Reply to Reviewer #1 for Manuscript of "Estimating hourly ground-level aerosols using GEMS aerosol optical depth: A machine learning approach" by O et al.

This manuscript is purposed on the PM concentration estimation based on the GEMS AOD observation data.

Because the GEMS is the geostationary orbit satellite, the idea in this manuscript has advantage of diurnal change monitoring of PM concentration from the satellite measurement.

⇨ *We thank the reviewer for recognizing the value of our study and for their thoughtful comments and efforts in reviewing our manuscript. Below, we provide point-by-point responses to the comments.*

⇨ *Here, we summarize the main revisions for the manuscript:*

- ***Data Update:*** *The dataset has been extended to December 2023, covering two years.*
- ***Additional Machine Learning Algorithm:*** *To enhance the robustness of our machine learning-based modeling, we have employed an additional algorithm, XGBoost (XGB).*
- ***Expanded Input Features:*** *We have considered additional chemical gas features, such as SO_2, NO_2, and O_3, in the modeling process.*
- ***The manuscript structure*** *is reorganized. Specifically:*
- *Data and Methodology sections are separated.*
- *Results with PM2.5 are reported in the main text.*

However, to understand and check this characteristics, this manuscript needs to some additional analysis. Especially, the machine learning method is not a perfect approach and its result can be changed by the input data selection.

For this reason, idealizing and analyzing the input variable for the machine learning method is essential to include the manuscript.

⇨ *We appreciate the reviewer's insightful comment. We agree that understanding the characteristics of the machine learning (ML) approach and its sensitivity to input data selection is crucial. To address this concern, we have conducted additional analyses focusing on the input data. Nonetheless, we would like to clarify that our primary aim is not to develop the best model for estimating PM concentrations but rather to evaluate the potential of new satellite data (GEMS AOD) as a valuable input for PM estimation. In this context, ML serves as an optimal tool due to its flexibility and ease in testing new data alongside other inputs that may contribute to improving model performance.*

For the detail, I listed to the below.

1) Introduction: For the readability of the manuscript, the author will be added to the brief explanation of sections in the final part of the Introduction section.

⇨ *Thank you for the suggestion. We have added the description on the sections in the end of the introduction: In the following sections, we first describe the data and its preprocessing in Section 2. Section 3 details the methodology, including the machine learning models employed for PM estimation. In Section 4, we present the results and discuss their implications, followed by conclusions and future research directions in Section 5.*

2) Adding the reference

- Because the study of satellite retrieved AOD was largely evaluated, the reference and related paragraph will be added before the paragraph of Line 33 (Before GEMS AOD study)

⇨ *We have added the following text in Introduction: [...] the accuracy of satellite AOD data needs to be validated to ensure their reliability for downstream applications, including PM estimation. Typically, this involves comparisons with ground-based measurements (e.g. Ogunjobi and Awoleye, 2019; Mangla et al., 2020). For instance, Choi et al. (2019) evaluated various satellite-derived AODs against ground-based measurements collected during the 2016 KORUS-AQ campaign in East Asia. Similarly, Cho et al. (2024) validated the performance of GEMS aerosol products against ground measured data. Both studies revealed a strong correlation between satellite and ground-based AOD measurements, demonstrating the utility of satellite-derived AOD for monitoring data-scarce regions.*

- L35: The references related to AOD definition and its retrieval method will be added, such as Go et al. (2020).

⇨ *Thank you for the suggestion. We have added more references including Torres et al., (2007), Go et al. (2020), and Cho et al. (2024): The GEMS aerosol retrieval algorithm (AERAOD) uses the optimal estimation (OE) method, which integrates satellite-observed radiances with initial estimates of aerosol properties, including AOD, derived from the two-channel inversion approach employed by the OMAERUV algorithm (Torres et al., 2007; Cho et al., 2024). The GEMS aerosol products provide final AOD at three wavelength channels with a nominal spatial resolution of 3.5 km x 8 km at Seoul. Further details about the GEMS aerosol retrievals can be found from NIER (2020) and Go et al. (2020).*

- L56: What is 'Korea Environment Corporation'? Is that 'Korean Environmental Institute'?

⇨ *K-eco is a government-affiliated public institution under the Ministry of Environment in South Korea, and it is responsible for managing AirKorea. We have modified the text as "[...] from the AirKorea real-time ambient air quality monitoring system operated by the Korea Environment Corporation, a government-affiliated public institution under the Ministry of Environment."*

- L57: For the PM concentration measurement, the author will be added the related references.

⇨ *The following reference is added: Hauck, H., Berner, A., Gomiscek, B., Stopper, S., Puxbaum, H., Kundi, M., and Preining, O.: On the equivalence of gravimetric PM data with TEOM and beta-attenuation measurements, J. Aerosol Sci., 35, 1135–1149, doi:10.1016/j.aerosci.2004.04.004, 2004.*

- L63: What is ARA? Need to clarify.

⇨ *Aerosol Retrieval Algorithm. We have clarified this in the reference list as "Algorithm Theoretical Basis Document (ATBD) for the GEMS aerosol retrieval algorithm."*

3) From the Section 2, this manuscript included both methodology and result parts. I suggest that this section separates the method section (e.g., Section 2) and Result section (e.g. Section 3), and the author will make sub-sections for the detailed explanation of each parts. In this version of the manuscript, method and result parts are too short to clarify the detailed machine learning method and the reason of variable selections. For this reason, the manuscript is not able to identify the difference of research compared to the several previous studies for PM estimation by the machine learning. The author have to include the table for the list of the selected variable and selection criteria.

⇨ *Following the reviewer's suggestion, we have restructured the manuscript to separate the methodology and results sections for greater clarity. The manuscript is now organized into Section 2: Data, Section 3: Methods, and Section 4: Results. Additionally, Section 4 has been divided into three subsections for 1. Direct comparison between GEMS and Aeronet AOD, 2. Use of GEMS AOD in estimating ground PM, and 3. Improving ML-based PM estimation.*

⇨ *To address the reviewer's comment, we have explained the reason for the variable and data selections in Sect. 2, with a table listing the selected variables: Those input variables are selected with reference to previous studies (Yang et al., 2020; Handschuh et al. 2022), including Seo et al., (2014), which conducted experiments in South Korea. [...] We obtain input data from reanalysis datasets, which are readily available across all areas within the GEMS satellite observation coverage. This ensures that the experiment conducted in this study can be easily extended to other locations, including other Asian countries, particularly in areas where meteorological measurements are unavailable. Additionally, reanalysis datasets provide consistent and reliable data updates over space and time. Nonetheless, it is well known that gases such as CO, NO_2, and SO_2 can influence PM formation mechanisms either directly or indirectly (Lee et al., 2024). Therefore, we also incorporate chemical data measured at the AirKorea stations as additional input features. In this way, we can evaluate the potential improvements in PM estimation using AOD when supplemented with additional information, and we report the corresponding results. The input variables used in this study are listed in Table 1.*

Table 1. List of data used in PM modeling

Category	*Name*	*Description*	*Data source*
Input data			

Aerosol data	*AOD*	*aerosol optical depth*	*GEMS satellite*
Meteorological	*BLH*	*boundary layer height*	*ERA5 reanalysis*
	RH	*relative humidity*	*ERA5-Land*
	TEMP	*air temperature*	*ERA5-Land*
	SP	*surface pressure*	*ERA5-Land*
	WS	*wind speed*	*ERA5-Land*
	WD	*wind direction*	*ERA5-Land*
Chemical	*CO*	*Carbon Monoxide*	*AirKorea station*
	SO2	*Sulfur Dioxide*	*AirKorea station*
	NO2	*Nitrogen Dioxide*	*AirKorea station*
	O3	*Ozone*	*AirKorea station*
Output data			
Particulate matter	*PM10*	*particles with a diameter of 10 micrometers or less*	*AirKorea station*
	PM25	*particles with a diameter of 2.5 micrometers or less*	*AirKorea station*

4) L59-L64: For the data colocation between ground and satellite pixels, temporal colocation is clarfied in the manuscript. However, the spatial colocation between Airkorea and GEMS pixel, the manuscript is explained only the 'nearlest pixel'. How to be selected the 'nearlest pixel'? Because the cloud contamination of the GEMS AOD value, some pixels have to be eliminated, and the spatial distance between two measurements will be far. Do you have criteria of the maximum distance? In addition, the ground observation stations are dense in the urban region, especially denser than the GEMS spatial resolution. In this case, the same GEMS pixel is duplicately selected in different AirKorea observation sites. In this case, how to correct the colocation method?

⇨ *Please note that GEMS L2 data is provided on a pixel basis, and we selected the ' nearest pixel,' not the 'nearest observation.' This means that GEMS AOD values at the* *nearest pixel are often missing due to issues like cloud contamination, leaving an average of 1990 data pairs (~11%) per station over the two-year period.*

⇨ *The average distance between GEMS pixels and AirKorea stations ranges from 1.5 to 2.7 km, while the average distance among AirKorea stations is 7.2 km. Therefore, we believe the impact of pixel duplication is minimal.*

⇨ *Further, we have confirmed that the spatial distances between GEMS pixels and ground stations do not significantly affect PM estimation performance, as shown in the figure below.*

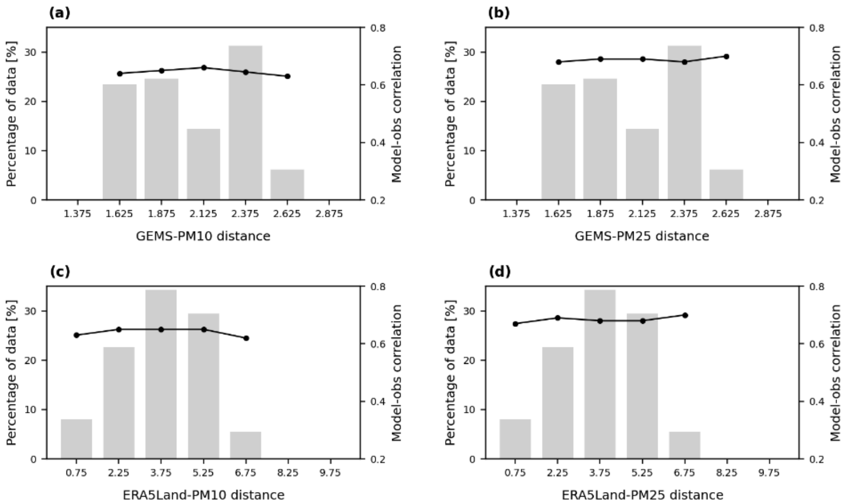


Figure S2. *PM estimation performance by distance between data sources and measurement stations. (a) and (b) show the distribution of distances between GEMS AOD observation pixels and PM measurement stations, along with average model performance (correlations) for PM10 and PM2.5, respectively. (c) and (d) depict the same, but for the distances between meteorological input data (i.e. ERA5-Land) and PM measurement stations.*

- In addition, for the colocation between observation and reanalysis dataset, what kind of interpolation method is used in this study? If you simply selected the 'nearlest grid', it may affect the uncertainty.

⇨ *We agree with the reviewer that simply selecting the "nearest grid" may introduce uncertainty. To address this, we have updated our colocation method to use Inverse distance weighting (IDW) is used. The updated method is explained as follows: Both gridded datasets are interpolated to the locations of AirKorea stations using inverse distance weighting (IDW) based on the four closest grid points. If data are missing in the nearest grid points (e.g., over ocean areas), the corresponding locations are excluded from the analysis*.

5) L132: Boundary Layer Height (BLH) is not a linear relation to correct the PM concentration from satellite AOD. The BLH is roughly changing the PM concentration. But its sensitivity is also changed by the columnar concentration of aerosols. Did the author check the sensitivity change of BLH for the relationship between PM concentration and satellite AOD? (Including the reference survey)

⇨ *In our machine learning approach, the model learns patterns and relationships directly from the data without explicitly assuming linear or non-linear relationships like the sensitivity of BLH. While the analysis of the sensitivity change of BLH is beyond the scope of our study, we acknowledge the importance of understanding these physical relationships.*

⇨ *To address this, we have expanded our explanation of the role of BLH as follows: "BLH is a good proxy with which to estimate the height of the aerosol layer and can help relate columnar satellite data to surface aerosol values. The relationship between AOD and PM is highly sensitive to variations in BLH conditions, as noted in previous studies (e.g. Zhang et al., 2016; Zheing et al. 2017). For instance, higher BLH facilitates greater vertical dispersion of aerosols, thereby reducing surface PM concentrations for a given AOD. While our machine learning approach inherently captures such complex interactions from the data, future work could explore the explicit sensitivity of BLH within the AOD-PM relationship to improve physical interpretability."*

6) Section 2.1: Although the supplement part include the PM2.5 result, the body of the manuscript is not shown the detailed analysis of PM2.5. I suggested that both the PM10 and PM2.5 estimation method and SHAP analysis will be included separately. Also, the detailed SHAP analysis results have to be included with the detailed analysis and explanations. If the author compares the difference between PM2.5 and PM10 estimation, it is possible to evaluate the contribution of aerosol types or absorptivity. In addition, for the explanation of PM10 concentration, the manuscript is confused about what is 'observed from AirKorea' and 'satellite retrieved PM concentrations'. The author will clarify the word for satellite-derived PM concentration and Ground-based observed PM concentration.

⇨ *We appreciate the reviewer's detailed feedback and suggestions. While it would be interesting to investigate aerosol types or absorptivity, SHAP values primarily indicate the importance of input features and do not provide direct evidence to analyze such detailed characteristics. Thus, we consider this aspect suitable for future research.*

⇨ *Following the reviewer's suggestion, we have included the results for PM25 in the main text, along with its SHAP analysis. Please note that a computational error in the SHAP analysis was identified and corrected, resulting in slight changes to the results. However, AOD remains one of the most important input predictors.*

⇨ *We have also expanded the explanation and discussion of the SHAP analysis as follows: "For both PM10 and PM25 estimations, AOD has a greater influence on model performance than most meteorological variables, as expected from its relatively strong correlation with ground aerosols (Fig. 2a), demonstrating the effectiveness of satellite information for ground aerosol simulations. The relationships between PM and meteorological variables are not straightforward (see Fig. S2), but SHAP analysis indicates that temperature (TEMP) is the most influential meteorological variable for both PM10 and PM25 estimations. [...] Similar results are observed for XGB (Fig S3).*

Among other meteorological variables, relative humidity (RH) emerges as one of the most important predictors for PM10 estimations. RH influences wet deposition of PM and also characterises seasonal variations in PM concentrations. For instance, in Korea, PM10 tends to increase due to wildfire emissions during dry seasons or from relatively coarse particles transported by Asian dust in the spring season. On the other hand, *BLH has a greater importance for PM25. As aerosols are primarily confined to the planetary boundary layer, BLH is a good proxy for estimating the height of the aerosol layer (Lee et al., 2024) and can help relate columnar satellite data to surface aerosol values (Handschuh et al., 2022; Gupta and Christopher, 2009a). The stronger importance of BLH for PM25 suggests its effectiveness in capturing the vertical distribution of finer aerosols. Previous studies have suggested a strong relationship between BLH and PM25 due to well-recognized positive feedbacks (e.g. Li et al., 2017; Su et al., 2022), which further underscores the importance of BLH. The significant influence of both BLH and RH on the relationship between AOD and PM has also been reported by Zheng et al., 2017.*

⇨ *To address the confusion regarding terminology, we have carefully reviewed and revised the manuscript to clearly distinguish between satellite-derived and ground-based observed PM concentrations.*

7) Figure 4 and 5: Re-arrange the time scale (24 hours -> Daytime)

⇨ *We have revised Figure 4 to reflect daytime data only, as suggested. Please note that Figure 5 has been removed in the updated manuscript.*

8) L161-L170: The author mentioned that the main reason for estimated PM underestimation is due to the GEMS AOD underestimation. However, this study's method made the machine learning model based on the GEMS AOD. If so, the uncertainty characteristics of GEMS AOD is adopted in the machine learning modeling. Another possibility of the estimated PM underestimation is the false selection of variables or lack of the variable for the machine learning method. From several previous studies, the PM concentration is not affected only by the meteorological components, but also by the chemical processes. The author has to check the variable selections.

⇨ *We agree with the reviewer that machine learning models should account for biases inherent in GEMS AOD. Following the reviewer's suggestion, we conducted additional experiments by incorporating chemical data as input variables for PM simulations. These experiments demonstrated that including chemical data can indeed improve the model's performance. However, the primary aim of our study is to use 'commonly-available' input data, such as reanalysis meteorological data and GEMS AOD, to ensure the generalizability of our approach to other Asian countries, where ground measurements may not be readily available.*

⇨ *Therefore, we have reported the results of these additional experiments as a test analysis under Subsection 3.3. This highlights the potential for improved model performance when supplementary data, such as chemical pollutants from ground stations, are incorporated. In fact, while model performance improves with the inclusion of additional input data, discrepancies between the measured and estimated PM concentrations persist. These discrepancies could be attributed to limitations in the quality of AOD data (e.g., biases compared to ground-based AOD measurements) and/or inherent limitations of the machine learning models.*

⇨ *The new result is reported in Fig. 6.*

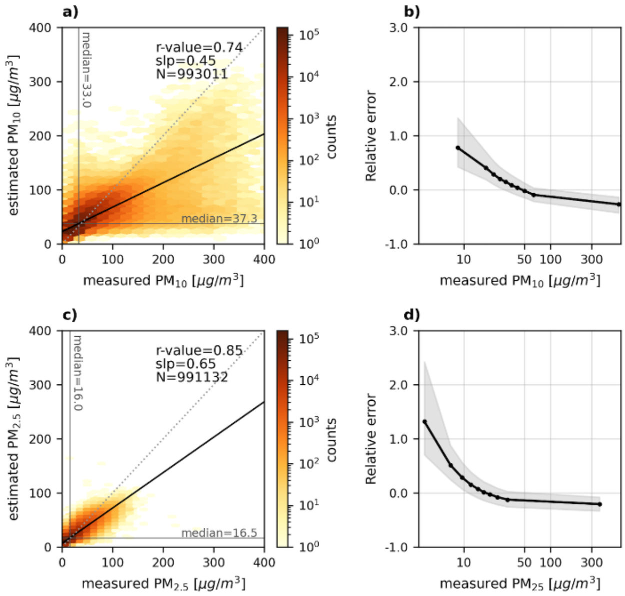
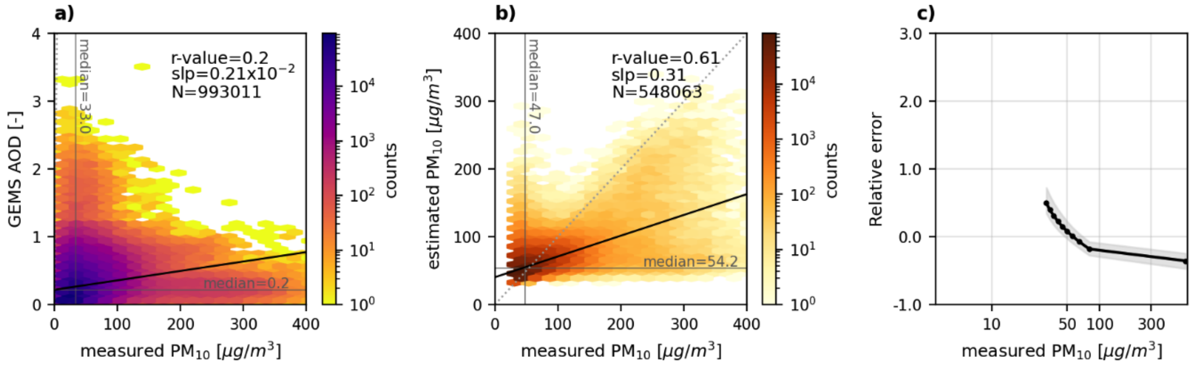


Figure 6. Performance improvement of RF models with the inclusion of pollutant data. *This figure is similar to Fig. 3, but shows results for models trained with additional locally available data (O_3, CO, SO_2, NO_2) from AirKorea stations. (a) Density scatter plot between measured PM10 and model-estimated PM10. (b) Relative errors are shown at each 10th percentile of measured PM10. Panels (c) and (d) correspond to (a) and (b), respectively, but for PM25.*

9) For the Machine learning adaptation, do you have the criteria of minimum concentration of observed PM and minimum value of satellite retrieved AOD? Low concentration of aerosol cases may be affecting the overall performance of estimation.

⇨ *Inspired by the reviewer's comment, we conducted additional experiments using ML models trained exclusively on high PM concentration cases. Unfortunately, this approach did not result in any noticeable improvement in model performance (see the figure below). Instead, we found that model performance can be enhanced by incorporating more informative input data, such as chemical pollutant measurements (see our responses to the comment above).*

⇨ *While there may be other approaches to further improve ML model performance, we consider this to be outside the scope of the current study. Our primary objective is to evaluate the usefulness of GEMS AOD in PM estimation and provide a baseline model performance for future studies.*



10) L179-L184 and Figure 6: For the statistical score, a detailed explanation will be needed. In addition, in Figure 6, 'n0', 'n1',' n2', 'n4', and 'n8' are not explained in the caption of Figure 6 and the body of the manuscript. The author has to clarify the explanation of Figure 6 and the mean of the statistical score.

⇨ *We apologize for the lack of clarity in the explanation. It refers to the number of neighboring stations providing the training dataset. We have updated the figure caption to clarify this. Additionally, please note that we have slightly modified the experimental setup to demonstrate the potential of satellite-derived AOD and machine learning for estimating PM concentrations at ungauged locations.*

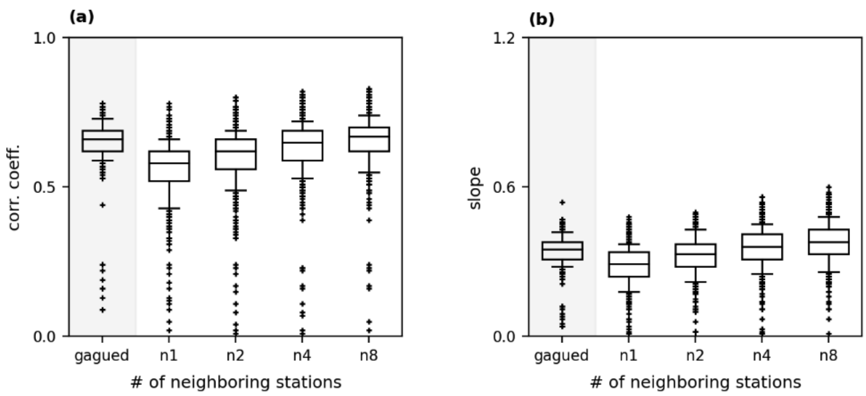


Figure 7. Potential of satellite data in PM estimation at ungauged locations. *(a) Correlation and (b) slope of the linear regression between the measured and estimated PM10 concentrations at each station. Data from the n closest neighboring sites are used to train the RF models, and the model performance is evaluated at the target station, where the training data is deliberately excluded for this experiment. The term 'gauged' indicates that the model is trained and tested at the same station (as shown in the main analysis in Fig. 2).*