Response to reviewer comments

Dear Editor and Reviewers,

We are very grateful for your time and valuable comments, which we found very helpful. We have addressed questions and comments raised by the reviewer in the revised manuscript with tracked changes. Please find our point-by-point response (in blue font) to the comments below. We hope our revisions have properly addressed your concerns.

Thanks again for your time.

Sincerely,
The authors

Reviewer 2

The revised version of Zhou et al. addresses many of the reviewer concerns and is better presented than the original version. However, I still have major concerns about the manuscript as presented.

Response: Thank you very much for your recognition and encouragement of our work, we have taken your comments into consideration and further edited our manuscript.

General comments

1. The first major concern is of the limited timespan of the data sets used in training and testing. I understand that the authors explained that the dataset contains many hours of data over 410 weather stations, but finer sampling of the same period cannot fully compensate for the fact that the data is constrained to one particular season (December - February) of a particular year. I would suggest to at least conduct the testing (similar to what's done in Figure 13) in other seasons to extend to another winter period.

Response: Thank you very much for your advice. We have extended the testing to winter of 2022, and the evaluation results of the testing are shown in Fig. 13.

We have also added the following text on line 254: "*While similar past studies for wind speed correction from NWP models usually use several years for training and at least one year for testing whereas our periods are shorter, the size of our data set is sufficient. For example, Sun et al. 2019 used a data set that contained 1827 days, from January 2012 to December 2016, using 143 grid points with a resolution of 0.5°*0.5° predicted by ECMWF, followed by 24 features for each sample, with a training set size of 1827*143*24 for each prediction time. Meanwhile, the size of our training set is about 2160*410*12. Therefore, even though it only took us a month to train, for this project, we trained millions of data; Second, the training data we used here was obtained through daily operational runs of numerical weather forecasting,* *so it would have taken several years to get an equal amount of training data*."

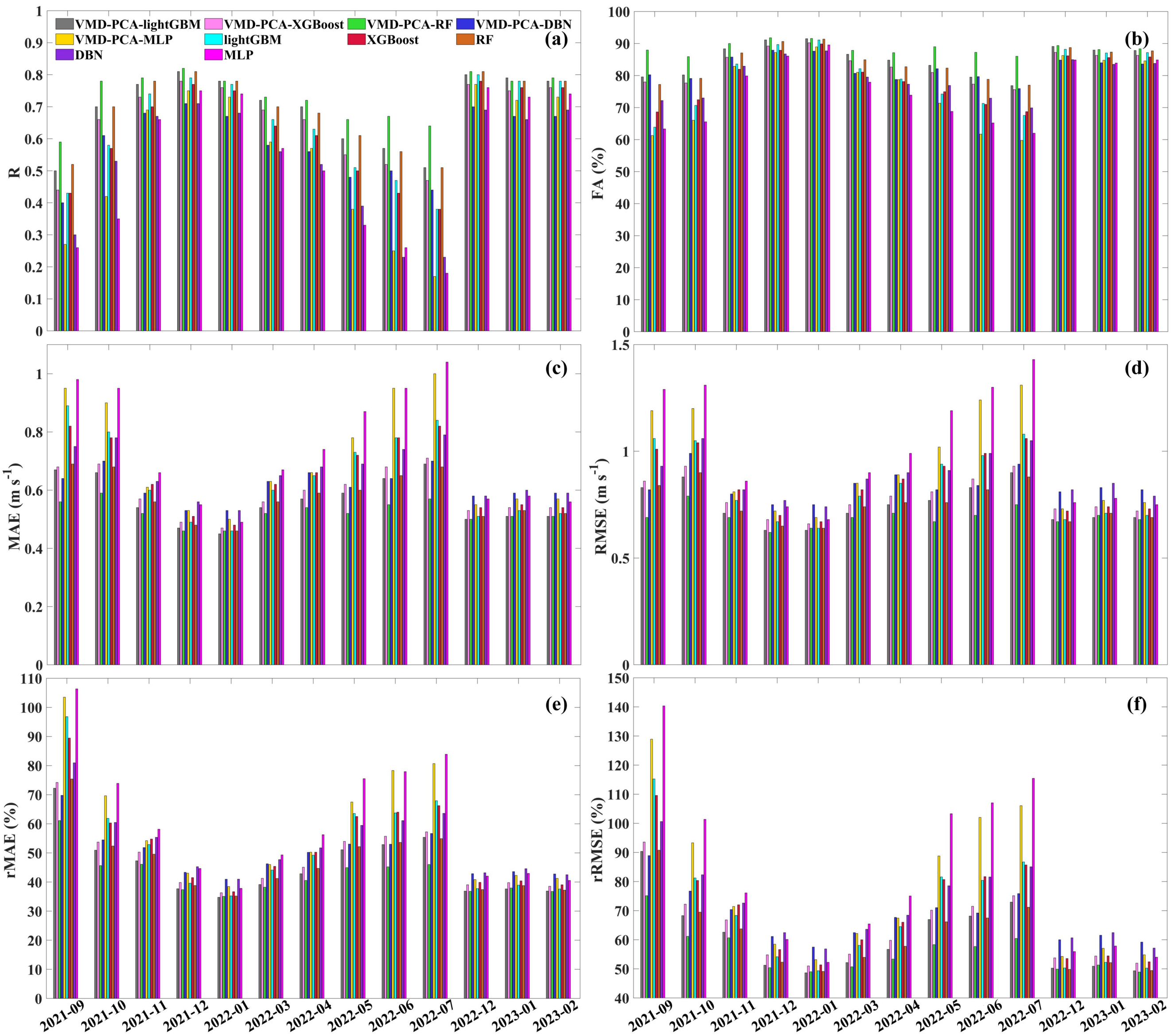

Figure 13: Evaluation histograms of 10-meter wind speed predicted by 10 models in different months in Experiment 1 and Experiment 2 ((a), (b), (c), (d), (e), and (f) represent R, FA (%), MAE (m s^{-1}), RMSE (m s^{-1}), rMAE (%), and rRMSE (%) respectively).

2. Regarding the worse performance of the model in different seasons as the training - this is to be expected, but do the authors anticipate that training over an entire year's data would improve the situation or distinct models would have to be built for each season in training?

Response: Thank you very much for pointing this out.

As we can see in the 13 months of testing from Fig. 13: VMD-PCA-RF evaluation indices are relatively stable across the 13 months, with a correlation coefficient R above 0.6, FA above 85 %, MAE below 0.6 m s^{-1}, RMSE below 0.8 m s^{-1}, rMAE below 60 %, and rRMSE below 75 %. The robustness of the other models was not stable during the 13-month test period. We try to get a best full-year wind speed correction model based on the training data specification reaching millions of levels. We recommend training one month for each season to build the model for two reasons: Firstly, the training data will be light and the training time will be shorter; secondly, the wind speed characteristics of each season are relatively strong, which contains time information, which is often not captured by machine learning models trained directly on a year's training data.

We have also added the following text on line 528: "*In cases where ample machine CPU and other hardware resources, as well as training time, are available, we recommend using VMD-PCA-lightGBM for modeling each season. However, when dealing with limited resources such as a laptop and constrained training time, we recommend using VMD-PCA-RF to train data for a single month, as this yields the most robust correction results.*"

3. The second major concern is regarding the presentation of the manuscript. The manuscript is 39 pages long but I personally would suggest both shortening existing sections as there are a lot of information that is not of general reader interest - e.g., Table 2 with the best hyper-parameters of the models would vary for future readers who would like to adopt a similar methodology. A lot of numbers are presented in the text, but some tables and figures present the same data, for example Table 3 & 4, and the Taylor diagrams in Figure 8(c), (f), and many of the same numbers are repeated in text. I would move the Table to Appendix as the Taylor diagram would visually show the information better.

Response: Thank you very much for pointing this out.

The manuscript is on the longer side, and after carefully reviewing and considering your suggestions, we think some details are essential for the reproduction of this work. For instance, for the hyperparameter setting of the model that you mentioned, the Bayesian tuning of the model parameters is time consuming, the listed parameters can be directly used in the parameter settings, reducing the time to re-select parameters. We also provide parameter settings with a set range reference for wind speed training data for other areas.

Nevertheless, we agree that some information in Tables 2 to 4 have been repeated in text, and therefore, we have taken your suggestion and moved Tables 2, 3, and 4 into the supplemental material.

4. Many of the figures and tables are also replicated for the December 2021 and January 2022 data, but many times they do not differ significantly. If there is no fundamental difference between the December and January data's characteristics, I suggest keeping only one set of Figures/Tables in text and refer to the other in the Appendix. The main takeaway point of the paper, in my view, is to understand the characteristics of each model in different situations (as presented by the manuscript - height intervals, different wind speeds, location, etc.) and the difference between Dec and Jan are not stressed. In any case, Figure 13 already shows many further months in a concise manner.

Response: Thank you very much for pointing this out.

We sincerely appreciate your valuable suggestions. Your point regarding no significant difference between the December and January data was especially true for Figures 9 and 10, and we agree it would be good to move one month (January 2022) to the supplemental material and keep just December 2021 to make our point. Thank you for helping us make the manuscript more concise.

At the same time, we have reviewed all other figures as well, and we found differences between the two months in other Figures/context to be significant enough to keep them both in the main body of the text. For instance, we can see from the scatter points in Figs. 6c, d, e, and f that the actual maximum wind speed in December 2021 is significantly greater than that in January 2022. It can also be seen from Fig. 12 that when WRF forecasts wind speed greater than 12 m s^{-1}, the correction error of various models is significantly lower in January 2022 than in December 2021. So we decided to leave Fig. 6 and Fig. 12 in the paper for a comparison.

We have also mentioned in the text the differences between the WRF forecast for January and December 2021: "*Moreover, the actual average wind speed in January 2022 deviates from the range of one standard deviation of the WRF forecast wind speed at 17:00 and 18:00 (Fig. S9a). This demonstrates that the wind speed forecast by WRF is inaccurate and exhibits substantial diurnal variation errors.*"

Additionally, we have mentioned in line 424 the differences between the WRF forecast for January and December 2021: "*In the regions of Hainan, Guangxi, and Guangdong, the number of sites with a RMSE for 10-meter wind speed forecast in December 2021 ranging from 5.6 to 6.0 m s^{-1} was significantly higher than in January 2022, especially in coastal areas (Fig. 10b, Fig. S10b).*"

5. As a paper in GMD, a potential reader looks to this paper for methodology and evaluation of using a VMD-PCA-RF method over a traditional numerical model approach. My suggestion is to focus on the major questions about methodology, model choice and comparison across machine learning models and with the WRF numerical model, and keep the paper concise.

Response: Thank you very much for pointing this out.

The purpose of our paper is to evaluate the pros and cons of post-processing methods for correcting forecast wind speed after the establishment of numerical weather prediction (NWP) models. Hopefully, potential readers will be inclined to choose VMD-PCA-RF method over a traditional numerical model approach (traditional error correction method) after reading the whole article, since only one month's data is trained to obtain the highest correction accuracy for the whole year.

We re-evaluated the manuscript based on your suggestion and found it difficult to reduce the length of the paper without leaving out what we think are important part of the story:

In the introduction of the paper, we dedicated a considerable amount of space to provide background information on the limited usage of BOA-VMD in traditional error correction methods by previous studies. Furthermore, the discussion regarding the effects of BOA-VMD-PCA and RF feature importance is also crucial.

As you agree, the methodology and model differences are important, so we introduced various methods in Section 2.2, including WRF numerical simulations, Variational Mode Decomposition (VMD), and Principal Component Analysis (PCA). We particularly focus on presenting the three key steps of the two sets of experiments required to construct the hybrid forecasting approach throughout the entire paper. We mention 10 models in the two sets of experiments. In brief, these 10 models can be divided into 5 base models (traditional error correction method). These 5 models represent different approaches, including regression models with simple neural network architecture (MLP), machine learning tree regression models (RF, XGBoost, LightGBM), and deep learning regression models (DBN). In experiment 2, we added VMD-PCA method to further excavate hidden features of wind speed data, and obtained five new VMD-PCA-models. Sections 3.1 to 3.4 compare the performance and spatiotemporal variations of these 10 models for December 2021 and January 2022. Additionally, Figure 13 displays six evaluation metrics over a 13-month period, concluding the best stability for the proposed model (VMD-PCA-RF).

The extra information we delve into are the importance of the effects of the BOA-VMD-PCA and RF features. Before introducing the best model, we extensively discuss the positive contributions of BOA-VMD-PCA to the calibration results, providing readers with an understanding of why we chose BOA-VMD-PCA over other model optimization methods (such as grid search) and wind speed feature extraction methods (like Empirical Mode Decomposition). A discussion on the effects of BOA-VMD-PCA is necessary. This discussion allows readers to understand that, "*despite the complexity of input factors being one source of uncertainty in the WRF prediction correction process, the 10-meter wind speed components introduced by VMD-PCA contribute positively to the correction results*." Therefore, it also explains that the correction effect of the VMD-PCA-models proposed in Experiment 2 is superior to the baseline model in Experiment 1.

Furthermore, we emphasize the importance of WRF forecast wind speed and site altitude in model correction, so that readers could hopefully agree that these two feature factors are highly significant in model training. Discussing RF feature importance is necessary because it would hopefully convince the readers that, "*as the 10-meter wind speed forecast by WRF increases, the instability of the 10-meter wind speed corrected by the 10 machine learning models gradually increased, and the correction accuracy gradually decreased (Fig. 12). This partly explains the higher importance of the 10-meter wind speed forecast by WRF. Moreover, when the height of the station increases or decreases at 1 km, the correction intensity tends to increase gradually. This further explains the higher importance of the height factor in the RF model training.*" Therefore, it also explains why in Experiment 2, we continue to retain two feature factors: WRF forecasted wind speed and station altitude, and further explore hidden information within the WRF forecasted wind speed.

All in all, these methods, model selection, and comparisons aid readers in reproducing our model training and evaluation process, which is crucial. However, it's also essential to help readers understand the proposed VMD-PCA-RF model's high robustness and to discuss the effectiveness of BOA-VMD-PCA and the importance of RF features. These discussions would be very helpful to the readers' understanding of the VMD-PCA-RF model.

6. A more minor concern is regarding language which is at times is unclear and probably could be shortened. For example, L228 "the first path involves increasing the meteorological variables possibly related to wind speed in the forecast field". I believe with "increasing" the authors mean that Experiment 1's purpose is to identify the most important predictors. For experiment 2 the authors write (L231) "...using the VMD-PCA algorithm to dig out the hidden wind speed characteristics of the 10-meter forecast wind speed, reduce the input of other meteorological factors such as WD_{10} and D_{2}, ..." which I believe is using VMD-PCA to process the input data into principal components. I feel Figure 2 can probably be simplified (with less text - which is also very small) to better explain the process.

Response: Thank you very much for pointing this out.

We have reduced the texts in Fig. 2 to make the messages more clear.

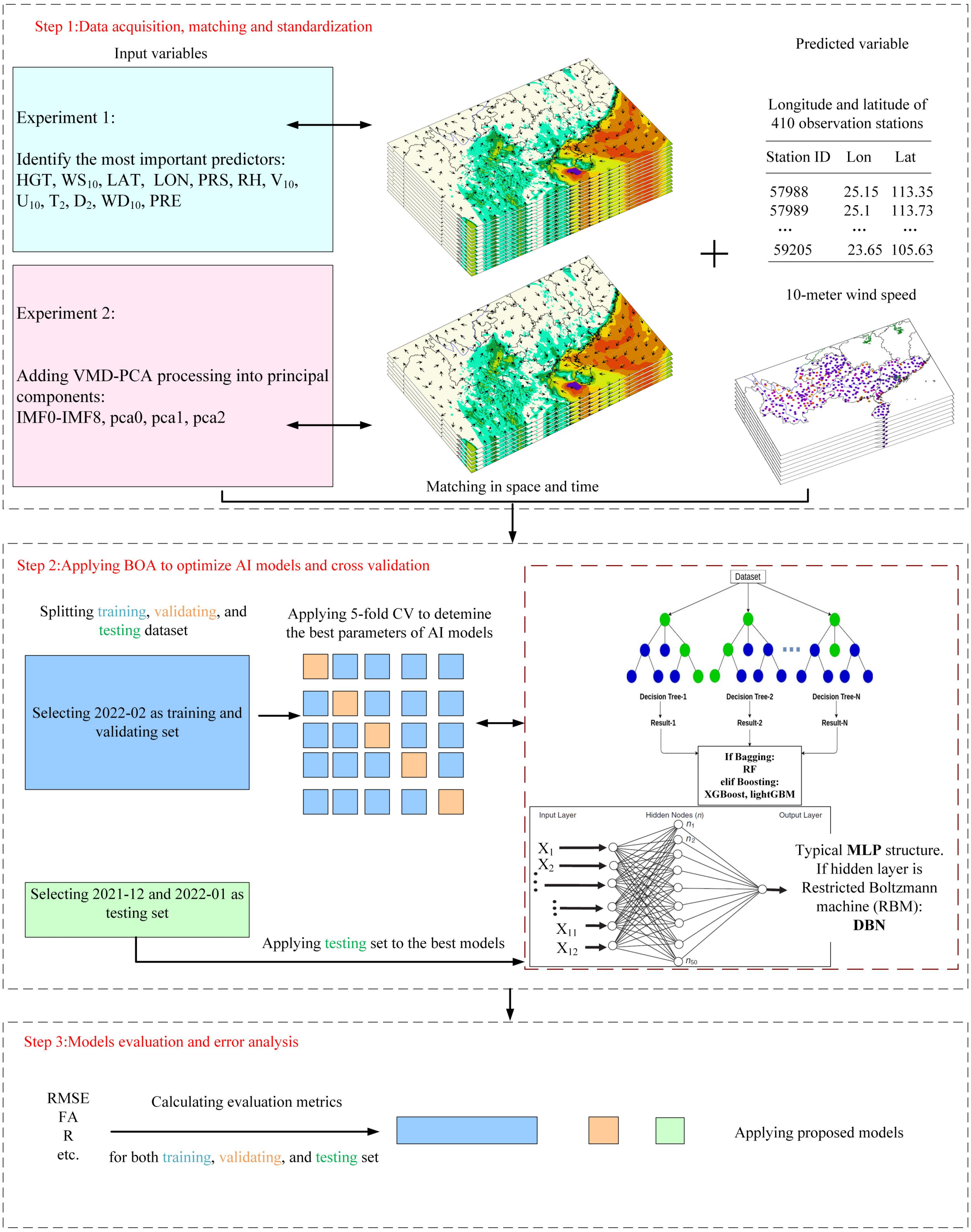

Figure 2. Flowchart of the AI model used to correct WRF-predicted wind speeds in the two main experimental pathways.

Specific comments:

There are a lot of potential points for improvement in the paper in the major comments above. I point out specific comments mainly on the figures.

1. L156: "the GFS data updates at 06:00 UTC ..." - I'm not sure what the authors mean here. Are the GFS boundary data only input into the WRF model once per day? That is an unusual configuration, as usually GFS data is used in 6-hourly or 3-hourly

intervals. Or do you mean the GFS data is updated on the website? In which case, such information is not very relevant and I would suggest removing it to avoid confusion.

Response: Thank you very much for pointing this out.

As we know, the GFS data is released daily at 00:00 UTC, 06:00 UTC, 12:00 UTC, 18:00 UTC with updated forecast data for the next 384 hours. We are using a 90-hour forecast every 3 hours, released at 06:00 UTC.

We have updated the text in L156: "We *use the WRF model in combination with daily GFS data resolution of 0.25° × 0.25°. The GFS data used by us is released at 06:00 UTC with forecasting every 3 hours for a total duration of 90 hours.*"

2. Figure 6, 7, 9: Is "Hour" local time or UTC?

Response: Thank you very much for pointing this out.

We have changed the titles of Figs. 6, 7, and 9 to:

"*Figure 6: The scatter density map compared with the actual 10-meter wind speed: (a) 10-fold cross-validation training set of VMD-PCA-RF model in February 2022, (b) 10-fold cross-validation validation set of VMD-PCA-RF model in February 2022. The 24-hour scatter map compared with the actual 10-meter wind speed: (c) WRF forecasts in December 2021, (d) VMD-PCA-RF model forecasts in December 2021, (e) WRF forecasts in January 2022, and (f) VMD-PCA-RF model forecasts in January 2022. (The time is UTC + 08:00.)*"

"*Figure 7: The scatter density map compared with the actual 10-meter wind speed: (a) 10-fold cross-validation training set of VMD-PCA-lightGBM model in February 2022, (b) 10-fold cross-validation validation set of VMD-PCA-lightGBM model in February 2022. The 24-hour scatter map compared with the actual 10-meter wind speed: (c) WRF forecasts in December 2021, (d) VMD-PCA-lightGBM model forecasts in December 2021, (e) WRF forecasts in January 2022, and (f) VMD-PCA-lightGBM model forecasts in January 2022. (The time is UTC + 08:00.)*"

"*Figure 9: VMD-PCA-lightGBM, VMD-PCA-RF, and WRF daily variation of predicted and actual wind speeds in December 2021 and January 2022. (The shading areas represent an interval of 1 standard deviation, which is a 68% confidence interval. The time is UTC + 08:00.)*"

3. Figure 9, 10, 11: Text is too small and really unreadable. Please take care of the figure presentation.

Response: Thank you very much for pointing this out.

We have increased the text size in the Figs.9, 10, and 11.

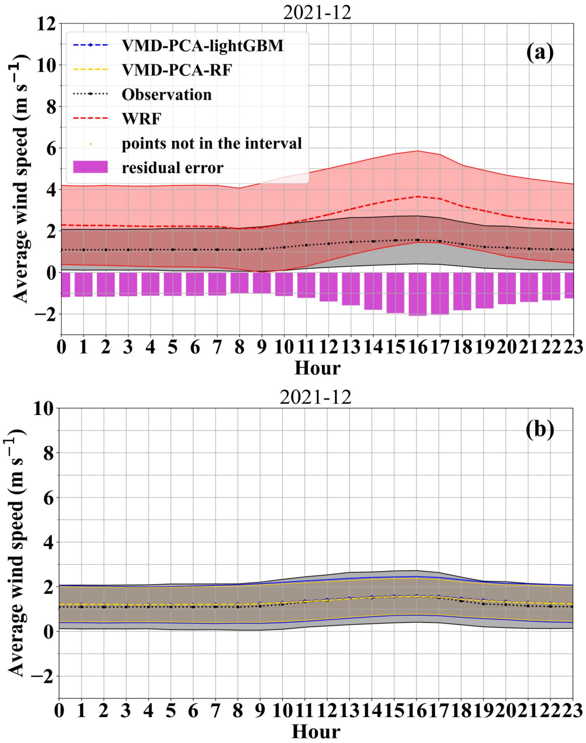

Figure 9. VMD-PCA-lightGBM, VMD-PCA-RF, and WRF daily variation of predicted and actual wind speeds in December 2021. (The shading areas represent an interval of 1 standard deviation, which is a 68% confidence interval. The time is UTC + 08:00.)

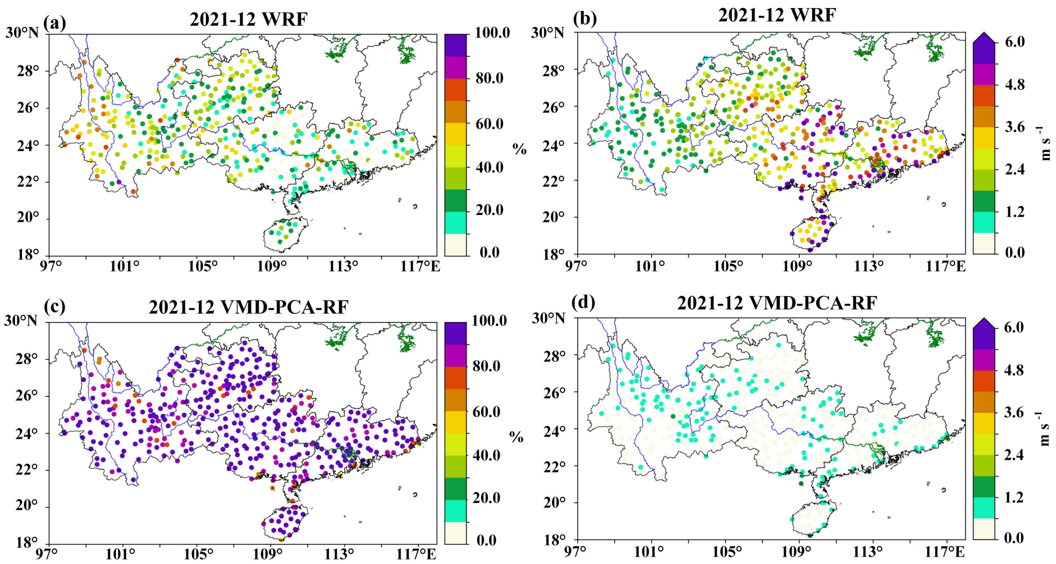

Figure 10. FA ((a), (c)) and RMSE ((b), (d)) distribution maps of VMD-PCA-RF and WRF models on 410 sites in five southern provinces in December 2021.

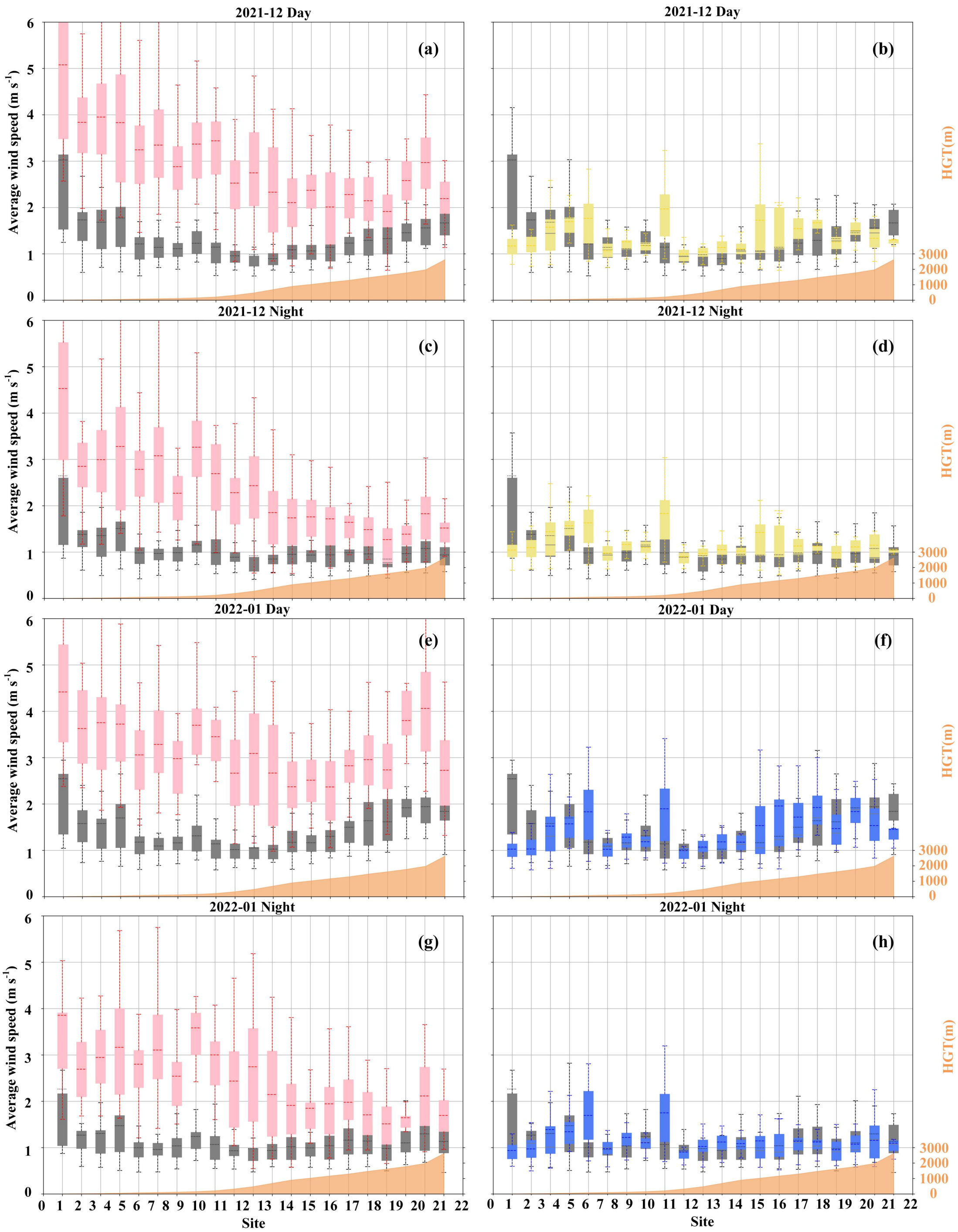

Figure 11. The boxplots of the predicted wind speeds of the VMD-PCA-RF (yellow), VMD-PCA-lightGBM (blue), and WRF (pink) models at 20 stations at different height intervals, and the boxplots of the actual wind speeds (gray).

4. Figure 13: Text is small. There are a lot of models in here and it's hard to plot all this in one plot clearly, but perhaps authors could experiment with only using markers or separating the individual bars. As it stands it is very hard to read any specific value for a particular model. Maybe show the top few models only.

Response: Thank you very much for pointing this out.

We have increased the text size in the Fig. 13.

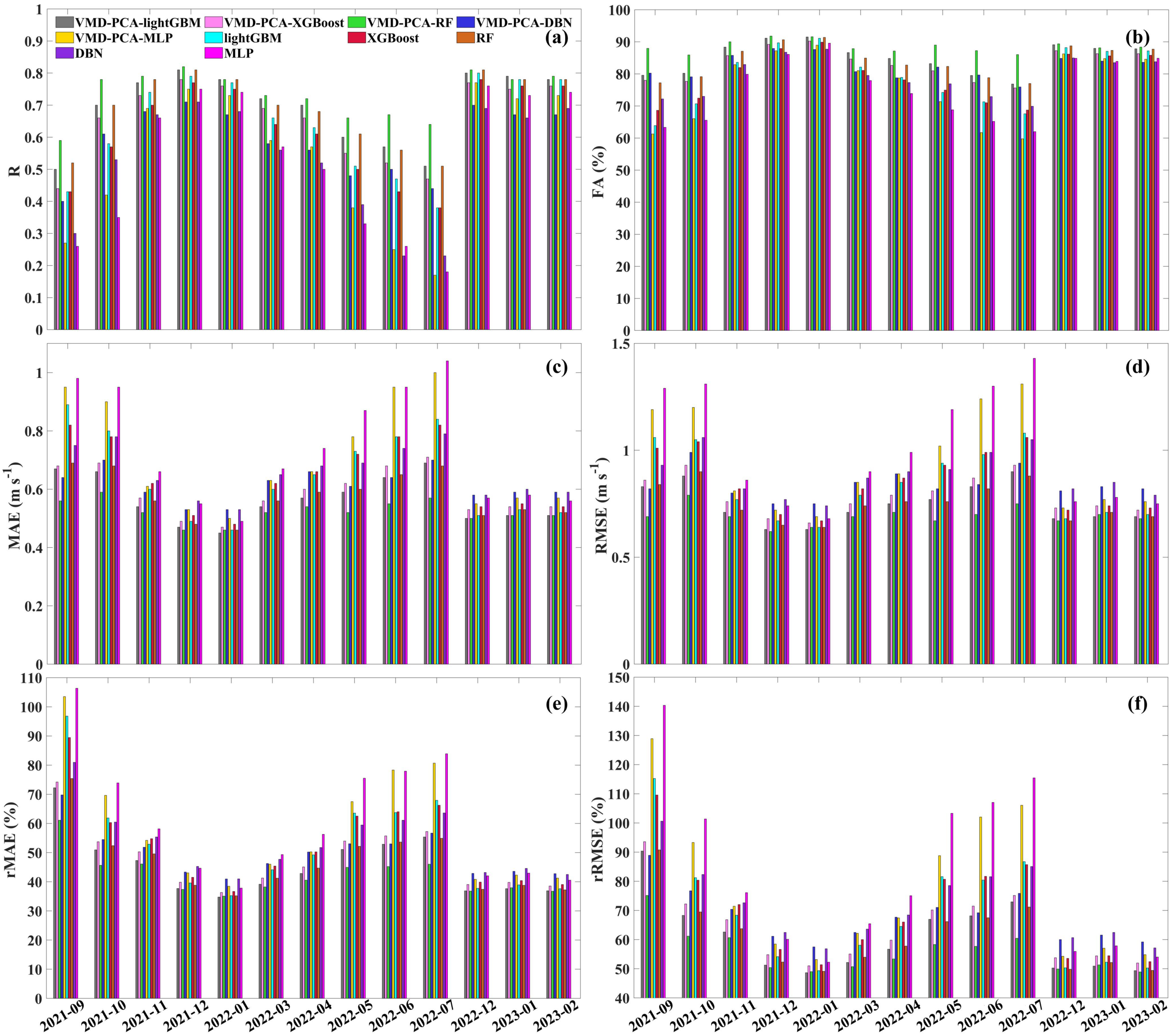

Figure 13. Evaluation histograms of 10-meter wind speed predicted by 10 models in different months in Experiment 1 and Experiment 2 ((a), (b), (c), (d), (e), and (f) represent R, FA (%), MAE ($m\ s^{-1}$), RMSE ($m\ s^{-1}$), rMAE (%), and rRMSE (%) respectively).

Reviewer 3

1. In the abstract, it still says "Global Prediction System (GFS)" which is not correct. The acronym is Global Forecast System.

Response: Thank you very much for pointing this out.

We have updated line 24-25 to the following text: "*We first construct WRF-predicted wind speeds using the Global Forecast System (GFS) model output based on* *prediction results.*"

2. Figure 1: It's pretty hard to see the provincial boundaries. Partly because of the black dots, and partly because of the dark blue background. Maybe they could be made thicker to better highlight the regions.

Response: Thank you very much for pointing this out. We have thickened the provincial boundaries to highlight the regions.

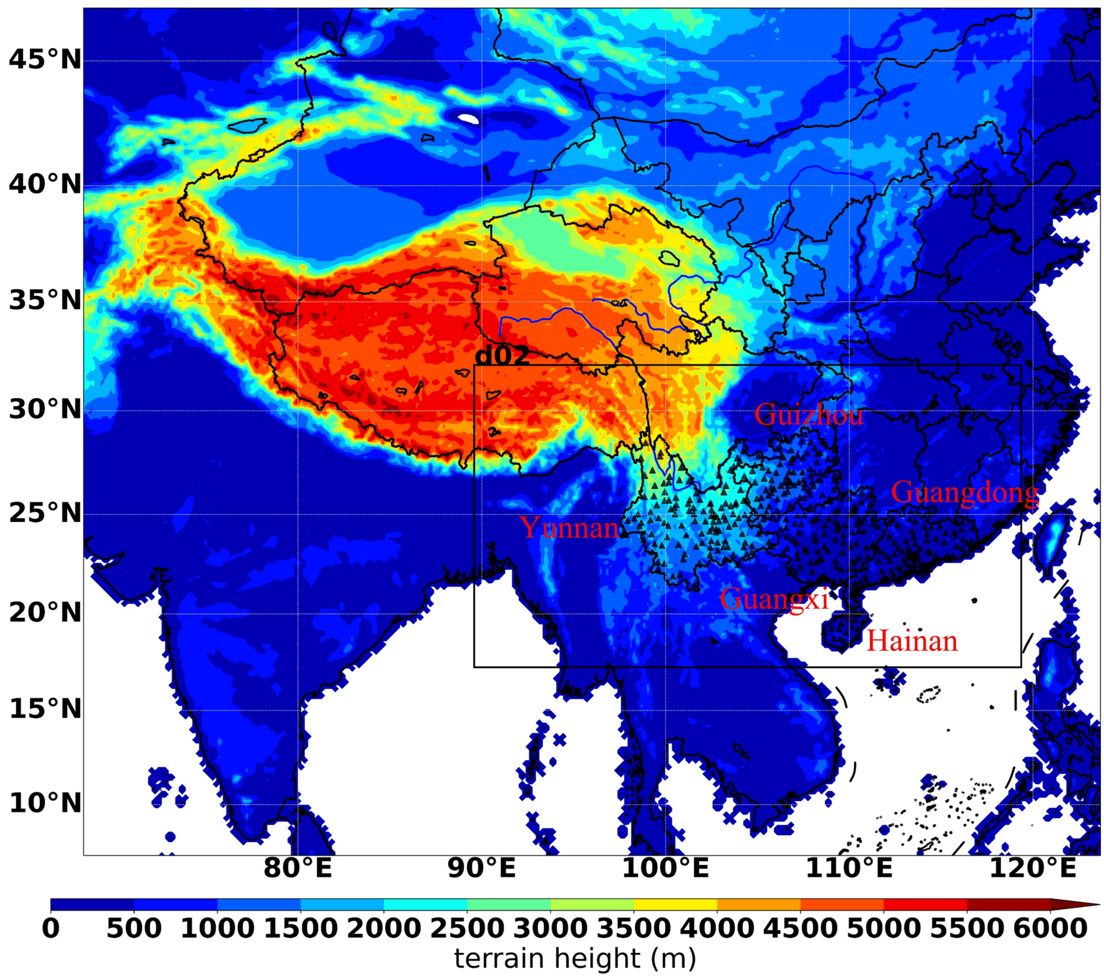

Figure 1. WRF model simulation area elevation diagram. (d02 represents the nested area of the second layer of the WRF model, and the black triangles represent the meteorological sites).