Response to Reviewer #1's Comments

The paper introduces the Phy-RF model, an innovative approach combining physical wind profile models with machine learning to extend wind profile estimations beyond the surface layer. Through comprehensive analysis and validation, the study demonstrates that the Phy-RF model offers a more accurate estimation of wind speeds at 100 meters, outperforming both the traditional power law method (PLM) and random forest (RF) machine learning algorithm alone. In general, this paper is well written. I recommend the publication of this manuscript, while I have some comments below for the authors to address.

Response: We greatly appreciated the reviewer's comments on our manuscript, which greatly improve the quality of our manuscript. We have made efforts to adequately address the reviewers' concern one by one. For clarity purpose, here we have listed the reviewer' comments in plain font, followed by our response in bold italics.

1. The paper compares the seasonal mean of ERA-5 and Phy-RF. It could also benefit from a direct comparison between the Phy-RF model outputs and ERA-5 wind speed data at 100 meters using scatter plots. Since the model uses ERA-5 as the input, such a comparison would be beneficial for evaluating the Phy-RF model's performance.

Response: Per your kind suggestion, we made a direct comparison between the PLM-RF model outputs and ERA-5 wind speed data at 100 meters using scatter plots, as shown in the Fig. S6. In addition, the acronym "Phy-RF" is modified to "PLM-RF" based on the reviewer #2's comment.

As a result, the following sentences has been added to the end of the first paragraph of Section 4.2 "Wind speed evaluation of the PLM-RF model" in the revised manuscript:

"In addition, the comparisons between the WS100 from ERA5 and from PLM-RF model for different periods are shown in Fig. S6. Although the output of the PLM-RF model has a good correlation with the WS_{100} from ERA5, there exist still some differences. Most of the WS_{100} from the PLM-RF model are greater than that of

ERA5 when the wind speed is high. This is because the Δα is introduced in the PLM-RF model, which makes the model tend to produce large output values.”

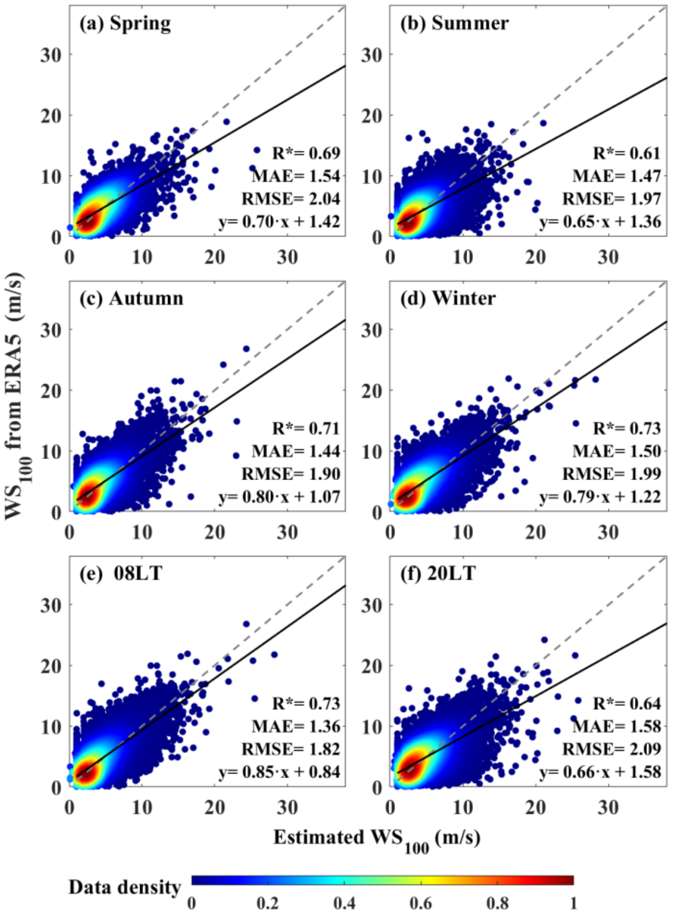

Fig. S6. Comparisons between the WS_{100} from ERA5 and from PLM-RF model in (a) spring, (b) summer, (c) autumn, (d) winter, (e) 0800 LT, and (f) 2000 LT.

2. While the study incorporates data from several ARM sites, there appears to be insufficient evaluation using these datasets. Since averaged profiles may not capture the full extent of discrepancies between the Phy-RF model estimates and observations, relying on mean profiles from lidar cannot accurately represent accuracy.

Response: Good points! To better reflect the performance of Phy-RF model, we investigated the diurnal variations of R^2, MAE and RMSE between the WS_{100} calculated by PLM-RF model and the WS_{100} observed by Doppler wind lidars at three ARM sites, as shown in Figure 13.

Accordingly, the accompanied descriptions have been supplemented in the revised manuscript. "To further evaluate the performance of the PLM-RF model, the diurnal variations of R^2, MAE and RMSE between the WS_{100} calculated by PLM-RF model and the WS_{100} observed by Lidar are investigated in Figure 13."

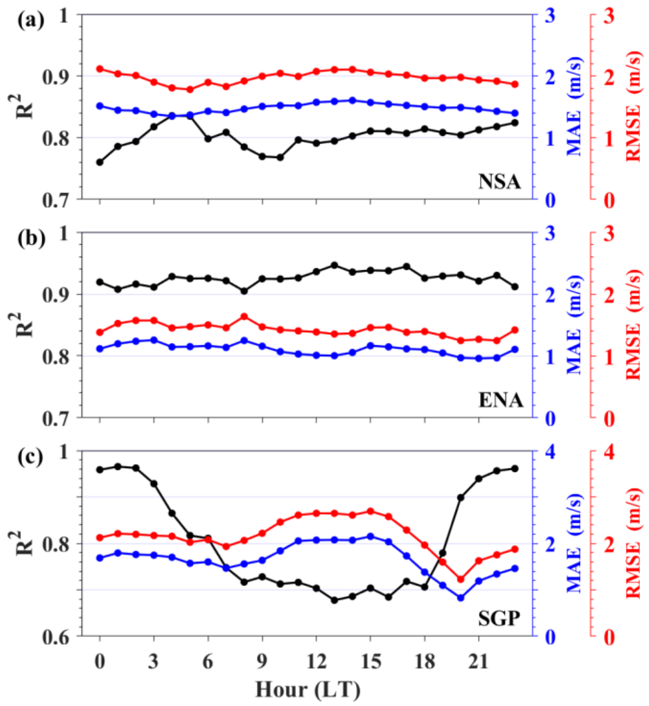

Figure 13. Diurnal variations of R^2, MAE and RMSE between the WS_{100} calculated by PLM-RF model and the WS_{100} observed by Lidar at (a) NSA, (b) ENA, and (c) SGP sites. The black, blue and red lines represent the R^2, MAE and RMSE, respectively.

3. More detailed statistical analysis, such as examining the mean absolute error (MAE), could enhance our understanding of the model's performance. I suggest including the MAE to evaluate the differences between Phy-RF, ERA-5, and Lidar.

Response: Good suggestion! In the revised manuscript, the statistical parameters such as coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE) are used to evaluate the differences between PLM-RF, ERA-5, and Lidar. The modifications can be seen in Fig.7, Fig.8, Fig.13 and Fig. S6-S8.

4. I noticed there are notable discrepancies in the mean profiles of Phy-RF and Lidar over the SGP. What factors lead to such biases? Meanwhile, Are such biases also included in the ERA-5?

Response: From the point of our view, the factors that may lead to such noticeable biases (at the very least) are as follows:

(1) The generalization of the PLM-RF algorithm depends on the training and test samples. However, the training and test samples of the PLM-RF model were obtained from soundings at 0800 and 2000 LT, which do not actually contain any in situ measurements from the period 1100 to 1500 LT. This means that the PLM-RF model has no generalization at noon, resulting in poor accuracy of the PLM-RF model during 1100 to 1500 LT.

(2) The SGP site are located over land, with significant diurnal variations in wind speed compared to NSA and ENA sties. Due to the lack of observational constraints from 1100 to 1500 LT, the low performance of the PLM-RF model is evident during the daytime at SGP sites.

These two factors lead to the noticeable discrepancies (at 1400 LT) in the mean profiles of Phy-RF and Lidar over the SGP site. Due to the biases caused by the model itself, it also occurred in the comparison with ERA-5.

Therefore, we added one sentence in the conclusions part: "Therefore, it is not recommended to use the PLM-RF model for the time period 1100 to 1500 LT over highland areas before including observation data to constrain the model."

5. The paper may include its discussion on the limitations of the Phy-RF model, particularly how it performs under extreme events.

Response: Per your kind suggestion, we discussed the limitations of the PLM-RF model in the conclusions part of this revised manuscript, which is shown as follows:

"The limitation of the PLM-RF model is that the performance of PLM-RF model is affected by diurnal variation and terrain. The generalization of the RF model depends on whether the training samples contain sufficient sample inputs. The training samples of the PLM-RF model do not contain in situ measurements from

the time period 1100 to 1500 LT, resulting in relatively poor accuracy during this period. Similarly, the RMSE of the wind profiles is relatively larger at highland areas, which is likely due to the fact that the influence of terrain was not considered in the construction of the PLM-RF model. Therefore, it is not recommended to use the PLM-RF model for the period from 1100 to 1500 LT over highland areas before adding observation data to retrain the model."

In addition, the Fig. 7 shows that the PLM-RF model has better accuracy and stability compared to PLM and RF. Especially under high wind speed events, the output of PLM is significantly low, while the PLM-RF model can effectively correct this underestimation. The modifications in the revised manuscript are as follows:

"Overall, the advantage of the PLM-RF model is that it can provide more accurate wind profiles than the PLM, especially when the actual wind speed is high."

6. Despite a worse performance compared to Phy-RF, the traditional PLM also seems good (Figure 7). It would be beneficial to discuss more clearly the potential applications and advantages of the Phy-RF approach, particularly in scenarios where the PLM and ERA-5 may fall short.

Response: Points have been well taken. Per your suggestion, we discussed the potential applications and advantages of the PLM-RF model in the conclusions part, which is shown as follows:

"Overall, the advantage of the PLM-RF model is that it can provide more accurate wind profiles than the PLM, especially when the actual wind speed is high. Moreover, the PLM-RF model is not affected by seasonal variation. This is because the RF model is data driven. The training sample of the PLM-RF model contains enough samples from four seasons. The PLM-RF model is recommended for areas with high wind speeds, such as coastal areas."