

Atmos. Chem. Phys. Discuss., referee comment RC2  
<https://doi.org/10.5194/acp-2022-634-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on acp-2022-634**

Anonymous Referee #2

---

Referee comment on "Estimating hub-height wind speed based on a machine learning algorithm: implications for wind energy assessment" by Boming Liu et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2022-634-RC2>, 2022

---

## **Revision of acp-2022-634**

### **General comments**

The work describes the use of Machine Learning algorithms to evaluate the potential of the use of wind energy in different locations in China. While the results are interesting, there are some revisions that need to be taken before it can be accepted. First of all, the Introduction section needs to emphasize better the research gaps that this work aims to fill, with clear and updated references to current literature. It could be interesting to know also if there were similar attempts previously and if yes, how this work is different; if not, why a trial was not made? The rest can be summarized. As for the methods, the data and instruments used need to be better described. Also, a description of the study site for a non-Chinese could be worthy. The description of the ML methods is not understandable for a non-expert. Limitations of the methods and of the data are never explained. Finally, as for the results: the discussion should be improved, citing relevant literature to explain them. An effort must be made to take the discussion to a higher scientific level, now it is limited to a qualitative description of the plots, but reasons for the findings are seldom given, often without citing relevant literature. Sometimes, the description of the processes is also not right, or at least superficial. This for instance applies to the effect of turbulence on wind speed. Or for instance to the factors that need to be taken into account into the WS120 estimation: why are they important? Or why are they not? A strong limitation of the work is that the comparison of observations with model estimations is carried out at a single location, whereas the retrievals are then used at eight different stations. It is not clear if the results obtained at the single station, from which a single ML algorithm was selected, also apply to the other stations, and why. As such, I suggest the authors for a thorough revision of their work, including but not limited to the specific comments available in the following.

### **Specific comments**

Line 9: What do you mean by "goal of carbon emission peak"? Revise.

Lines 10-13: The reader may not know what is the "traditional power law method" (and therefore in which sense it relies on the constant coefficient to estimate the high-altitude wind speed) and the variety of factors on which it depends. You should explain better which are those factors or at least some of them.

Line 17: Add "of" before "results". I would also add "with the observations" before "show". Change "show" to "shows".

Lines 18-20: Rephrase: "Based on the WS120 from the RF model, the diurnal variations of WS120 and of wind power density (WPD) were then estimated."

Line 23: Change "by" to "based on".

Line 25: This is not the unit for wind speed (could be so for WPD).

Lines 23-26: But for the reader these names are not meaningful. Please indicate the characteristics of the cities (e.g., geographical location).

Line 28: Change "into" to "for".

Lines 33-34: Well, not only carbon dioxide!

Lines 34-35: The link with the previous sentence is missing; and also it is not clear what you mean by "depletion of fossil energy".

Line 39: Delete "had".

Lines 39-40: It would be interesting to know a percentage contribution of this source to the total capacity.

Lines 41-42: The link with the previous sentence is not that clear.

Line 54: Perhaps you mean that China is currently facing an increasingly serious energy and climate situation?

Line 56: Again, I cannot understand what is this "carbon emission peak" strategy.

Line 58: Change "has been flourished" to "is flourishing":

Lines 59-64: This is not linked with the previous sentence. A break is needed.

Lines 65-66: This is not true as the boundary layer height varies with day, land use, meteorological conditions, and so on. Please revise.

Lines 68-70: You are making a jump to this theory, without introducing the reader to it and to its meaning.

Lines 70-72: Quite generalistic and without details.

Line 73: Change "was" to "is".

Line 101: Change "developed" to "increased".

Lines 100-107 and 109-115 and 117-120 and 122-132: Are these data available somewhere?

Lines 109-115: So are the synoptic (or I assume so based on the statement in the abstract, not repeated here) stations set in the same place as those from the RWP network? This is not clear. The instruments and the data used should be better described.

Lines 122-124: This seems more like an advertisement rather than a description. Please provide more details.

Line 126: Which surface parameters?

Line 126: The spatial resolution is very rough compared to the granularity of point observations.

Line 130: How did you obtain data at eight stations from gridded data?

Lines 127-130: It is not clear how these parameters affect wind speed. Explanations are needed here or somewhere in the manuscript.

Lines 134-136: Rephrase: "In this section, we introduce firstly the classical PLM method to retrieve the WS120 based on 10-m wind speed measurement. Then, we describe the three ML algorithms used to retrieve WS120. We finally present the method for evaluating wind energy."

Line 138: Change "assumed" to "assumes".

Line 139: Change "has been" to "is".

Line 140: Change "formulae" to "formula".

Lines 143-144: And what is the value for non coastal locations?

Line 153: Add "presented" before "as follows".

Lines 155-197: To be honest, if one is not expert in those techniques, the explanation is not well understandable. Figures for this Section could be moved to the Appendix, but the explanation of the methods must be rewritten.

Lines 236-237 and 237-238: Not clear: revise.

Line 241: Change "significant improvement" to "significantly improved".

Line 243: Change "duo" to "due" and "to it considers" to "to the fact that it considers".

Line 250: Change "By contrast" to "Conversely".

Lines 248-257: The explanation is not clear: revise. Also, I don't understand the need to discuss the difference (a sort of mean bias) when you were discussing the RMSE and R values. Also, it would be needed to understand if the fitting and comparison of model estimations with observations vary with hour of the day, season, or other factors. Also, the discussion could be improved because for instance from Figure 5 I can observe that: RF model is the best but tends to overestimate small values and underestimate high values; similar discussions also for the other models.

Lines 259-260: If it is obvious, why do you need to discuss it?

Lines 260-263: Can you explain the reason of those seasonal variabilities?

Lines 270-271: The mechanism is much more complicated and also variable because of the presence of complex terrain, buildings, sea-land interfaces.

Lines 292-294: Isn't this obvious?

Line 293: Change "On the whole" to "Overall".

Line 313: Change "maximums" to "maximum".

Table 1: Are you sure about the unit for Altitude (km)? Some altitude values are really high.

Table 2: This Table is not needed, this explanation can be given in the main text.

Figure 4: Please discuss somewhere the importance of the parameters and the reason.

Figure 5: What is the color of the scatterplot representing?

Figure 7: The plot is quite strange and is not well presented in the main text and in the caption.

Figure 8: The probability distribution look quite far from the fitted distributions. Please discuss.

Figure 9 and 10: Please adjust the scales for the y-axes.