Response to RC1 on Data-driven farm-wide fatigue estimation on jacket foundation OWTs for multiple SHM setups

-The methodology section should improve a lot. Section 2.2.1 discusses Tier one which is about the estimation and validation of thrust load. Then Section 2.2.2 is focused on feature selection so that it reviews different techniques but does not clearly answer the question that which technique is applied to this work. The reader expects to realize what you have applied in the methodology section. Please provide references for the techniques that you have mentioned there. Maybe it is better to name the Section methodology, then creating two subsections Input data and the Proposed Estimation Algorithm. Please improve the structure to make it easier for readers to follow the work.

The text on the methodology section is indeed confusing. The paragraphs: "The main methodology of the present contribution can be understood as a *two-tier* neural network model. The first tier concerns itself with the generation and processing of relevant 10-minute features, which will serve as inputs for the second tier. In the first tier, an ANN model is utilized to estimate the thrust load on a 1s-basis. This model has to be trained, validated and cross-validated before being considered fit for employment (upper white-background section of Figure 4). After this, the 1s thrust load, along with SCADA and accelerations from the SHM accelerometer, is processed into a variety of 10-minute metrics (*cf.* grey section of Figure 4). Between the 10-minute feature generation tier (grey section of Figure 4) and the DEL prediction tier (pale yellow section of Figure 4), the 10-minute metrics undergo a dimensionality reduction procedure based on a feature selection algorithm, which allows to train, validate and cross-validate a 10-minute ANN FA DEL estimation model for a smaller amount of relevant input features. The motivation behind a 10-minute approach lies with the common framework for data processing (also 10-minute), the aim of including environmental effects and vibration levels, but also the issues inherent to working with different sampling frequencies (1*Hz* SCADA, 12.5*Hz* accelerometer) and possible time-delays. After training, validating and cross-validating the DEL ANN model (middle white-background section of Figure 4), we can observe this second tier in pale yellow in Figure 4}. Albeit the current contribution's methodology can be seen globally in Figure 4, the sections which are under a white background consist of training, validation, cross-validation and determining which features ought to be used. Thus, these tasks will only be performed once. When a final model is achieved, only the sections with the grey and pale yellow background are required to produce estimation - this will be picked up again by subsection 3.4." as been added and the author hopes that it is sufficiently enlightening to ensure the continuity of the current structure of section 2.

-Please explain why ADAMAX is used for optimizing ANN training. What are other alternatives and what is making ADAMAX different in this context?

In the manuscript it is explained that ADAMAX is selected through hyperparameter tuning, which is a commonplace term in machine learning to refer to trial-and-error. The choice on an ADAM based optimizer is also explained as being a common practice in the machine learning community: "A common, well-performing choice…". This optimizer was compared to a number of other optimizers available in keras and selected for being the best performing. The sentence: "This particular optimizer was selected through hyperparameter tuning by comparison with other available optimizers, such as Stochastic Gradient Descent (SGD), Root Mean Squared Propagation (RMSProp) and Adadelta (Zeiler, 2012), an extension of the Adaptive Gradient Algorithm, Adagrad." was added to the manuscript.

-Is it possible to provide a confidence interval for the estimation error by applying different techniques like what is used in the following publication: Online condition monitoring of floating wind turbines drivetrain by means of digital twin, FK Moghadam, AR Nejad - Mechanical Systems and Signal Processing, 2022

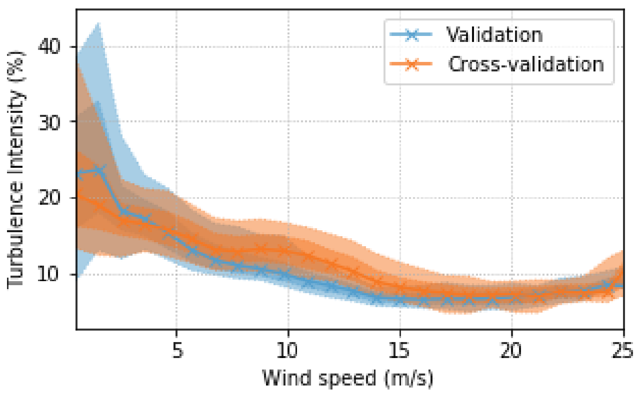
This suggestion is taken wholeheartedly, but it is the authors' opinion that it would further increase the size of an already big article. Nevertheless, the authors continue their work on this research and will attempt to follow this suggestion in future work.

-Please list the general specification of the under investigation turbines, including the rated power and speed.

The sentence: "These turbines have a rated power of 6 MW and a maximum of 12 RPM." has been added to line 118.

-Could you give a comment on Figures 6 and 13 results, and explain in the text how the error is correlated to wind speed?

The discussion of how the error is correlated with wind speed is present for Figure 6 in: "The increase of MAE with the wind speed is expected, as higher loads are attained for higher wind speeds (and the MAE relates to the absolute value of the loads)." As for Figure 13, it is said: "we can observe how the model performs worse for a different turbine for mid-range speeds (8-18 m/s) rather than for higher speeds. This is possibly related with the worse performance of the model under wake." This comment indeed highlighted a deficiency in the explanation of how the DEL prediction error is correlated with wake, which doesn't show as a linear correlation as for Figure 6. The sentence: "This can be further verified by inspecting Figure 14, where the turbulence intensity is plotted against the wind speed. As turbulence intensity (TI) is given by $TI\ (\%) = (\sigma(u)/\bar{u}) \times 100)$, where $\sigma(u)$ represents the standard deviation of wind speed and $\bar{u}$ the mean wind speed, the values of turbulence intensity for lower wind speeds are rather unstable and thus, disregarded in this analysis. In this figure we can see how, for wind speeds between 5 and 18m/s the turbulence intensity of the cross-validation turbine is noticeably higher than for the validation (and training) turbine. We can then reasonably assume that, precisely because the cross-validation turbine faces higher turbulence (being located under more severe wake) than the training turbine, the prediction error for the cross-validation turbine will be highest in the regions were the gap between turbulence intensities is greater". and the following figure, with the caption: "Turbulence intensity (%) *vs.* wind speed (*m/s*) for validation and cross-validation turbines." were added.



-The highest error in DEL does not correspond to the highest error in thrust load estimation. Could you add a comment on that? Is it possible to do a sensitivity analysis to see which parameter contributes the most to the DEL error?

Thrust load and total DEL don't have a 1:1 relation – the computation of total DEL, although dependent of the thrust load, is also the fruit of complex interactions between thrust, SCADA and accelerations. It should also be noted that Fig. 6 and Fig. 13 are comparing two different metrics: Fig. 6 presents the MAE (in % of maximal training thrust) computed at every 10 minutes and Fig. 13 has the Error (as a percentage of the maximal training DEL). The suggestion of performing a sensitivity analysis to see the contribution each parameter has on the DEL error is indeed much appreciated and taken wholeheartedly, but the author feels that the current manuscript is already rather extensive. Nevertheless, the research on this topic is ongoing and this suggestion will surely be included in future work.

-Could you move the schematic diagram presented in Fig. 15 related to the farm-wide DEL to the methodology section?

The author hopes that the discussion of the first comment regarding the methodology makes this suggestion unnecessary. The sentence: "Figure 15 schematizes the pipeline for the farm-wide model, consisting in the parts of Figure 4 which are kept for producing DEL estimations - the 10-minute feature generation and DEL predictions tiers, in grey and pale yellow, respectively". has been added to page 24 line 548.

-Line 125 "To avoid excessive drifts in this transformation a lower frequency bound of 0.1Hz is used.", please elaborate more or possibly provide a reference.

It is explained in "Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms" how, for low frequency strains, induced by the thrust-loading there's a concern with the conversion from accelerations to strains which requires a double integration, risking blowing up the low frequency noise in the measured accelerations, resulting in large errors in the low frequency components of the predicted strains, requiring a high pass filter to prevent this (see Section 3.5, page 15-17, K. Maes et al., Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms, Mech. Syst. Signal Process. (2016)). This reference has been added to the manuscript.

-Line 152 "In the current contribution we focus primarily on the DEL estimation in the FA direction, as this is considered most relevant for the current jacket foundations." Could you provide a reference?

The sentence was altered to: "In the current contribution we focus primarily on the DEL estimation in the FA direction, as measurements reveal it to be the most relevant for the current jacket foundations."

-Line 170 "The thrust load can be obtained from measurements by low-pass filtering the bending moment timeseries (with an upper frequency bound of 0.2 Hz)." Why is the cut-off frequency chosen to be 0.2 Hz?

It is explained in "Performance monitoring and lifetime assessment of offshore wind farms based on SCADA data" how up until 0.2 Hz the frequency spectrum can be interpreted has being in a quasi-stactic (thrust load) region. Above 0.2 Hz begins the low frequency region driven by wave dynamics and the first eigenfrequency. (See Fig. 5.5., page 112, N. Noppe, Performance monitoring and lifetime assessment of offshore wind farms based on SCADA data). This reference has been added to the manuscript.

-Site-to-side in figure 2 should change to side-to-side.

Altered.

-Multiple grammatical errors. e.g. “a intermediate” in Line 27 that should be corrected.

The manuscript was further proof-read, and errors corrected.