

## Review of egusphere-2023-2727: “Extending wind profile beyond the surface layer by combining physical and machine learning approaches” by Liu et al.

### General comment

This study introduces a Phy-RF method to extend wind profiles beyond the surface layer, overcoming limitations of the traditional model based on the Monin–Obukhov similarity theory. By combining the power law model (PLM) with the random forest (RF) algorithm, the Phy-RF method addresses errors in the PLM above the surface layer attributed to the  $\alpha$  setting. Comparing performance over China, the Phy-RF model outperforms PLM and RF, demonstrating better accuracy and stability. Temporally, it is not significantly affected by seasonal variations but shows limitations during specific time periods. Spatially, the model performs worse in highland areas due to the absence of consideration for terrain factors.

After some minor revisions, I am in favor, that this paper gets published in ACP. The text in general should be carefully checked during the English language copy-editing process.

### Specific comments

1. The acronym Phy-RF confuses me a bit, because you state in line 11-12: “...we propose a novel method that combines the power law method (PLM) with the random forest (RF) algorithm to extend wind profiles beyond the surface layer, called the Phy-RF method.”

Why you do not use PLM-RF as acronym if it is based on PLM. RF-PHY acronym (Radio Frequency Physical Layer) is also used in the wireless communication sector, and you may want to separate the name of your new method better.

2. In the introduction, I find the absence of a concise overview and reference of the Prandtl-layer, which encompasses the initial tens of meters within the atmospheric boundary layer.
3. The Root Mean Squared Error (RMSE) quantifies the accuracy of a regression model in predicting the response variable's value in absolute terms, whereas R-Squared measures how effectively the predictor variables account for the variability in the response variable.

I encourage the authors to also have a look and include R-Squared or Adjusted R-Squared metrics in their model evaluation. If the authors stick with RMSE, I think they must better justify their decision.

4. Just a curiosity, now you focused on different land-cover types, can you make a statement about the performance of the Phy-RF model above water surfaces yet?

The emphasis was on comparing the performance over China. Do you plan to investigate a more global performance estimate of the models in the future?

## Technical corrections

1. Lines 25-26: "These findings have great implications for the weather, climate and renewable energy." The noun of the sentence is missing. Maybe add in the end of the sentence the word "sector" or "research."
2. Lines 37-38: "..., in which can be assimilate into atmospheric models to produce global wind profile products." This sounds a bit wrong. I suggest writing: "Satellite observations, such as those from Aeolus, can provide horizontal line-of-sight wind profile data that can be assimilated into atmospheric models to generate global wind profile products."
3. Line 41: I suggest to better formulate the following part: "...ground-based observations like wind tower, wind profile radar, and wind profile lidar..." into "...ground-based wind measurements from towers, radar or lidar-based profilers...".
4. Line 90ff: I recommend not to describe the color bar in the text here. Just refer to the Figure 1 and maybe specify the land cover types in the caption of this Figure, if needed.
5. Line 157: Here you talk about previous studies, but no references are given. Either add the references again or reformulate.
6. Delete full stop in title of Section 4.1.
7. Wording in title of Section 4.2 wrong. I guess should be "Wind speed evaluation of the Phy-RF model".
8. Paragraph 346-349: First sentence misses something g in the end like e.g. "sector": "...implications for the weather, climate and renewable energy sector.". Please also reformulate the second part. What is meant by "limitations of data time"? This is not clear to me.
9. Figure 3: I suggest adding the units in the y-axis's captions. The scatter points, for my impression overlap to strong here and this could lead to misinterpretations by the reader.
10. Figure 4: The figure does look too pixelated, please increase the resolution of this figure. In 4(c) correct "Wooldland" to "Woodland". The typo ("wooldland") is also in the caption of the figure.
11. Figure 5: Here please write in the caption the meaning of all variable acronyms.