

Interactive comment on “The Super-Turbine Wind Power Conversion Paradox: Using Machine Learning to Reduce Errors Caused by Jensen’s Inequality” by Tyler C. McCandless and Sue Ellen Haupt

Anonymous Referee #2

Received and published: 22 March 2019

Review: This paper takes the very interesting paper by Jensen and applies it to the over-prediction or under-prediction of power in wind turbines, with increasing inaccuracy as the Standard Deviation of the wind increases. The paper also looks at this for real turbine data and shows the same effect. The authors then use a machine learning method to remove this effect. I am not sure the machine learning section is necessary but the paper provides a novel contribution and makes for an interesting discussion.

Section 2.3: this analysis would be more interesting with some percentage errors. What is the mean error and what is the maximum error? It might be small for this set,

C1

it looks like it from figure 3, but this could be larger for other datasets. Perhaps some discussion on this would be interesting?

Section 2.4: Can you justify the SD of 2ms^{-1} ? Is this value high or low for real data sets?

Section 3: It would be good to track how many datapoints there were at the beginning, I can work it out but easier if it is in the paper.

Section 3: I am not sure of setting the powers above 2020kW to a maximum, they should be removed as they are erroneous, or remain as they demonstrate an offset in the reading.

Section 3: I am not sure about setting the negative powers to 0, they should be removed or remain for the same reasons as above.

Section 3: What did the dataset look like? Some statistics to define the scatter and shape of the data would be very beneficial. What is the correlation of the speed to power? While I believe that the change is due to Jensen’s inequality it would be nice to ensure that this was the case.

Figure 7 shows some interesting distributions for the data and it would have been interesting to see these earlier.

Section 5: It would be good to better justify the use of random forest regression, what other methods are available and why this one?

Section 5.2: It would be nice to see the training and testing errors to look for over- or underfitting.

Technical corrections: The use of the term super-turbine is not clear, I have not seen this term used before. The definition appears to be hidden in Figure 1, but I am still not sure if it is the authors term or a general term. In addition the exact meaning of the term is not clear. p.1 Line 36: I am not sure this is Artificial Intelligence, more machine

C2

learning. P.2 Line 13: For completeness it would be nice to state that this is only for increasing convex and concave functions. P.3 Line 15: It would be clearer if the line read 6-13 ms^{-1} . p.3 Line 22: It is quite clear that Gauss is the correct distribution as this is not a long-term study. The additional comments make the discussion more confusing. p.3 Line 26: A 10th order polynomial fit would seem to be high, it is generally good practice to use the lowest order fit possible, is 10th order really the lowest order that can be used in this scenario? p.9 Line 1: Reference is not formatted correctly. Figure 2: Please use the journal referencing style of the website Figure 6: The quality of this figure seems poor.

Interactive comment on Wind Energ. Sci. Discuss., <https://doi.org/10.5194/wes-2018-74>, 2019.