Response to Comments on the Manuscript:

"An optimized semi-empirical physical approach for satellite-based $PM_{2.5}$ retrieval: embedding machine learning to simulate complex physical parameters"

Response to Comments of the Editor:

We would like to thank the editor for his precious time. A response to the comment follows.

Response: We have carefully read the reviewers' comments and feel particularly grateful to them for their wise suggestions. According to the comments, we have made further adjustments to our manuscript, especially in the grammar of the article. And all changes are highlighted in yellow color in the manuscript. Other minor problems have also been responded to and revised one by one.

Response to Comments of Reviewer #2:

We would like to take this opportunity to gratefully thank the reviewer for his/her constructive suggestions for improving the paper. An item-by-item response to the comments raised by the reviewer follows.

Comments:

1. Lines 73-75: The sentence is confusing and may contain a grammatical error. It may be helpful to include a comma between ‘S’ and ‘and’ in order to improve clarity. The suggested sentence would be: “Based on 355nm-band radar observations, Raut and Chazette (2009) introduced a specific extinction cross-section to simplify the expression of S, and the PM2.5 concentration was estimated.”

Response: Thank the reviewer for pointing out this problem. According to the comment, we have added a comma between ‘S’ and ‘and’ in the sentence mentioned above (Page 3 Line 75). We hope it is clear now.

2. Line 294: It could be revised to “Then, we calculate….” The use of imperative sentences in the manuscript is not common or formal in academic writing.

Response: We gratefully thank the reviewer for this comment. According to the reviewer’s suggestion above, we have changed the statement to “Then, we calculate $PM_{2.5}$ according to the corresponding process” (Page 12 Line 294). Also, we have checked the usage of other imperative sentences in the article.

3. Lines 335-336: Does it mean you feed Phy-DL FMF to the trained model, which is based on S-FMF? If so, I would suggest conducting an additional experiment to test the performance of VEf using S-FMF and Phy-DL FMF under experiment 1 (e.g., Isolated-validation with inputs of Phy-DL FMF) and including the results in Section 5.4.3.

Response: Thank the reviewer for the comment. In our experiment, Phy-DL FMF is introduced into the RF model to replace S-FMF. The basis of the above replacement is that the accuracy of Phy-DL FMF is relatively consistent with that of S-FMF. The previous studies (Lines 146-148) have made a comprehensive comparison between two FMFs, and Phy-DL FMF shows a high accuracy (R = 0.78, RMSE = 0.100). Therefore, data replacement can be implemented to achieve point-to-surface extension.

For the additional experiment mentioned by the reviewer, the current experimental results can already prove the performance of Phy-DL FMF (Page 13 Lines 312-327, Section 4.1). In experiment 4.1, the estimated value of VE_f was obtained by inputting Phy-DL FMF and was compared with the true value of VE_f (obtained from S-FMF), which is to some extent an independent validation process. At the same time, the isolated validation in experiment 4.1 exactly verified the accuracy of VE_f results by inputting Phy-DL FMF.

4. Line 477, Table 5: It will be helpful to include a comparison based on the same valid DOY (e.g., the intersection of valid DOY of MODIS FMF and Phy-DL FMF) to ensure a fair comparison.

Response: Thank the reviewer for the comment. Different DOY can exactly compare the performance of two FMFs. Section 5.1 compares the $PM_{2.5}$ accuracy using two FMF data in 2017. Although the specific validation time of two FMF varies, the overall accuracy of the $PM_{2.5}$ estimation (which can be regarded as the average accuracy over the year) shows that the Phy-DL FMF increases R to 0.68 (MODIS FMF: 0.38) with low uncertainty (Page 25 Lines 585-588). It is to some extent a fair comparison when focusing on the overall accuracy of the estimated $PM_{2.5}$. See Pages 24-25 Lines 579-588 for detailed descriptions.

If we take the intersection of valid DOY of MODIS FMF and Phy-DL FMF to compare the accuracy as mentioned by the review, it can be found that there are too few valid DOY for the intersection, which will lead to unreliable comparison results.

5. Lines 532-534: This sentence "Analyzing the model construction….." is difficult to read and understand. I recommend revising it for better clarity and flow. There are similar issues with other sentences in the manuscript that the authors may want to address before publication.

Response: Thank the reviewer for pointing out this problem. For clarity, we have changed the statement to "From the perspective of model construction" (Page 23 Line 532). According to the reviewer's suggestion, we have carefully checked the entire article for other identical errors. The modifications include 1) Page 14 Lines 340-341, "Here, RF-PMRS simulates VE_f based on RF, and replaces the polynomial of the PMRS method."; 2) Page 16 Line 374, "To visually compare the optimization effect, Fig. 6 plots the $PM_{2.5}$ bias distribution patterns for two methods."; 3) Page 20 Lines 465-466, "the experiment takes the BJ and BC sites as examples (in 2017), and then compares the $PM_{2.5}$ accuracy…". We hope it is clear now.

6. Lines 573-574: It would be helpful to include a comparison between Phy-DL FMF and S-FMF to support the statement made here. Alternatively, references could be included to provide additional support.

Response: Thank the reviewer for the suggestion. The previous studies (Yan et al., 2022) have made a comprehensive comparison between two FMFs, and Phy-DL FMF shows a high accuracy (R = 0.78, RMSE = 0.100). Therefore, data replacement can be implemented to achieve point-to-surface extension. According to the comment, we have added references to the article (Page 24 Lines 573-574) to provide additional support.

References:

Yan, X., Zang, Z., Li, Z., Luo, N., Zuo, C., Jiang, Y., Li, D., Guo, Y., Zhao, W., Shi, W., and Cribb, M.: A global land aerosol fine-mode fraction dataset (2001--2020) retrieved from MODIS using hybrid physical and deep learning approaches, Earth Syst. Sci. Data, 14, 1193-1213, https://doi.org/10.5194/essd-14-1193-2022, 2022.

Response to Comments of Reviewer #3:

We would like to take this opportunity to gratefully thank the reviewer for his/her constructive suggestions for improving the paper. An item-by-item response to the comments raised by the reviewer follows.

Comments:

Comment #1: The authors may not correctly understand my comment. In the original introduction, they introduce three kinds of method to calculate PM2.5. The second one is the univariate/multivariate regression. The third one is the semi-empirical ML approach. The second one could be the baseline. Although the semi-empirical method is the physics informed, it could be necessary to compare it with the baseline to prove that the accuracy is better than the end-to-end ML method.

Response: Thank the reviewer for the comments. The baseline method in our experiment is the semi-empirical physical method (i.e. the third one mentioned by the reviewer), not the end-to-end ML method (i.e. the second one mentioned by the reviewer). To be specific, we used the original semi-empirical physical model (PMRS) as the baseline. Then we proposed an optimization method called RF-PMRS, which embeds ML into PMRS. All experiments in the article have demonstrated the improvement of the proposed method (RF-PMRS) compared to the baseline method (PMRS).

The purpose of our article is to utilize ML to improve the expression of complex parameters in semi-empirical physical models and the experiment does not involve end-to-end ML models. As for the method comparison, since the semi-empirical physical method and the ML method have different mechanisms (model-driven and data-driven, respectively) and require different variables, it is necessary to consider various aspects of the two types of methods and their results cannot be easily compared. The semi-empirical physical model only uses four variables and does not rely on ground stations. But the method categories (such as RF) and input variables selected by the ML method affect the $PM_{2.5}$ estimation results, and the modeling relies on dense ground sites.

Therefore, different ML models, input variables, and ground values will affect the credibility of comparative experiments with the semi-empirical physical model. In the future, we will continue to explore the accuracy comparison issues of ML methods and semi-empirical physical methods, as mentioned by the reviewer.

Comment #7: The authors didn't reply how the aerosol type affect the prediction. They mentioned several times about the aerosol type. It could be the important factor for final PM2.5.

Response: The PMRS method (Zhang and Li, 2015) does not consider the aerosol type as an input factor, and our work is improved based on PMRS, so this variable is not considered. The aerosol type corresponds to the spatiotemporal distribution pattern of $PM_{2.5}$ in different regions, and **it is only used as a reference for the experimental selection of AERONET sites for RF modeling.** Referring to previous articles, Zhang and Li (2015) have elaborated on the characteristics of particle volume size distribution (PVSD) of different aerosol types in detail. **So the aerosol type does not participate as an input factor in $PM_{2.5}$ prediction.** On the other hand, the purpose of RF-PMRS is to explore a universal model from the obtained point-matching data pairs and generalize it to the space-time continuous surface data for VE_f derivation. The modeling considers different sites with four typical aerosol types and other spatiotemporal variables, which may optimize the fitting process. Therefore, **"how the aerosol type affects the prediction" is not the research direction of this article.** In the future, we will take this constructive suggestion into account, which may improve the accuracy of our method.

Thank the reviewer for the comment. For clarity, we have added a discussion on aerosol types in the article (Page 23, Lines 532-534).

References:
Zhang, Y., and Li, Z.: Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation, Remote Sens Environ, 160, 252-262, https://doi.org/10.1016/j.rse.2015.02.005, 2015.

Comment #12: The correct reference for RF should be "Ho, Tin Kam. "Random decision forests." In Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278-282. IEEE, 1995."

Response: Thank the reviewer for pointing out this problem. According to the comment, we have made changes to the reference for RF cited in the article (Page 12 Line 278; Page 31 Lines 750-752). We hope it is clear now.

Comment #16: I think it could be helpful if the authors add the discussion in the manuscript.

Response: Thank the reviewer for the comment. The article already includes a discussion part, which is divided into four main sections (Page 20 Line 462). Section 5.1 to Section 5.3 are supplementary experiments that demonstrate the superiority of the data and ML model used in our experiment. Section 5.4 focuses on the overall performance of the RF-PMRS method, which analyzes:

A. the universality of RF-PMRS to answer why it applies to North China (Section 5.4.1);
B. limitations of the validation experiments and reasons (Section 5.4.2);
C. variable uncertainty and it is carried out from five aspects (Section 5.4.3).

For specific descriptions, please see Section 5 in Pages 20-25. And we hope it is clear now.

Thanks again to the reviewers for giving us a chance to revise and improve the quality of our article. In all, we find these comments quite helpful. We wish this revision will be acceptable.

Thanks to the editor for his consideration. If you still have any questions about our study, don't hesitate to contact us.