Post-process correction improves the accuracy of satellite $PM_{2.5}$ retrievals - Reply to referees

Andrea Porcheddu, Ville Kolehmainen, Timo Lähivaara, Antti Lipponen

We would like to thank the reviewers for reading carefully the manuscript and giving their comments. Below we reply to each of the comments.

1 Answers to reviewer #1

1.1 There are many researches that focus on using AOD to estimate $PM_{2.5}$ through machine learning approaches. Compared with them, what is the innovation of this study? I understand that this study corrects the ratio between $PM_{2.5}$ and AOD derived from MERRA2 and applies the improved ratio to satellite AOD to estimate surface $PM_{2.5}$ concentrations, which is different from other researches that estimate $PM_{2.5}$ directly. Although this is a new approach, what is the advantage of this study. Compared with previous researches, can the new approach provide better $PM_{2.5}$ estimation?

The novelty of this study is to employ the post process correction where we use machine learning for correcting the geophysical model based AOD-to-$PM_{2.5}$ conversion ratio instead of directly predicting the conversion ratio. The rationale for this selection is improved accuracy over the conventional approach of learning the conversion ratio directly. Figure 1 shows the results obtained with the conventional approach of learning directly the AOD-to-$PM_{2.5}$ ratio. The R^2, RMSE and MAE with the proposed post-correction approach (left image) are better than with the conventional direct estimation (right image).

1.2 This exclude the $PM_{2.5}$ concentrations that are larger than 80 $\mu g/m^3$. This would reduce the importance of this study as the research community is more interested in heavy polluted scene. The authors excluded that condition that $PM_{2.5}$ concentrations that are larger than 80 $\mu g/m^3$ due to imbalanced data (only a small set of data with $PM_{2.5}$ concentrations that are larger than 80 $\mu g/m^3$). Can the problem is solved through bagging or other approaches?

The $PM_{2.5}$ values beyond 80 $\mu g/m^3$ were excluded from the study due to sparsity of high value data in the region of interest (central Europe, 2019) considered in this study. Similar cutoff for high $PM_{2.5}$ values has been applied, for example, in Ibrahim et al. (2022)). To address the lack of high $PM_{2.5}$ data, one could think of producing synthetic training data using machine learning models such as TVAE (Tabular Variational Auto Encoder) or CTGAN (Conditional Tabular Generative Adversarial Network) (Xu et al., 2019): we tried already to use them to balance the data but they did not improve the results. In future studies

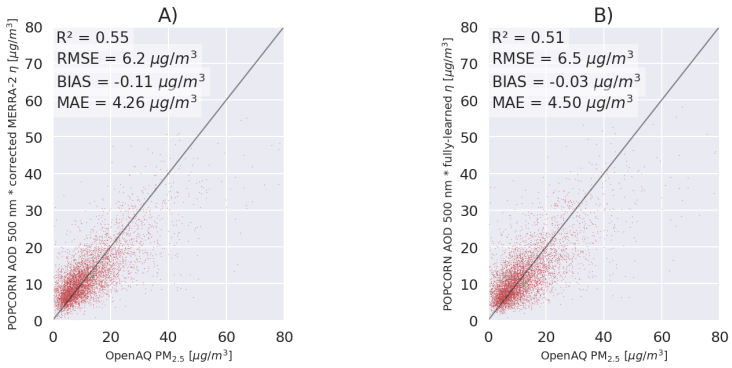

Figure 1. A) Post process corrected $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements. B) Fully-learned NOODLESALAD $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements.

the approach could be extended to more global training data and include data, for example, from India and China, where higher values of $PM_{2.5}$ exist more frequently.

1.3 More information of satellite data is needed. What is the temporal resolution and swath of the sensor?

Two Sentinel-3 satellites currently flying provide revisit times of less than two days for OLCI and less than one day for the SLSTR instrument at equator. Swath width of the OLCI instrument is 1270 km. SLSTR swath width is 1420 km for the nadir view and 750 km for the oblique view. In our study, we base our aerosol information on the official Sentinel-3 Synergy Land data product and the characteristics of that data set matches our satellite overpass data. The information have been added to the manuscript.

1.4 This study only demonstrates the validations in Fig. 4. It would be more interesting to show the fitting (training) as well.

Fig.2 shows the results on the training set. The metrics are better than on the test set (as it is expected to be), but not very different since we used the early stopping technique as regularization to avoid overfitting on the training set. We remark that in this study we separated part of the available stations to be used as independent validation data for the methods. For operational use, the model could be trained using all the available data in the training and test data sets.

1.5 In Fig.4, why monthly mean shows larger bias than instant estimations?

It is not strictly necessary that the monthly bias is smaller than the instantaneous bias. We have added mean absolute error (MAE) as an additional metric in the figures. MAE shows improvement for the monthly data over the instantaneous data.

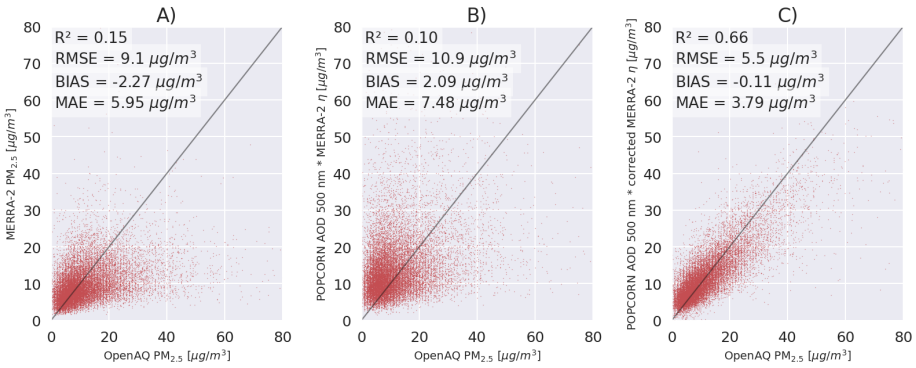

Figure 2. A) MERRA-2 $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements per single-overpass. B) Uncorrected NOODLESALAD $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements per single-overpass. C) Corrected NOODLESALAD $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements per single-overpass. These results regard the training set.

1.6 Latitude and longitude are missing in the top panels of Figs. 5 and 6.

Latitude and longitude have been added to the panels.

1.7 Line 78-80: I cannot understand. More details are needed for the method description.

Some OpenAQ stations report 24 hour average $PM_{2.5}$ every hour based on the last 24 hours.

In this work, we used the 24 hour averages given every hour to estimate hourly $PM_{2.5}$. This processing was done station-by-station using a Tikhonov regularized (with regularization parameter value 0.05) least-squares fit to unfold the time integrated data into hourly estimates.

In practice, the hourly $PM_{2.5}$ estimates were computed using the formula

$$PM_{2.5,1h} = \left(A^T A + \alpha I\right)^{-1} A^T b, \tag{1}$$

where the system matrix

$$A = \begin{bmatrix} \frac{1}{24} & \frac{1}{24} & \cdots & \frac{1}{24} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{24} & \cdots & \frac{1}{24} & \frac{1}{24} & 0 & \cdots & 0 \\ & & & \vdots & & & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \frac{1}{24} \end{bmatrix}, \tag{2}$$

is 24 hour time averaging operator and the data vector

$$b = \begin{bmatrix} PM_{2.5,24h,1} \\ PM_{2.5,24h,2} \\ \vdots \\ PM_{2.5,24h,N} \end{bmatrix}, \tag{3}$$

$$PM_{2.5,1h} = \begin{bmatrix} PM_{2.5,1h,24} \\ PM_{2.5,1h,25} \\ \vdots \\ PM_{2.5,1h,N} \end{bmatrix} \tag{4}$$

contain the hourly 24 hour averages $PM_{2.5,24h,N}$ of the station data and $\alpha = 0.05$ is the regularization parameter. The solution vector $PM_{2.5,1h,N}$ contains the unfolded 1 hour $PM_{2.5}$ at timestep N, respectively.

We have added this explanation in the revised manuscript.

1.8 Why CALIOP data are used. This is monthly mean data, but $PM_{2.5}$ and AOD has strong diurnal variation. Can the CALIOP data help to improve $PM_{2.5}$ estimation?

We tested the approach without CALIOP data in the training and the result is shown in fig.3 (left image: CALIOP included, right image: without CALIOP). The model which uses CALIOP data results in slightly better accuracy, indicating that use of CALIOP data is warranted.

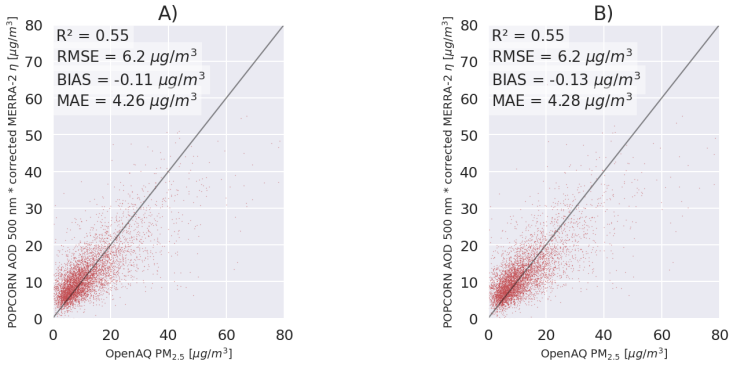

Figure 3. A) Corrected NOODLESALAD $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements per single-overpass (using CALIOP data). B) Corrected NOODLESALAD $PM_{2.5}$ predictions against OpenAQ $PM_{2.5}$ measurements per single-overpass (without CALIOP data).

1.9 The unit of RMSE is missing throughout the paper.

The unit of RMSE has been added to the paper.

2 Answers to reviewer #2

2.1 Abstract: The study lacks major conclusions and quantitative descriptive results.

The abstract has been modified as follows (extension highlighted in red):

Estimates of $PM_{2.5}$ levels are crucial for monitoring air quality and studying the epidemiological impact of air quality on the population. Currently, the most precise measurements of $PM_{2.5}$ are obtained from ground stations, resulting in limited spatial coverage. In this study, we consider satellite-based $PM_{2.5}$ retrieval, which involves conversion of high-resolution satellite retrieval of Aerosol Optical Depth (AOD) into high-resolution $PM_{2.5}$ retrieval. To improve the accuracy of the AOD to $PM_{2.5}$ conversion, we employ the machine learning based post-process correction to correct the AOD-to-PM conversion ratio derived from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis model data. The post-process correction approach utilizes a fusion and downscaling of satellite observation and retrieval data, MERRA-2 reanalysis data, various high resolution geographical indicators, meteorological data and ground station observations for learning a predictor for the approximation error in the AOD to $PM_{2.5}$ conversion ratio. The corrected conversion ratio is then applied to estimate $PM_{2.5}$ levels given the high-resolution satellite AOD retrieval data derived from Sentinel-3 observations. The region of study is central Europe during the year 2019. Our model produces $PM_{2.5}$ estimates with a spatial resolution of 100 meters at satellite overpass times with R^2 = 0.55 and RMSE = 6.2 $\mu g/m^3$. The corresponding metrics for monthly averages are R^2 = 0.72 and RMSE = 3.7 $\mu g/m^3$. Additionally, we have incorporated an ensemble of neural networks to provide error envelopes for machine learning related uncertainty in the $PM_{2.5}$ estimates. The proposed approach can produce accurate high resolution $PM_{2.5}$ data that can be very useful for air quality monitoring, emission regulation and epidemiological studies.

2.2 The introduction is very short and lacks a comprehensive review of numerous previous studies on converting AOD to $PM_{2.5}$ using machine learning models.

The introduction has been modified as follows (extension highlighted in red):

Poor air quality is one of the most serious environmental health risks of our time. In September 2021, the World Health Organization (WHO) released Global Air Quality Guidelines, revealing clear evidence of the damage air pollution inflicts on human health at even lower concentrations than previously understood (World Health Organization, 2021). WHO estimates that exposure to air pollution causes 7 million premature deaths every year. A key indicator in monitoring air quality and epidemiological studies is the $PM_{2.5}$ parameter, which is the dry mass concentration of fine particulate matter with an aerodynamic diameter of less than 2.5 micrometers (micrograms of particulate matter per cubic meter of air). Fine particulate matter originates from vehicle emissions, coal burning, and industrial emissions, among many other human and natural sources. Epidemiological studies link long exposures to high $PM_{2.5}$ levels to many severe illnesses, such as stroke and cardiovascular

and respiratory diseases (e.g. Pope and Dockery, 2006; Cohen et al., 2017). On a global scale, the magnitude of the $PM_{2.5}$ exposure-related risk for human health is enormous as more than 90% of the world's population lives in areas with annual mean $PM_{2.5}$ levels exceeding the new WHO 2021 air quality guideline of 5 micrograms per cubic meter (Health Effects Institute, 2019).

While the knowledge of the health effects of pollution increases continuously, the epidemiological estimates still have significant uncertainties due to the lack of accurate global air pollution data (Hammer et al., 2020). Networks of ground-based observation stations produce accurate pointwise observations of $PM_{2.5}$ and certain chemical components such as ozone, sulfur dioxide and nitrogen dioxide. These ground station measurements produce relatively accurate data, but the networks consist of only a few thousand irregularly located observation stations, mainly in developed countries, leading to the insufficient spatial coverage of the $PM_{2.5}$ data. To better monitor and understand air quality and pollution sources near real-time global observations of air quality are needed. The only way to get spatially resolved air quality data is to utilize satellite retrievals.

Satellite retrievals of $PM_{2.5}$ are often based on satellite AOD retrievals and AOD-to-PM conversion ratio (Health Effects Institute, 2019; van Donkelaar et al., 2013; Zhang and Kondragunta, 2021; Geng et al., 2015). AOD is a columnar optical quantity, whereas $PM_{2.5}$ is the mass concentration of dry aerosol particles at some single point, typically at the surface level. Many factors affect the AOD-to-PM conversion ratio, including the aerosol vertical extinction profile, aerosol type and size distribution, and relative humidity. These factors are typically unavailable from a single data source, such as data provided by the instruments onboard a satellite, so a simulation-model-based AOD-to-PM ratio is often used. The simulation-model-based AOD-to-PM conversion ratio is typically computed based on meteorology, chemical transport models (CTM) and auxiliary satellite data such as lidar-based aerosol vertical profiles. The $PM_{2.5}$ retrieval at a given location and time is then calculated as a product of the retrieved satellite AOD and the AOD to $PM_{2.5}$ ratio. The current state-of-the-art $PM_{2.5}$ retrieval algorithm also contains a post-processing step where the retrieved spatial $PM_{2.5}$ estimate is fitted to the ground-based $PM_{2.5}$ station data by a linear geographically weighted regression (van Donkelaar et al., 2016).

Many previous studies use machine learning techniques to convert AOD to $PM_{2.5}$ levels. In particular, (Ibrahim et al., 2022) used a variant of Random Forest called Extremely Randomised Trees (ET) to estimate $PM_{2.5}$ across Europe. (Stafoggia et al., 2019; Schneider et al., 2020) used Random Forest regressors in a multi-stage approach to estimate $PM_{2.5}$ at ground stations when only PM_{10} measurements were available, to impute AOD values when not accessible and to finally predict $PM_{2.5}$ values across Italy and Great Britain. (Handschuh et al., 2023) considered multiple Random Forest models to evaluate $PM_{2.5}$ levels across Germany using 4 different AOD datasets.

In this paper, we propose a novel approach for high-resolution satellite-based retrieval of $PM_{2.5}$. While the previous studies use machine learning to learn the AOD to $PM_{2.5}$ conversion directly, we take a novel approach where we train the model to predict the approximation error in the geophysical model based conversion ratio. Our approach retrieves $PM_{2.5}$ at the spatial resolution of 100 m. It is based on the machine learning post-process correction approach, which we developed for the correction of approximation errors in satellite retrievals (Lipponen et al., 2021) and employed for high-resolution spectral aerosol optical depth (AOD) retrieval (POPCORN AOD) from SENTINEL-3 SYNERGY data (Lipponen et al., 2022). In our algorithm development work, we take the spectral, high-resolution Sentinel-3 POPCORN AOD (Lipponen et al., 2022) as the starting point.

Our $PM_{2.5}$ retrieval is based on the AOD-to-$PM_{2.5}$ conversion ratio applied to the POPCORN AOD. The AOD-to-$PM_{2.5}$ ratio is estimated by machine learning techniques utilizing a fusion of collocated ground station-based in-situ $PM_{2.5}$ data, MERRA-2 reanalysis model AOD and $PM_{2.5}$ data, spectral AERONET AOD, satellite-observed spectral top-of-atmosphere reflectances, meteorology data and various high-resolution geographical indicators representing, for example, population density and land surface elevation. Utilizing these data, we employ the post-process correction approach to the estimation of AOD-to-$PM_{2.5}$ ratio (Lipponen et al., 2021, 2022; Taskinen et al., 2022) and then the high-resolution $PM_{2.5}$ retrieval is obtained as the product of the post-process corrected AOD-to-$PM_{2.5}$ ratio and POPCORN AOD. By using an ensemble of neural networks, we can also provide error envelopes for the machine learning related uncertainty in the $PM_{2.5}$ estimates. The approach is tested with Sentinel-3 data from central Europe in 2019.

2.3 The use of MERRA2-2 for calculating $PM_{2.5}$ is criticized for its inaccuracies and omission of certain species like Nitrate. It is suggested to consider using GEOS-CF data, which provides $PM_{2.5}$ simulations at a higher resolution of 0.25 degrees.

We thank the referee for the suggestion. We agree that GEOS-CF would be a suitable and good model data to consider in our methodology. Our methodology developed is not restricted to any single model or satellite data. Some criteria in selecting the model data for our work were long time series and widely used model in scientific literature and we therefore ended up selecting MERRA-2. In our future work, we will consider using GEOS-CF data as it has somewhat better spatial accuracy and more relevant species for air quality applications.

2.4 The spatial resolution of high-resolution indicators such as roads and nighttime lights needs clarification.

- NASA Black Marble Night Lights: We use the 500 meter resolution data.
- OpenStreetMap roads: The original data is vector data with a typical accuracy of orders of meters. We have re-projected the OpenStreetMap road data to 100 meter resolution before use.

We have added the resolutions used to the revised manuscript.

2.5 The excessive number of variables selected raises questions about their relevance and contribution to the network model. It would be beneficial to employ importance analysis methods to identify and eliminate redundant variables. This process will streamline the model and improve its efficiency and interpretability.

We used SHAP analysis (Lundberg and Lee (2017)) to estimate the feature importance after model training. A bar plot can be seen in fig.4 for the first 26 features found with the SHAP analysis. Table 1 contains all the input variables listed by importance (SHAP value). Since all the variables had non-negligible SHAP values, indicating some information content in them, we decided to keep them all. The input variable table in the manuscipt has been modified so that now the variables are listed in order of importance by the SHAP values.

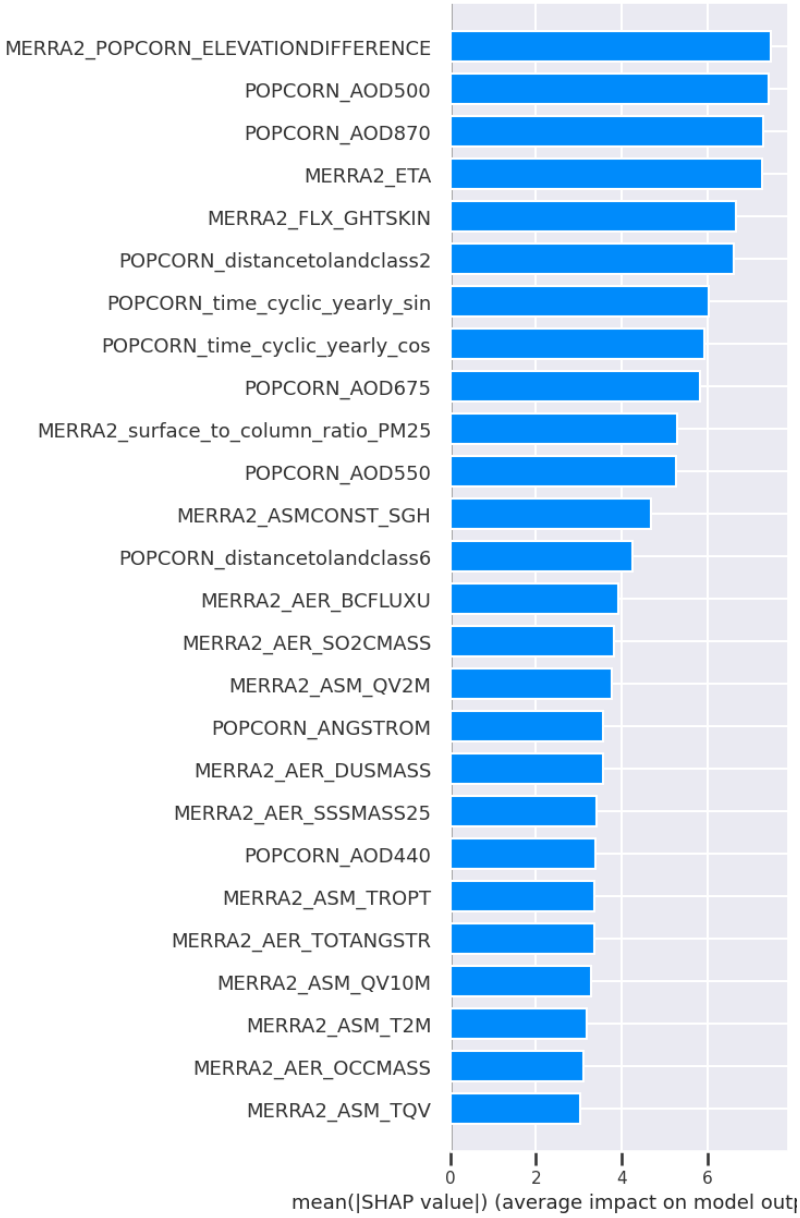

Figure 4. Bar plot of the SHAP values for the first 26 input variables in order of importance.

2.6 Figure 3: It is unclear how the training and validation stations are divided. Additionally, the proximity of stations may lead to correlation issues, affecting the independence of training and testing samples spatially.

We are not the first ones to use site-validation, please see the review of different validation methods used in the literature Tang et al. (2024). Essentially we divided randomly the stations into training set, validation set and test set, then used the related data

accordingly. For what regards the correlation issues, there should be only a minimal effect since we are operating at resolution of 100 m and there are not two or more stations in the same pixel.

2.7 Section 3.4: The rationale for choosing the neural network model over other more powerful machine learning and deep learning models is not provided. The advantages of this model should be discussed.

Using a fully connected neural network compared to other common models in $PM_{2.5}$ prediction like random forest is very suitable in the case one is working with assumption of independent pixels and a high number of data samples and features. In this case, our study is limited to the chosen ROI and period of time (2019) so we have roughly 20000 points in the training set: our plan is to extend the ROI and the period of time in future studies so to have more data in the training set, where the fully connected neural network should show all its learning capability. Comparing the fully connected neural network to other deep learning models, we think it is the most suitable architecture for the task at hand since we are dealing with tabular data and a fair amount of features. One could for example think of using a convolutional neural network and reorganizing the data samples into appropriate matrices but still the number of features is small and we don't know if we would benefit deploying a convolutional network (there's no computational burden to justify this approach and the fully connected neural network should be able to find proper representations of the input data in its hidden layers).

2.8 Figure 4: While the accuracy has improved, the correlation remains relatively low (only 0.63), compared to previous studies achieving higher accuracy with AI (R2 higher than 0.8). The significance of the study is questioned, and comparison with previous studies to assess improvement is recommended.

The R^2 coefficient is low compared to other studies but we should compare our manuscript to papers who deal with, for example. the same ROI, spatial resolution and study time period. Our RMSE is comparable to other studies that consider European countries or similar ROI (Schneider et al. (2020); Ibrahim et al. (2022); Handschuh et al. (2023)). The R^2 coefficient is low but the number of data samples at hand is low too: as mentioned before, we are considering the year 2019 and compared, for example, to Ibrahim et al. (2022) we have 10% of data. Our ensemble of fully connected neural networks would benefit from having more data and the metrics would improve further. Also the data and the preprocessing choices are different, so the studies are not directly comparable.

MERRA2_POPCORN_ELEVATIONDIFFERENCE	POPCORN_AOD500	POPCORN_AOD870
MERRA2_ETA	MERRA2_FLX_GHTSKIN	POPCORN_distancetolandclass2
POPCORN_time_cyclic_yearly_sin	POPCORN_time_cyclic_yearly_cos	POPCORN_AOD675
MERRA2_surface_to_column_ratio_PM25	POPCORN_AOD550	MERRA2_ASMCONST_SGH
POPCORN_distancetolandclass6	MERRA2_AER_BCFLUXU	MERRA2_AER_SO2CMASS
MERRA2_ASM_QV2M	POPCORN_ANGSTROM	MERRA2_AER_DUSMASS
MERRA2_AER_SSSMASS25	POPCORN_AOD440	MERRA2_ASM_TROPT
MERRA2_AER_TOTANGSTR	MERRA2_ASM_QV10M	MERRA2_ASM_T2M
MERRA2_AER_OCCMASS	MERRA2_ASM_TQV	MERRA2_FLX_QLML
MERRA2_AER_SUFLUXV	MERRA2_FLX_USTAR	MERRA2_AER_SO4CMASS
POPCORN_distancetolandclass17	MERRA2_AER_DUCMASS	MERRA2_AER_BCSMASS
MERRA2_AER_BCSCATAU	MERRA2_AER_DUEXTTAU	MERRA2_FLX_EFLUX
MERRA2_AER_SO4SMASS	MERRA2_FLX_EVAP	MERRA2_FLX_NIRDR
MERRA2_FLX_HFLUX	POPCORN_ASTERDEM	MERRA2_AER_SUANGSTR
MERRA2_ASM_TROPPB	MERRA2_AER_BCFLUXV	MERRA2_FLX_TLML
MERRA2_FLX_QSTAR	POPCORN_time_cyclic_daily_sin	MERRA2_AER_DUSCATAU
MERRA2_FLX_PBLH	POPCORN_distancetolandclass7	POPCORN_distancetolandclass12
MERRA2_AER_OCSCATAU	MERRA2_AER_TOTEXTTAU	POPCORN_distancetolandclass15
MERRA2_ASM_TROPPV	MERRA2_SURFACERH	MERRA2_FLX_RHOA
MERRA2_AER_BCEXTTAU	MERRA2_FLX_FRCLS	MERRA2_AER_DUEXTT25
MERRA2_ASM_T10M	MERRA2_ASM_TS	MERRA2_FLX_SPEED
MERRA2_AER_BCANGSTR	MERRA2_AER_DUSCAT25	MERRA2_AER_OCFLUXU
MERRA2_CTMCONST_FRLANDICE	MERRA2_AER_DUCMASS25	MERRA2_AER_OCEXTTAU
MERRA2_FLX_FRCAN	MERRA2_ASMCONST_FRLAND	MERRA2_AER_SSCMASS
MERRA2_AER_TOTSCATAU	MERRA2_AER_BCCMASS	MERRA2_CTMCONST_FRACI
MERRA2_AER_DUSMASS25	POPCORN_distancetolandclass16	POPCORN_CALIOP_MASK_AOD_90_Percent_Below
POPCORN_time_cyclic_daily_cos	POPCORN_distancetolandclass4	MERRA2_AER_DUANGSTR
MERRA2_FLX_SPEEDMAX	MERRA2_CTMCONST_FRLAND	MERRA2_FLX_HLML
MERRA2_AER_DUFLUXV	MERRA2_AER_OCANGSTR	MERRA2_FLX_TAUY
MERRA2_FLX_FRCCN	MERRA2_PM25	MERRA2_ASMCONST_FRLAKE
POPCORN_distancetolandclass8	MERRA2_AER_SSFLUXV	MERRA2_AER_SUFLUXU
MERRA2_FLX_CDQ	POPCORN_distancetolandclass13	MERRA2_FLX_TSTAR
MERRA2_FLX_CN	MERRA2_ASM_V50M	MERRA2_AER_SSSCATAU
MERRA2_FLX_QSH	MERRA2_FLX_Z0H	MERRA2_ASM_PS
MERRA2_AER_SSEXTTAU	MERRA2_FLX_TCZPBL	MERRA2_AER_OCSMASS
MERRA2_FLX_TSH	POPCORN_distancetolandclass3	MERRA2_SURFACEELEVATION
MERRA2_ASM_TROPQ	MERRA2_FLX_CDH	MERRA2_FLX_PGENTOT
MERRA2_ASM_U10M	MERRA2_FLX_ULML	MERRA2_ASM_TOX
MERRA2_AER_DMSCMASS	POPCORN_distancetolandclass1	POPCORN_distancetolandclass14
MERRA2_FLX_TAUX	MERRA2_ASMCONST_FRLANDICE	MERRA2_AER_SUSCATAU
MERRA2_AER_DUFLUXU	POPCORN_distancetolandclass10	MERRA2_FLX_PREVTOT
MERRA2_CTMCONST_FROCEAN	MERRA2_ASM_TQL	MERRA2_ASM_U2M
MERRA2_ASM_DISPH	MERRA2_FLX_PRECTOT	MERRA2_AER_SO2SMASS
MERRA2_FLX_CDM	MERRA2_FLX_Z0M	MERRA2_ASM_windspeed
POPCORN_distancetolandclass11	MERRA2_FLX_DISPH	MERRA2_AER_OCFLUXV
MERRA2_FLX_PRECTOTCORR	MERRA2_ASM_TROPPT	MERRA2_FLX_PRECLSC
MERRA2_FLX_BSTAR	MERRA2_ASM_TO3	POPCORN_CALIOP_MASK_AOD_63_Percent_Below
MERRA2_FLX_PRECCON	MERRA2_ASM_TQI	MERRA2_ASMCONST_FROCEAN
MERRA2_CTMCONST_PHIS	POPCORN_distancetolandclass5	MERRA2_CTMCONST_FRLAKE
MERRA2_FLX_TAUGWX	MERRA2_FLX_PRECANV	MERRA2_ASM_V2M
MERRA2_ASMCONST_PHIS	MERRA2_FLX_NIRDF	POPCORN_distancetolandclass9
MERRA2_ASM_SLP	POPCORN_BlackMarble	POPCORN_distancetoroad_upwind
MERRA2_AER_SSANGSTR	MERRA2_FLX_VLML	MERRA2_AER_SSSCAT25
MERRA2_ASM_winddirection	MERRA2_FLX_TAUGWY	MERRA2_AER_SSFLUXU
MERRA2_AER_SUEXTTAU	MERRA2_ASM_V10M	MERRA2_AER_SSCMASS25
MERRA2_FLX_PRECSNO	MERRA2_AER_SSEXTT25	MERRA2_AER_DMSSMASS
MERRA2_FLX_RISFC	MERRA2_AER_SSSMASS	MERRA2_ASM_U50M
MERRA2_FLX_FRSEAICE		

Table 1. List of input variables used in our model ordered by SHAP value (from left to right and from top to bottom).

References

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al.: Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, The Lancet, 389, 1907–1918, 2017.

Geng, G., Zhang, Q., Martin, R., Donkelaar, A., Huo, H., CHE, H., Lin, J., and He, H.: Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model, Remote Sensing of Environment, 166, https://doi.org/10.1016/j.rse.2015.05.016, 2015.

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., et al.: Global estimates and long-term trends of fine particulate matter concentrations (1998–2018), Environmental Science & Technology, 54, 7879–7890, 2020.

Handschuh, J., Erbertseder, T., and Baier, F.: Systematic Evaluation of Four Satellite AOD Datasets for Estimating PM2.5 Using a Random Forest Approach, Remote Sensing, 15, https://doi.org/10.3390/rs15082064, 2023.

Health Effects Institute: State of global air 2019, 2019.

Ibrahim, S., Landa, M., Pešek, O., Brodský, L., and Halounová, L.: Machine Learning-Based Approach Using Open Data to Estimate PM2.5 over Europe, Remote Sensing, 14, https://doi.org/10.3390/rs14143392, 2022.

Lipponen, A., Kolehmainen, V., Kolmonen, P., Kukkurainen, A., Mielonen, T., Sabater, N., Sogacheva, L., Virtanen, T. H., and Arola, A.: Model-enforced post-process correction of satellite aerosol retrievals, Atmospheric Measurement Techniques, 14, 2981–2992, 2021.

Lipponen, A., Reinvall, J., Väisänen, A., Taskinen, H., Lähivaara, T., Sogacheva, L., Kolmonen, P., Lehtinen, K., Arola, A., and Kolehmainen, V.: Deep-learning-based post-process correction of the aerosol parameters in the high-resolution Sentinel-3 Level-2 Synergy product, Atmospheric Measurement Techniques, 15, 895–914, 2022.

Lundberg, S. M. and Lee, S.: A unified approach to interpreting model predictions, CoRR, abs/1705.07874, http://arxiv.org/abs/1705.07874, 2017.

Pope, C. A. I. and Dockery, D. W.: Health Effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air & Waste Management Association, 56, 709–742, https://doi.org/10.1080/10473289.2006.10464485, 2006.

Schneider, R., Vicedo-Cabrera, A. M., Sera, F., Masselot, P., Stafoggia, M., de Hoogh, K., Kloog, I., Reis, S., Vieno, M., and Gasparrini, A.: A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain, Remote Sensing, 12, https://doi.org/10.3390/rs12223803, 2020.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., and Schwartz, J.: Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model, Environment International, 124, 170–179, https://doi.org/https://doi.org/10.1016/j.envint.2019.01.016, 2019.

Tang, D., Zhan, Y., and Yang, F.: A review of machine learning for modeling air quality: Overlooked but important issues, Atmospheric Research, 300, 107 261, https://doi.org/https://doi.org/10.1016/j.atmosres.2024.107261, 2024.

Taskinen, H., Väisänen, A., Hatakka, L., Virtanen, T. H., Lähivaara, T., Arola, A., Kolehmainen, V., and Lipponen, A.: High-Resolution Post-Process Corrected Satellite AOD, Geophysical Research Letters, 49, e2022GL099 733, 2022.

van Donkelaar, A., Martin, R. V., Spurr, R. J., Drury, E., Remer, L. A., Levy, R. C., and Wang, J.: Optimal estimation for global ground-level fine particulate matter concentrations, Journal of Geophysical Research: Atmospheres, 118, 5621–5636, 2013.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M.: Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors, Environmental Science & Technology, 50, 3762–3772, https://doi.org/10.1021/acs.est.5b05833, 2016.

World Health Organization: New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution, https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution, [Online; accessed 12-April-2023], 2021.

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K.: Modeling Tabular data using Conditional GAN, 2019.

Zhang, H. and Kondragunta, S.: Daily and Hourly Surface PM2.5 Estimation From Satellite AOD, Earth and Space Science, 8, e2020EA001 599, https://doi.org/https://doi.org/10.1029/2020EA001599, e2020EA001599 2020EA001599, 2021.