

## Response to reviewer comments

We are sincerely grateful to the editor and reviewers for their valuable time spent on reviewing our manuscript. The comments are very helpful and valuable, and we have addressed the issues raised by the reviewers in the revised manuscript. Please find our point-by-point response (in blue font) to the comments (in black font) raised by the reviewers.

Referee #1:

### COMMENTS TO THE AUTHOR(S)

This paper used the random forest models to predict typhoon-associated air quality in the Guangdong-Hong Kong-Macao Greater Bay Area. This work is interesting and meaningful to the readers of AMT. Generally this paper is well organized. The methods and conclusions are reliable. But I still have some suggestions below.

Response: Thank you so much for your efforts for reviewing our manuscript. The comments are very constructive, and we have now further revised our manuscript in light of these comments.

Detailed comments:

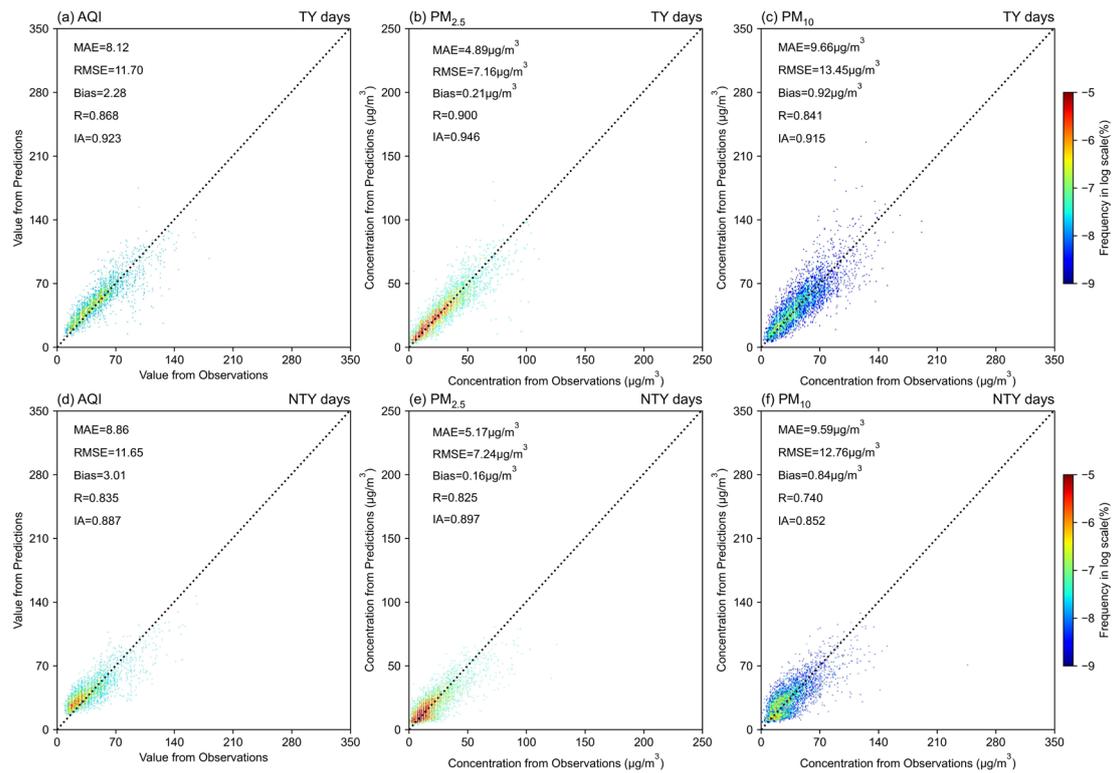
1. This paper used data from 36 air quality monitoring stations in 10 cities of the GBA that is heavily polluted. Why did you not consider the observation data from the rural regions of Guangdong Province?

Response: Many thanks for your kind suggestion. Relative to rural areas, both the population and stations are denser over urban area. Our objective is to explore the performance of the machine learning model over urban areas in the Guangdong–Hong Kong–Macao Greater Bay Area. We hope this study can help improve the prediction and assessment of typhoon-associated air quality over cities in the GBA.

Are the 36 air quality monitoring data are used for both model training and RF model evaluation? I think it may be more reasonable if you could choose parts of the 36 sites for model training and the others for RF model evaluation.

Response: Thank you for your comments. Yes, the data for 36 monitoring stations are used for both RF model training and testing. We divide the whole dataset into the testing set and training set to ensure that model will not see the training data in the testing stage, and 10-fold cross-validation is applied to minimize the data proportion bias. Besides, we employed an independent sample during typhoon Danas (outside of the training and testing dataset) to test the RF model's generalization ability. Per your suggestion, we added more samples from three new stations to evaluate RF model (lines 87-88). Figure 4 shows the result of three testing stations in Guangzhou, Shenzhen and Hong Kong on both TY and NTY days. Lines 176-179 in section 3.1.1 show the results for these three stations, the  $R$  (RMSE) values for AQI, PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub> and O<sub>3</sub> on TY days are 0.868(11.70), 0.900 (7.16  $\mu\text{g}/\text{m}^3$ ), 0.841 (13.45  $\mu\text{g}/\text{m}^3$ ), 0.496 (5.38  $\mu\text{g}/\text{m}^3$ ), 0.538 (27.94  $\mu\text{g}/\text{m}^3$ ) and 0.878 (22.45  $\mu\text{g}/\text{m}^3$ ), respectively. The result indicates that the RF model successfully captures the correlation between typhoon's location and monitoring stations' location. Though the input stations changed, the model still produces accurate prediction based on the relative position of the station and the

typhoon.



The results of SO<sub>2</sub>, NO<sub>2</sub> and O<sub>3</sub> are shown in Fig. S4, which is consistent with the results of AQI, PM<sub>2.5</sub>, PM<sub>10</sub>.

2. In section 3.1, did you compare the model performance of the RF model with the other traditional air quality models, e.g., CMAQ, WRF/Chem.

Response: Thanks for your suggestion. Compared with traditional numerical approach, ML methods have the advantage of lower computational cost, which give ML models huge application potential. Therefore, we aim to explore the ability and application of the machine learning model in the present study. The numerical models such as CMAQ, and WRF/Chem will be compared with RF model in our future work.

Figure 5: Why the data are only available over the seas?

Response: Thank you for comment. We followed your suggestion and edited the figure. The data are available in the land now, and the results are illustrated by scattering points instead of shading as following. Because we added a new figure, it is Fig. 6 now.

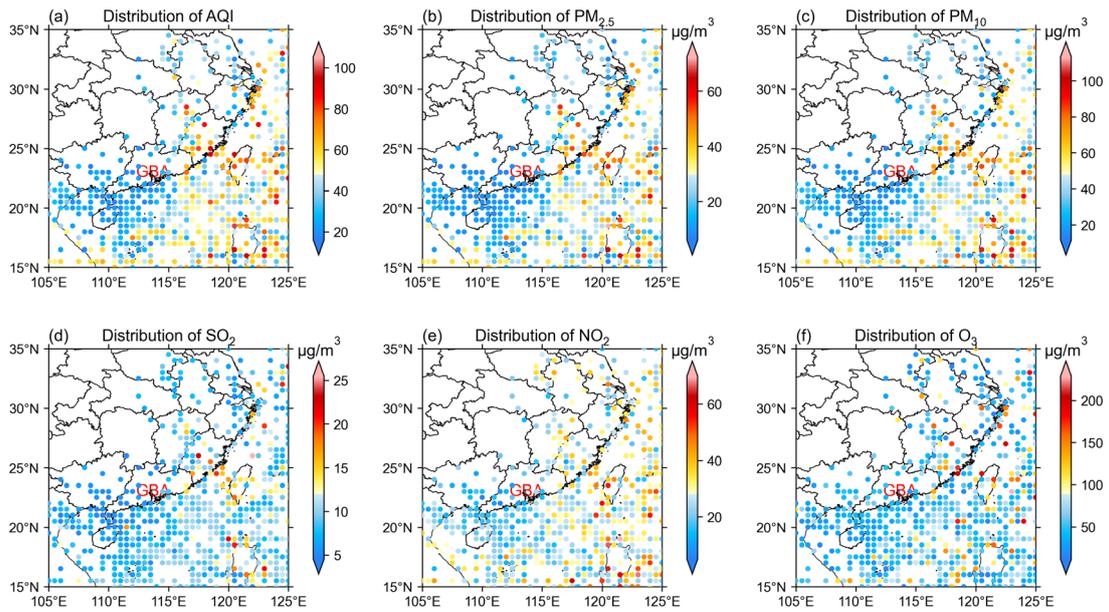


Figure 6: The same data at the air quality monitoring sites are first used for model training and then for model evaluation?

Response: Thanks for your suggestion. They are different and independent. We have added this point at lines 256-258 in section 3.3 as follows:

This paper takes typhoon Danas (2019) as an independent case to analyze the model's ability to predict typhoon-associated air quality over the GBA. For better evaluation of RF model, typhoon Danas's data have been removed from the dataset in training and testing steps.