Utilizing Physics-Based Input Features within a Machine Learning Model to Predict Wind Speed Forecasting Error

Daniel Vassallo, Raghavendra Krishnamurthy and Harindra J.S. Fernando

REVIEW

GENERAL COMMENT:

The paper by Vassallo et al. presents an interesting contribution that uses a ML approach to predict wind speed forecasting error. The paper is overall well-written, and the methods used described in detail. The final discussion of the results is also well-laid.
Nevertheless, I have some major points that require clarification before the paper can be recommended for publication. In addition, the analysis presented in the current version of the manuscript, though interesting, could definitely be expanded given the unique dataset the authors are considering.

MAJOR COMMENTS:

1. I find the choice of the authors of (extensively) describing the methods used before the data a bit confusing. I personally had to go back to the methods section after reading the data section to make sure I got everything right. I would recommend switching the order of the two sections.
2. More clarification is needed on what averaging time is used in the calculations of the variables considered in this work. For example, what averaging time is used to calculate the Reynolds decomposition for turbulent averaging, for example for TKE, TI, friction velocity? Why did you choose it? How does that conciliate with the different lead times of the ML models?
3. In addition, the authors should better clarify how the random split of train-test set mentioned in the paper is implemented. Do you mean that you are randomly picking 25% of data for testing, and then using those times stamps, once the algorithm has been trained, to predict wind speed 10-min, 1-hour, 3-hour ahead of each of the randomly selected time stamps in the testing?
4. With such a huge data set as the one used in this analysis, I feel like the results shown could be greatly expanded, as a lot of additional analysis relevant to the topic could be made. After all, you are using 2 sonic anemometers out of an array of almost 200. For example, how does the performance of the used ML algorithms vary with atmospheric stability? Or with height? Do you find that different input features are more relevant close to the surface compared that at let's say hub-height? Or how do the results vary in different complex terrain locations, for example comparing results from the valley and from the ridge tops? Please consider adding more analysis to this piece of work.

SPECIFIC COMMENTS:

5. Abstract: introducing the symbols of each feature are not necessary in the abstract.
6. Figure 2-a: the map is not super clear.
7. Figure 2-b: not really needed.
8. Page 4: was wind speed at Perdigão really stationary? Over which time scales? Please clarify.
9. P.6 l.10: rephrase as “a feature set that utilizes all input features is tested”.
10. Did you apply any cross-validation for your ML models? If not, why?
11. P.8 l.1: please specify what you mean by “Sensors at 20 and 100 m AGL were chosen based on data availability.”
12. Please state the native time resolution of the sonic data you are using.
13. Have sonic anemometer data been filtered for tower wake effects? These effects would artificially increase turbulence (and reduce wind speed) for some wind direction bins, thus invalidating the quality of quite some data.