

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

**Anonymous Referee #1**

Received and published: 8 November 2017

A well written manuscript that details the estimation of thrust loading on wind turbines using 1s SCADA data.

Although many, the requests for corrections and clarifications are minor. Positive feedback is also added to point out relevant pieces in the contribution. Congratulations to the authors on such good work!

Thus, I add a notation to facilitate working the comments: \*must, +clarify, -minor, !great.

All the best,

\*\*\*\*\*

- title, it states 'high frequent' whereas it uses 1s; somehow this could be made explicit,

C1

you never know if another person is already working on 10Hz, so make it part of the title '... frequent (1s) ...'

+ abstract, expand if possible by the number of words, otherwise in the introduction, to justify why were the simulations needed and under which criteria 15% error is assessed (both in time and quality). I kind of infer from the lack of direct measurement, but it is better if explicit.

- introduction, explain the, typically non-written, motivation behind the use of 1s, which a priori represent more information over 10m data; was it defined according to methodology, simply accessible or the like.

+ introduction, modelling in a time-series domain instead over 10m statistics have advantages. What goals did you envisioned? Where is the knowledge gap?

+ page 1, line 21, is 10m SCADA meant sampled data or statistics?

+ 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.

- page 3, Fig. 1, increase size of fonts (a), circles in (b) did not show in print.

+ page 3, line 1, indicate how many samples did the data contain or its duration if record was continuous.

+ page 3, line 9, (also page 5, line 19) either here or in the discussion, time-lag analysis would have indicated the statistical duration of lag between series. It is of interest to understand its impact on the model (see page 6, line 13 regarding the use of previous records) and on the estimation of thrust (excitation) after moments at the bottom of the tower (response) is recorded.

!+ page 4, Fig 2c, looking at it reminds me of heteroscedacity (even more when looking at Fig 4c and 6c). Have you consider it? I think is of great value that you included

C2

residuals next to results! It helps to understand limitations of the model, bring it to the discussion.

- page 4, line 4, Pearson was used, OK fair, but why not to consider Spearman or information coefficient to account for non-linearities?

- page 6, line 9 (then line 13), was the architecture and the number of previous records optimized? are they used in all channels?

- 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?

- 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.

\* page 7, line 2, a reference next to '... proposed by the software...' is needed.

\* page 7, line 2, if I understood correctly, simulations were meant to be site-specific, please share more information about the model used and whether it was validated or not.

- page 7, line, the comparison of measurements to simulations is not clear in the overall aim of the manuscript (a statistical description to assess if both processes behind the data are similar or not will lead to other focus), it does not make it worst, but when the reader misses the point, it might have missed the relevance of such an effort

+ page 7, line 14, one of the weakness (to increase the rate of the paper) is that estimations are verified versus simulated values, which rises the question on its relevance. Would the evaluation assess measured  $M_{tn,m}$  data as a previous step, then the inductive procedure would appear more robust.

- 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

C3

quantiles, rises my question)

- page 9, line 19, there is a typo 'With a(n) median...' just found it by chance.

\* 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.

- page 10, line 19, please give an numerical range to your definition of significance (a reference would be even better), people will understand it differently.

- page 11, line 1, I would rather specify the use of 1s SCADA data.

! the tone of the paper is entertaining and well balanced, congratulations!

\*\*\*\*\*

---

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

C4