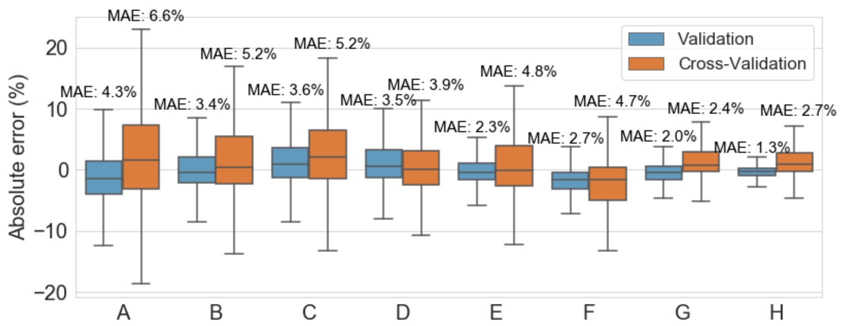

Figure 10

Line 574, Figure 18: Reference to the mean acceleration. Mean acceleration is usually zero as it oscillates around zero. Please clarify if this parameter is indeed the one used for plotting Figure 18.

The mean acceleration does indeed usually oscillate around zero – this is the case of the high-quality SHM accelerometers installed at three points of the tower. However, as the farm-wide implementation requires the use of low-quality accelerometers installed at nacelle level (present for all sites), this doesn't hold true, as can be attested by Figure 3 (green line). Indeed, these sensors present the absolute value of accelerations. In order to avoid confusions, the reference to absolute accelerations was added whenever 'low-quality accelerations' are mentioned in the text and captions.