

Interactive comment on “The Power Curve Working Group’s Assessment of Wind Turbine Power Performance Prediction Methods” by Joseph C. Y. Lee et al.

Ding Yu (Referee)

yuding@tamu.edu

Received and published: 11 November 2019

I am Dr. Yu Ding, Professor at Texas A&M University and the author of book "Data Science for Wind Energy." Despite being published only very recently, I am pleased to see that the authors cite this in their paper. It is also a great please to be asked to comment on this paper.

I personally participated in some of the PCWG meetings and had quite some discussions with Andy Clifton (when he was with NREL) and Jason Fields. I laud the collective effort of PCWG, especially the tremendous amount of work done initially by Peter Stuart and later by Taylor Geer for putting the group together, although I have unfortunately

C1

not had the chance to meet either of them yet.

The work done by PCWG is important and will have a long-term impact for wind industry, particularly nowadays when government subsidies to wind energy are reducing or going away altogether. Performance enhancement and market competitiveness is the key to making wind energy self-sustainable. That is the common goal for all of us working in this area.

I am glad that the PCWG eventually summarizes this work for publication. When I was writing Chapter 5 of my book, I had been looking for a formal source to cite the work of PCWG. Had this paper been published earlier, I would have cited it in the book. Nonetheless, this is still not late and the paper will have good impact on wind researchers and practitioners alike. I do not have any serious concerns, barring some suggestions and comments below, and believe the paper should be published.

1. One thing rising to the level of a concern is the use of NME. I find it troublesome to use NME as the primary criterion to indicate fitting/prediction quality. Two functions, f and g , can be everywhere different, and the difference can be infinitely large, yet the NME can still be zero. This is different from NMAE, because when NMAE goes to zero, it guarantees the two functions, f and g , will converge to each other. In comparison studies or decision making, NME can be used but generally not as a primary criterion, and its use has to be carefully administered to serve a limited purpose, due to its fundamental theoretical flaw, namely when NME goes to zero, it does not guarantee anything concerning the similarity between f and g .

In the paper, it is said that the use of NME is for the long-term purpose while the use of NMAE is for short-term, and also that NME has a direct impact on P50. These statements are questionable. P50 means the 50-percentile (or the 0.5-quantile) of a probability distribution, which corresponds to the median of the probability distribution. To estimate a median, one minimizes the NMAE not NME. As a matter of fact, no statistical analysis or any objective function used in statistical modeling/machine learn-

C2

ing/data science optimizes NME (because of NME's fundamental flaw). The statement saying that NME has a direct impact on P50 is not grounded in well-established scientific literature. This paper cited Clifton (2016) as the basis for this statement. Based on my understanding of Andy's technical rigor and prowess, I will be surprised if Andy supported the current statement (although he is a co-author of this paper).

To say that NME is for long-term while NMAE is for short-term is not principled either. In fact, NME can be particularly misleading when it is evaluated through a long-term average. Suppose that NME is calculated through a monthly average and its value is zero, but the day-ahead power prediction and the actual power can be different everyday, and the daily differences can be unboundedly large. NME being zero simply means that the prediction is more than the actual half of the times and less than the actual the other half of the times, but the more and the less cancelled each other. On surface, this zero NME indicates a perfect power curve model but an owner/operator using this model will lose money everyday (when doing day-ahead bidding). At the end of the month, the owner/operator suffers a huge monetary loss, and will puzzle how a "perfect" power curve model gets him/her into such a mess. On the other hand, if using NMAE, this problem won't happen.

As a side note, rigorously speaking, P50 is not the average (or mean) but the median. To estimate the mean (average), one should optimize an objective function based on RMSE (i.e., squared error loss). Of course, when a distribution is symmetric, like Gaussian, the two statistics, median and mean, become the same. Still, it is better to be rigorous when using the terminologies.

2. The paper identified PDM as an effective way to make accurate power prediction. This is not surprising. PDM is the same as multi-dimensional binning, which is a non-parametric method having sound data science foundation. The main limitation of PDM is its lack of scalability. One-dimensional PDM is not problem, 2D is fine, but 3D may be as far as PDM can go. If each variable is split into 20 bins, 3D PDM has 8,000 bins, and 4D PDM has 160,000 bins, and 5D PDM has over 3 million bins. But there are

C3

only 50,000 or so 10-min data points for a whole year. In the end, the number of bins that need data points increases so fast in a multi-dimensional data space and they will far exceed the number of data points available, an issue known as the curse of dimensionality. This argument was made in our paper, Lee, Ding, Xie, Genton (2015) "Kernel Plus method for quantifying wind turbine upgrades," *Wind Energy*, 18: 1207-1219. It is the fundamental reason why I believe that data science/machine learning methods will in the end win out, because they can handle the scalability better (See Chapter 5, *Data Science for Wind Energy*). I hope that the authors are amendable to make a comment on the issue of "curse of dimensionality".

3. When PCWG started in 2012, the trend of data science methods and its practice was not yet popular or widely aware of outside the CS/Tech field. Back then, PCWG designed its specific way for intelligence sharing for understandable reasons. It is still a good practice. Nevertheless, after seven years, I hope that PCWG can adopt some of the tactics used by other academic/industry communities for identifying the best practice. For instance, the image processing challenge of 2013 administrated by CVPR identified the deep convolutional neural network (CNN) as the best tool for image processing (although people still try to understand, to this very day, why so). What the CVPR community did is to release a large amount of training image samples and reserve the test samples and ground truth for validation. The challenge administrator did not impose any restriction in terms of methods or programming languages. The testing results are done and averaged over thousands of test samples to effectively reduce biases.

If PCWG is interested, it can do something similar to CVPR's image processing challenge. Specifically, PCWG can go through the following steps:

a. Collect enough datasets with wind/environmental inputs and power outputs from a wide variety of turbines. Normalize the power to $[0, 1]$ and sanitize datasets to remove any farm/company identifying information.

C4

b. Split the data into training and test. Release the training datasets, as well as the input data of the test sets. Both in pure text files.

c. A participating method, identified by a unique ID, known only to the participant and PCWG, makes prediction of power on the test set using the provided input and then produces a single column of power predictions.

d. PCWG's website engine takes the prediction and compute the error metric using the (never released) actual power output values. The prediction performance is released using the participant's ID. Again, I hope the error metric is not NME but a more principled choice.

4. The last comment of mine is about the different purposes that the IEC standard power curve method and a predictive power curve (used by owners/operators) serve. In my book, I have written the following, and wonder how PCWG in general, and the authors of this paper in particular, think about the difference.

Chapter 5, page 150, Data Science for Wind Energy, "Recall that we mention earlier in this chapter that the IEC method's intention is to provide a benchmark for verification purpose, rather than providing a method for real life performance assessment or wind power prediction. Consider the following analogy in the context of vehicle fuel economy. At the time of sale, a new car is displayed with a published fuel economy, in the unit of miles per gallon. The published fuel economy value is obtained under a standardized, ideal testing condition, which cannot be replicated in real-life driving. A car's real-life fuel economy based on someone's actual driving is almost always worse than the published one. In the wind power production, engineers observe something similar—using the IEC binning power curve often leads to a conclusion of under performance, which is to say that the actual power production falls short of the prediction.

Still, car buyers and car manufacturers feel that the fuel economy obtained under the ideal condition provides a reasonable benchmark, offering some ballpark ideas of how fast a car consumes its fuel. However, for consumers who care very much about the

C5

real-life fuel economy, such as in commercial driving, they are not advised to use the published fuel economy value, as using the published value will surely lead to biased estimations. The same conclusion should have been extended to the IEC method, but in actuality, in the vacuum of robust, reliable, and capable power curve models, the IEC binning method is routinely used in the tasks or for the missions it is not designed for."

Once again, I commend the great work done by the authors and the remarkable effort in summarizing PCWG's intelligence sharing into a formal publication. I support its publication, although I hope the authors will take some of my comments into consideration while revising the paper (in particularly, NME). I look forward to continual interaction with the group and contributing to this important area.

Interactive comment on Wind Energ. Sci. Discuss., <https://doi.org/10.5194/wes-2019-69>, 2019.

C6