Wind Energ. Sci. Discuss.,
https://doi.org/10.5194/wes-2017-46-AC1, 2017


eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

# *Interactive comment on* "High frequent SCADA-based thrust load modeling of wind turbines" *by* Nymfa Noppe et al.

**Nymfa Noppe et al.**

nymfa.noppe@avrg.be

Received and published: 22 December 2017

Thank you for your nice and interesting comments. We agreed to most of them and already discussed them during the project. Following you can find our answers and how we suggest to adjust the manuscript.

Both reviewers suggested to revise the title of the manuscript. We agree with them and therefore the title will be modified to "Modelling of quasi-static thrust load of wind turbines based on 1 second SCADA data".

From several comments of both reviewers, we realized the motivation behind the thrust



load estimation and the use of 1s SCADA data is not obvious. As 1s SCADA data was accessible for this research and is becoming the standard in industry, the main goal was to explore its possibilities to estimate fatigue loads. Given the sampling frequency of 1 Hz, only low frequent loads (lower than 0,5 Hz) can be estimated. In addition it was soon clear the first order tower dynamics were not present in the SCADA signal. Therefore the load spectrum that can be estimated based on 1s SCADA data is limited to the quasi-static thrust load.
Previous research within our group showed fatigue loading can be estimated using a combination of strain gauges and accelerometers. For several reasons accelerometers are preferred over strain gauges, but they are not suited for quasi-static loads. The strain gauges are thus crucial to capture the quasi-static part of the loading. The research presented in this manuscript aims to replace the strains gauges by the SCADA-based approach. In future research the proposed thrust load estimation can be combined with the use of accelerometers to estimate the total fatigue loading. This motivation will be explained better in the introduction.

Both reviewers asked for clarification regarding the time delays mentioned in the paper. During this research, an autocorrelation was performed between a thrust signal and several SCADA signals for multiple periods. Results showed the observed time shift changed for different signals but also for different periods. The first attached figure (CorrelationCoefficient_vs_timeshift) shows the correlation coefficient between a thrust signal and 4 SCADA signals of 2,5 months, where a time lag of -15 to 15 seconds was considered between the signals. In general, the differences in correlation coefficient are fairly low. We decided the added value of including this figure to the manuscript was small, but based on the comments following sentence will be added to page 5, line 20: *When calculating the autocorrelation between the thrust load signal and shifted SCADA signals, the biggest time shift was found for the pitch signal and corresponded to -3 seconds.*
Since the maxima never exceed 5 seconds, it was chosen to include the 5 previous

timestamps for every selected SCADA parameter as an input to the neural network. Moreover a small sensitivity analysis (trial and error) revealed that including less previous timestamps in the input set of the neural network made the obtained results worse, there was also no gain beyond 5 seconds.

Both reviewers pointed out that Pearson correlation only applies for linear correlations. We fully agree and calculated mutual information too during this research to consider non-linear correlations. Results were comparable to those obtained with Pearson correlation. Since results were comparable and for sake of simplicity it was decided at the time to only include Pearson correlation in the manuscript. Since this rightfully raises questions, the results for mutual information will be included as well in the revised version of the manuscript. The most important observation in these results is that all selected parameters are clearly correlated to the thrust load based on the complete dataset, whereas the Pearson correlation of the pitch angle to the thrust load was low when all data was considered.

Some comments given by both reviewers concerned the topology of the neural network. This topology was chosen by the authors in the beginning of the project. Three hidden layers were chosen because three operational states can be distinguished for a wind turbine ('non-operating', 'operating below rated power' and 'operating at rated power'). Moreover 4 neurons were chosen in each layer because 4 input parameters were selected. The second attached figure (MeanRelativeError_vs_different_topologies) shows the possible gain in mean relative error of the test set by using a different topology. In this case only topologies are considered with the same number of neurons in each layer. These results show the error is not influenced a lot by the topology, as long as more than one neuron is used. The topology was not optimized afterwards because results were already satisfying. Following sentence will be added to the contribution: *By choosing a different topology the mean relative error*



*of the test set improved with maximum 0,2% if more than one neuron is chosen in each layer.*
All the other needed settings for the neural network are kept as suggested by the Neural Network Toolbox of MATLAB.

Based on several comments, we understand that the motivation behind the use of simulations is not fully clarified in the manuscript. During the project we decided to use simulated data, which provided a controlled and reproduceable environment, in order to understand the real measurements better. However the simulated turbine (and site) is not the same turbine, although comparable, as the real one. Therefore the simulated thrust load cannot be compared to the measured thrust load. We included the results based on simulated data to show the method works for different types of turbines and to provide a reproduceable environment to optimize the technique in the future. If desired we can make the simulation data publicly available. Several adjustments will be made to the manuscript. The abstract will be modified to motivate the two-step approach of both simulation and real-world measurements. Moreover section 5 and section 6 will be merged to one section "Results". This section will start with following text:
*The proposed approach is validated using two different datasets. The first one is obtained by simulation in FAST, while the second one is obtained thanks to a measurement campaign performed at an offshore wind turbine. The dataset obtained by simulations was included to illustrate the approach in a controlled and reproduceable environment.*

Smaller comments are addressed in the following.

*"+ abstract, expand if possible by the number of words, to justify under which criteria 15% error is assessed (both in time and quality)."*

The error is assessed between the direct simulations/measurements of the thrust load and the predicted result. The abstract will be adjusted.

*"+ page 1, line 21, is 10m SCADA meant sampled data or statistics¿'*
Statistics, will be clarified in paper.

*"+ page 2, line 30, more information about the filtering process is needed. Although it is data dependent and designer driven, a fair reproduction from the manuscript will be impossible without this step. Please indicate what %-data were removed."*
Values outside the interval [mu-3*sigma;mu+3*sigma] are removed. In total, less than 6% is removed. This will be added to the paper.

*"- page 3, Fig. 1, increase size of fonts (a), circles in (b) did not show in print."*
Font size and caption will be adjusted.

*"+ page 3, line 1, indicate how many samples did the data contain or its duration if record was continuous."*
The length of the investigated data-set (one year) will be mentioned in the paper.

*"!+ page 4, Fig 2c, looking at it reminds me of heteroscedacity (even more when looking at Fig 4c and 6c). Have you consider it?"*
This is an interesting point. We'll add a comment on the increased variability of the modelling error with the wind speed to the text. It is however not a path we have investigated further, but we are open for suggestions.

*"- page 6, line 11, when the inputs are mentioned it is now clear that 10-minute data*



*was only used to explore correlations, or I did get it wrong?"*
The 10-minute statistics are indeed only used to explore correlations.

*"- page 6, line 20, without knowing the number of data points in your data set, the question is why was hold-out preferred over cross-validation? check Hastie, the elements of statistical learning for a discussion."*
The training set used for the offshore wind turbine consisted in over one million data points after filtering. Given the big dataset hold-out was preferred to reduce computational cost. Moreover, during the project we decided to use the default settings proposed by the Neural Network toolbox of MATLAB. Since we were already satisfied with the result using these default settings, we chose not to optimize the settings. Therefore, optimization of the method is still possible and using cross-validation instead of hold-out can lead to better results but was not investigated during this research.

*"* page 7, line 2, a reference next to '... proposed by the software...' is needed. "*
The reference will be added.

*"- page 16, line 15, the decision to represent 90% of the data, was it for clarity of the graph or for other reason? The use of different language (see Fig 2, referring to quantiles, rises my question)"*
It is exactly the same, the sentence will be rephrased.

*"- page 9, line 19, there is a typo 'With a(n) median...' just found it by chance."*
Thanks, it will be adjusted.

*"* page 10, line 2, I don't agree on the statement 'these offsets won't influence a*

*fatigue assessment', as the distribution of offsets and variance is not included in the analysis, this needs to be provided next to its verification. Furthermore, since damage would have an exponent in the operation, a small difference would be increased dramatically."*
We consider a fatigue assessment independent from offsets assuming this fatigue assessment is performed according to common practise in industry. This means performing cycle counting and the Miner's rule. Since this is based on the size of the cycle and independent from the mean level, we concluded offsets won't influence a fatigue assessment. This will also be clarified in the manuscript.

*"- page 11, line 1, I would rather specify the use of 1s SCADA data."*
This will be specified in the paper.

Thank you very much for your nice comments and helpful review!

---

Interactive comment on Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2017-46, 2017.



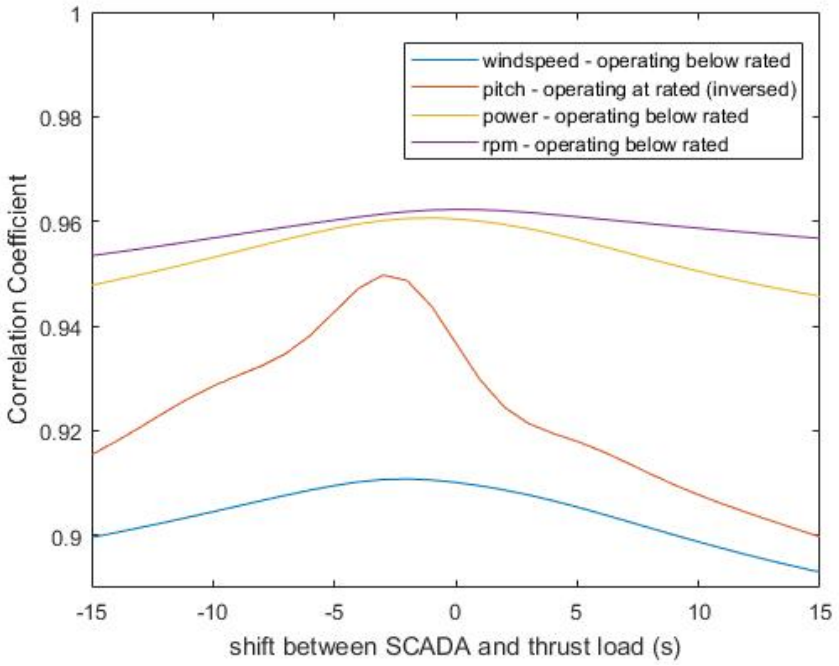


**Fig. 1.** CorrelationCoefficient_vs_timeshift

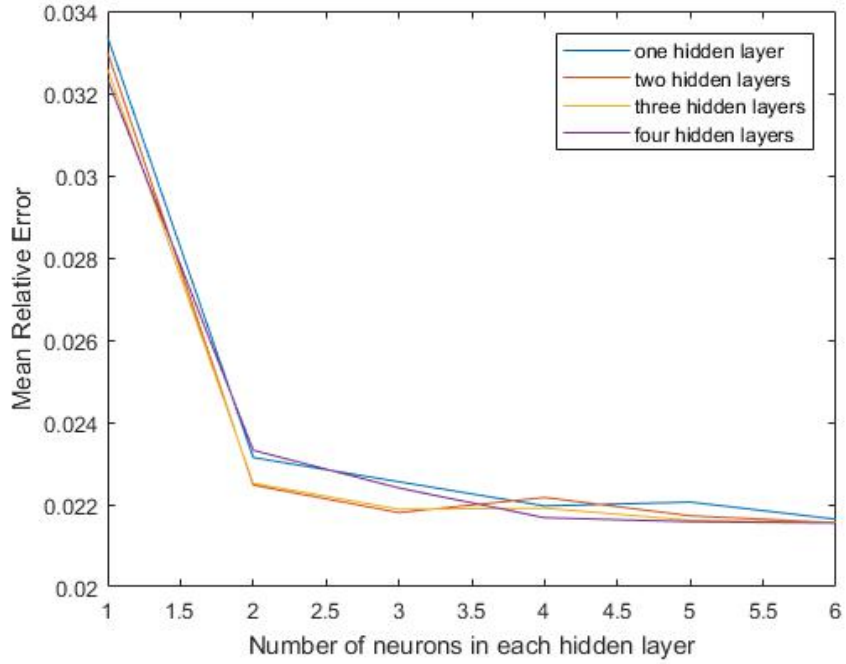


**Fig. 2.** MeanRelativeError_vs_different_topologies