Response to RC2 on Data-driven farm-wide fatigue estimation on jacket foundation OWTs for multiple SHM setups

Comments:

At the top of Figure 4, I read the expression "Measured thrust load". I would not agree that thrust is strictly measured, it is rather estimated from the strain measurements.

This is correct. The sentence: "*N.b for ease of writing, the thrust load deduced from the strain gauges is mentioned as measured thrust load, and to differentiate from the ANN's thrust load predictions.*" Was added to the caption of Figure 4.

At the top of Figure 4, I read that you provide a validation set for the thrust load. This indicates that the thrust has been validated independently somehow. Might you please elaborate on this point? Were validation performed via an aeroelastic simulation? In order to get a good estimate of thrust, the strain gauges need to be calibrated for a know load level. How was this done, in order for one to trust that the Thrust estimates are correct?

The aforementioned figure was indeed problematic, as it led the read to believe that the thrust load was being validated. In fact, the terms 'validation' and 'training' of Figure 4 relate to the validation and training of the thrust load ANN model – this has been corrected in the figure to show 'Training & Validation' for both the 1s SCADA and thrust load.

On page 9: I suggest that authors explain the basis by which they selected the features of the time series.

The reason which features were included is related with Vera-Tudela's study. In the text (line 220) it is highlighted that Vera-Tudela's paper signals spectral moments, skewness and kurtosis has features of interest, as they aren't that commonly used, but the other features are also present in this reference. This sentence has been altered to: "The selection of which engineering input features should be calculated can be traced to Vera-Tudela and Kühn (2014)".

Section 2.2.2: another powerful feature selection approach is https://pypi.org/project/BorutaShap/

This is indeed a good suggestion and, has the current research of the group will again involve feature selection, it will surely be added to the list of techniques.

Page 12 line 289-290: "carefully selected as to be representative of all operating conditions". What are the criteria for such a selection? Turbine operating in partial load, full load and transition? Turbine operating at max Cp and rated power? Operating conditions covering all possible combination of pitch-tip Speed Ratio? Please explain what are all the representative operating conditions.

The sentence: "…, namely parked, run-up and full load" was added in line 302.

Page 12 line 292: why do you correct the thrust for the air density? Are you calculating the thrust coefficient?

The answer to this question can be found in Noppe, N., Weijtjens, W., and Devriendt, C.: Modeling of quasi-static thrust load of wind turbines based on 1 s SCADA data, Wind EnergyScience, 3, 139–147, 2018: "According to Baudisch (2012), thrust loads are influenced
by air density. While changes in the depicted SCADA variables happen within seconds, air density changes on a different timescale (several hours). Instead of including air density
in the set of input parameters, it is accounted for as a correction of the modeled thrust load $\widehat{F_T}$: $\widehat{F_{T,corr}} = \rho \widehat{F_T}$". Both references have been added to that section of text.

Page 13 line 299: ANN as used in this article cannot extrapolate, i.e. they cannot make correct predictions when the input are not within the range of the training set. How do the authors ensure that the validation data (3 months of data outside of the training period) fall within the range of the training set?

The training data was ensured to be statistically representative of all conditions faced during 1-year worth of data. This is further discussed in d N Santos, F., Noppe, N., Weijtjens, W., and Devriendt, C.: Input parameter selection for full load damage neural network model on offshore wind structures, in: Proceedings of 16th EAWE PhD Seminar on Wind Energy, 2020 for the DEL model. The sentence: "… carefully selected as to be statistically representative of all operating conditions …" has been added at line 301.

Page 13 line 300: the cross-validation applied to a different turbine. Has the cross-validation set been chosen in such a way to reflect the conditions that occurred in the original training set? i.e. you cannot cross-validate on a set where the other turbine is known to be in a waked condition, right?

In relation to wake, even though the training turbine is mostly under free flow, it also includes training periods in which it is under wake. The training turbine is mostly free-flow facing, but can be understood as under partial wake. The assumption was that the present of partial wake would be taken into account by the model and suffice – apparently this assumption doesn't hold.

Table 1: you want to consider adding this to the mix of methods:
https://pypi.org/project/BorutaShap/

Answered above.

Page 15: please note that random forest feature importance can be derived either based on mean decrease in impurity or based on feature permutation. Both methods could give slightly different results. Mind you that the results could also be affected by dependent input features. Please make sure you do not bias your feature selection based on the above. Kendall's Tau takes such dependence into account, and probably why it gives the lowest MAE in figure 7 at the cost of high number of features

The author acknowledges the reviewer's comment and will keep it in mind for future research. In the present contribution it was only used for comparison and not for the final results.

Section 3.2.2: Table 2: once 1-second SCADA data is made available, by default then the 10min-SCADA can be calculated. Did you consider a scenario where both 10min and 1Hz SCADA are made available to the model?

The 10-minute mean of 1s SCADA is the same as 10-minute data. What changes is simply the availability of features (with 1s SCADA, apart from mean, you can have spectral moments, etc.)

Section 3.2.2: it is perfectly acceptable to propose several scenarios of various data sources, and check their effect on the model predictive error. However, wouldn't a more principled approach involving sensitivity analysis and sensors selection optimization with value of information be more adequate? Please discuss.

One of the driving research questions behind this contribution was to understand the models' performance degradation when accelerations aren't included in the model. It was also driven by a desire to understand what some real-world sensor setups would allow to achieve – a goal that is mirrored by a common operator concern, namely, is it worth installing X or Y sensor, and what am I gaining/missing out with each installation. The author isn't entirely acquainted with value of information analysis, but from its understanding value of information requires a cost-benefit analysis on sensor uncertainty, which is currently out of scope. Nevertheless, the author will begin shortly collaboration with an external partner aiming precisely at quantifying the uncertainty of these models and the gain in reducing those uncertainties.

Page 20, line 472: please elaborate on the 80-20 train-test split. Do you respect the temporal evolution of the data or do you perform randomized split?

The train-test split was randomized (this information has been added to line 486).

Page 21, line 489: "…for a different turbine". Please specify where this other turbine located with respect to the reference turbine where the model was trained.

Added: " – located in the turbines' wake –" and "The training turbine is located at the north-western edge of the farm and the cross-validation turbine located in the middle of the farm (*cf.* subsection 3.4., OWT 7 and OWT 35 Figure 17)" at lines 489-491.

Page 22, line 505-508: this might the case, but the question is whether the training set of the reference turbine included any data corresponding to wake? Indeed not all wakes/partial wakes/multiple wakes are created equal because of dependence on atmospheric stability, turbulence and shear. It is worth discussing this issue. The logical consequence of this is that your training set for the reference turbine should include a much larger amount of data in order to take into account the various wake effects in order to generalize to another turbine…

This is indeed true, albeit the training turbine is under partial wake. For reference, the training turbine was OWT 7 and the cross-validation turbine OWT 31. Naturally that, the larger the training dataset, the better the training turbine will be able to adapt and generalize. One of the questions that arise is whether the use of a transfer function is able to bridge this higher need of data. Or a population-based approach, for which a future approach is planned.

Page 22, line 505-508: 8-18 m/s according to the reference turbine where the model was trained or according to the different (second) turbine? If the second turbine were wake affected, then its wind speed should account for the velocity deficit.

The binned wind speeds of the validation turbine are of this turbine. The same for the cross-validation turbine.

Page 22, line 505-508: Avendaño-Valencia, L. D., Abdallah, I., and Chatzi, E.

The sentence "Avendaño-Valencia et al. have worked in this direction, concluding that the fatigue life of OWTs under free-stream inflow can be quite distinct from OWTs under wake (Avendaño-Valencia et al.)" has been added to lines 523-526.

Figure 16: what is mean by mean_DEL? DEL is a short term measure of fatigue conditional on wind speed. Is the mean calculated by weighting according to the pdf of the wind speed? Please elaborate how the mean is computed in this case.

The mean here represents the simple arithmetic mean of every 10-minute DELs. Added "(arithmetic mean of all 10-minute DELS)" to caption

Figure 18: same comment as above.

Same as above.

Figure 16: DEL across various wind turbines in the farm will be highly influenced by the mean wind speed (at each wind turbine), turbulence and shear. Maybe you would want to plot Figure 16 for various wind speed bins (e.g. 6-9m/s, 10-14m/s and >15m/s). You will notice that the DEL across wake-free and wake affected wind turbines will change quite a bit.

This is indeed a good point, but it would add yet another topic to an already rather lengthy paper and the main objective of subsection 3.4 was to merely present a farm-wide application of the contribution's methodology, without going too much into detail. Nevertheless, it is in the researcher's plans to dedicate a proper study to farm-wide wake.

General comment:

There are multiple grammatical and orthographic mistakes, and a proof-read is necessary.

Revised

It is not mandatory, but you might want to consider comparing the performance of your method to other state of the art methods dealing with a similar subject matter.

This comparative aspect has been briefly explored in the introduction section (1.2), from a methodological POV.

Methodological Suggestion:

A more direct approach avoiding the two-tier approach proposed in this article might be to use variational auto-encoders neural networks such as proposed here:

https://onlinelibrary.wiley.com/doi/full/10.1002/we.2621

https://link.springer.com/chapter/10.1007%2F978-3-030-12075-7_21

These references have been added to the introduction section (line 104): "As for *Mylonas et al.*, it used conditional variational auto-encoder neural networks to estimate the probability distribution of the accumulated fatigue on the root cross-section of a simulated wind turbine blade, making long-term probabilistic deterioration predictions based on historic SCADA data (*Mylonas et al.,2020, 2021*)".