Review 1:

The manuscript describes the application of a supervised machine learning algorithm (lightGBM) for the retrieval of AOD, fAOD, and cAOD over Europe. However, the method presented for aerosol retrieval is not new, and I have some main concerns about this study. Firstly, the claimed high-resolution (0.1 degree) aerosol product is questionable. Secondly, the validation of the proposed model shows severe overfitting.

Thank you for taking the time to review our manuscript. We appreciate your feedback and comments. We understand your concerns regarding the novelty of the method presented for aerosol retrieval and the validation of the proposed model, however, our goal in this paper is to provide improved fAOD and cAOD products over Europe. To that end our methods focus on addressing limitations related to coarse spatial resolutions or the large data gaps of current products, rather than introducing a completely new machine learning framework. The quality of fAOD and cAOD products are key to better model particulate matter of different diameter ranges (e.g., PM2.5 and PM10), which is needed for future epidemiological studies. To make the lightGBM method more suitable for our goal, we developed three techniques, including distance weighted loss function, minimum directional distance, and white-noise data augment to improve the models. As for the other two main concerns (spatial resolution and validation), we have revised our manuscript to avoid any misunderstanding and we have done some additional sensitivity analyses to show the robustness of our predicted aerosol dataset as detailed below.

Major concerns:

1. The study claims that their AOD products were generated at a spatial resolution of 0.1 degrees. However, it should be noted that the key input variable, MAIAC AOD, only has a spatial resolution of 1km, and was eventually excluded from the models. Other variables used in the study have a lower spatial resolution than 0.1 degrees. Therefore, it is questionable whether the resulting product is truly a 0.1 degree product.

Response: Thank you for your review and for raising these concerns. We apologize for the confusion regarding the resolution of our inputs. In the submitted manuscript we inadvertently omitted specifying the resolution of 0.1 degrees of the ERA5-land surface variables, which contributed to the resolution of the final product. In the revised manuscript we included the description of the variable inputs both from ERA5 and ERA5-land (Lines 138-160), and added the information about the resolution in Table S2.

2. Additionally, Figure 11 shows that the developed AOD (B1) does not provide better details than the CAMS AOD (0.75 degrees) and MERRA-2 AOD (0.625 degrees * 0.5 degrees).

Response: Regarding the differences between QML AOD and CAMS and MERRA-2 AOD, we recognize that it is difficult to highlight the differences between products in continental maps showing 18-year averages, as originally shown in Figure 11. A zoom of Figure 11 focusing on Italy and Spain, where MAIAC AOD has fewer missing gaps, shows that the QML AOD exhibits substantially more details compared to CAMS AOD and MERRA-2 AOD (Figure R1). It is remarkable to verify that the spatial variability of the QML AOD is close to that of the MAIAC AOD, despite that MAIAC is omitted by the model. While MAIAC AOD appears to show slightly more details, likely due to its higher 1km resolution, the fewer observations available in MAIAC may also partly contribute to these differences. Figure R1A illustrates that the limited observations in MAIAC AOD can introduce some biases when calculating 18-year averages for comparison, as opposed to solely comparing the overlapping available period of both MAIAC and AERONET data. These biases do not exist in QML AOD, because QML AOD have full coverage in 18 years.

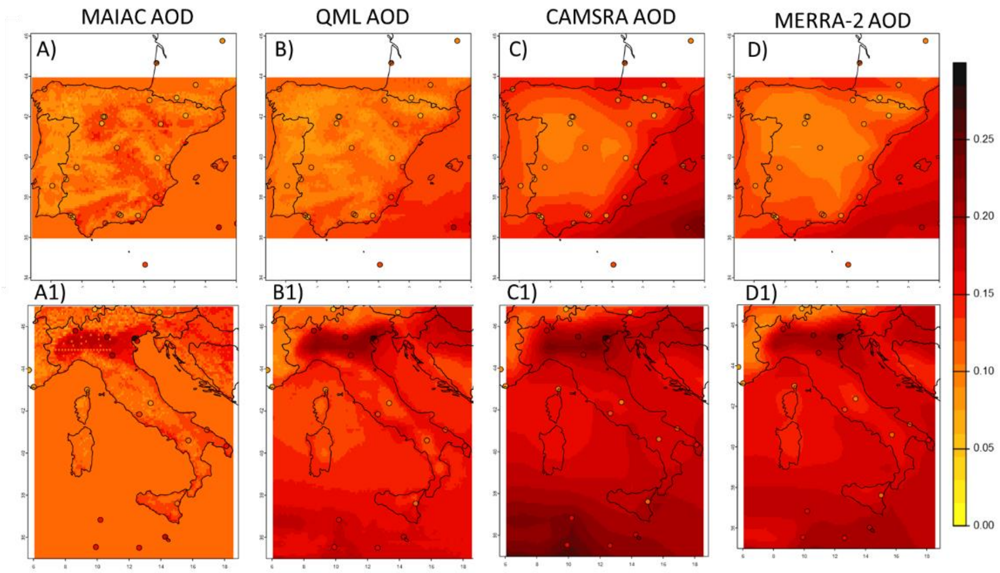

Figure R1. 18-year averages (2003-2020) of different AOD in Spain(A-D) and Italy(A1-D1). We illustrate the results for these regions because MAIAC AOD has few missing gaps in southern Europe (e.g. low cloudiness). Notedly, the CAMS AOD and MERRA-2 AOD here have already been spatially interpolated to 0.1 degrees.

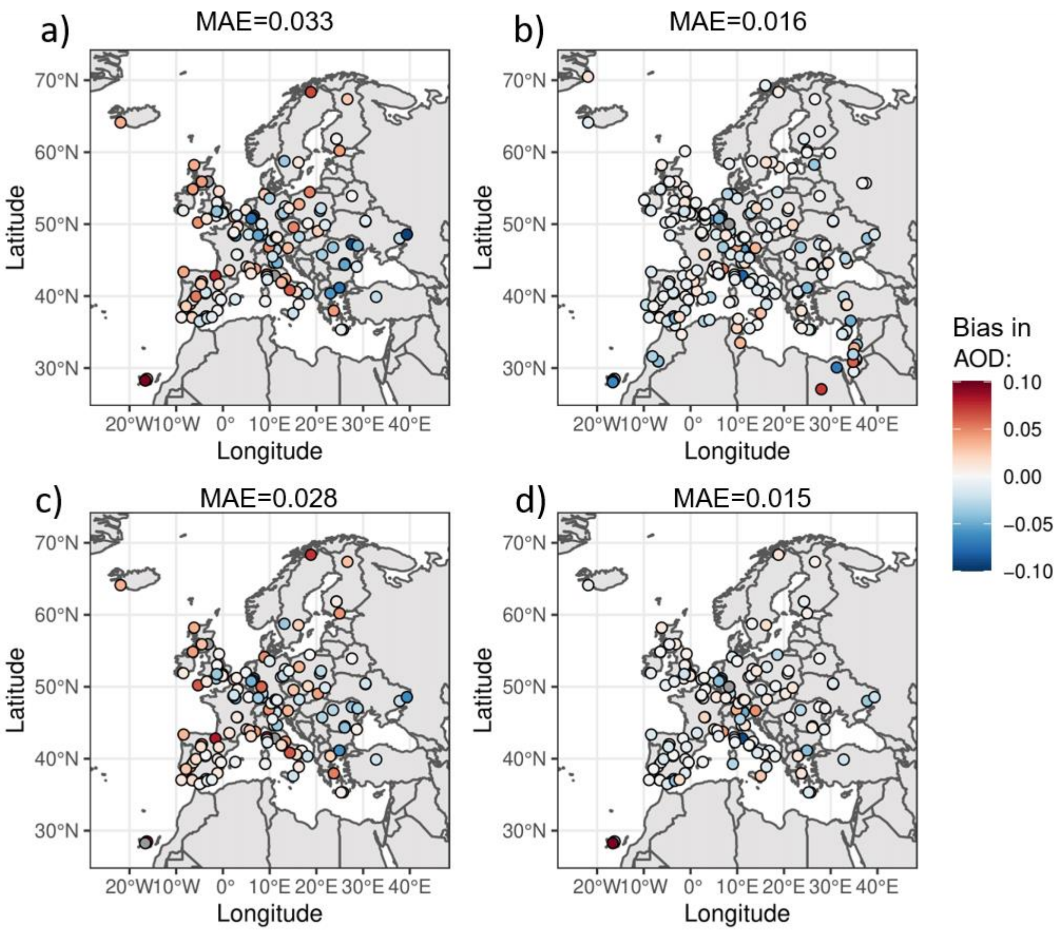

Figure R1A. The bias between MAIAC and AERONET AOD 18-year (2003-2020) averages when AERONET data is available(a); the bias between QML and AERONET AOD 18-year averages when AERONET data is available (b); the bias between MAIAC and AERONET AOD 18-year averages only when both MAIAC and AERONET data are available (c); bias between QML and AERONET 18-year averages only when both MAIAC and AERONET data are available (d). MAE is the mean absolute error.

To further validate our product, Figure 4 in the manuscript clearly demonstrates that QML AOD exhibits a better fit with AERONET AOD at the daily level compared to other products, as evidenced by the higher spatial cross-validation R square (0.68 compared to 0.36-0.52) and lower NRMSE (21.25% compared to 31.24%-32.96%).

Additionally, we have observed that most products show a good fit with AERONET AOD when calculating the 18-year averages (see Figure R2). However, the differences in performance among the four products become more pronounced when comparing AOD products with daily AERONET AOD. This suggests that while the predictions in 18-year averages may appear quite similar, their daily estimations can vary significantly. For instance, Figure R3 (randomly selected day) demonstrates that the differences among the four products become more noticeable at the daily resolution. While there is a general similarity in the patterns across these products, there are substantial differences in certain locations. Notably, QML AOD continues to exhibit more details than CAMS AOD and MERRA-2 AOD even at the daily scale.

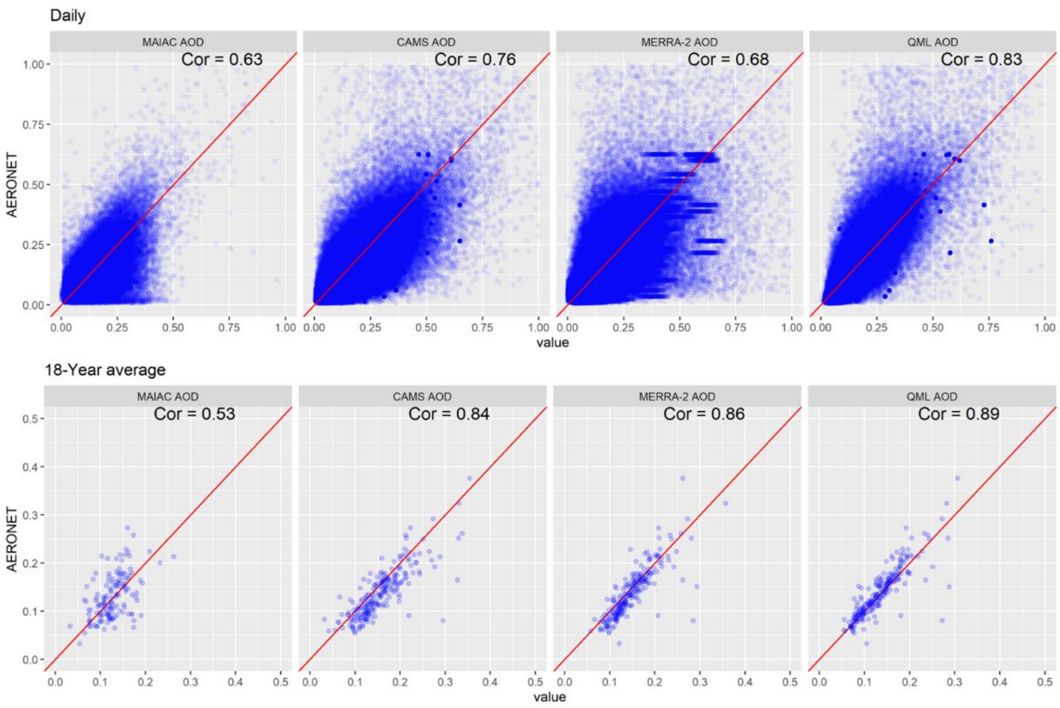

Figure R2. The scatter plots of different AOD products against daily AERONET data and 18-year averaged AERONET data. Red line is y=x.

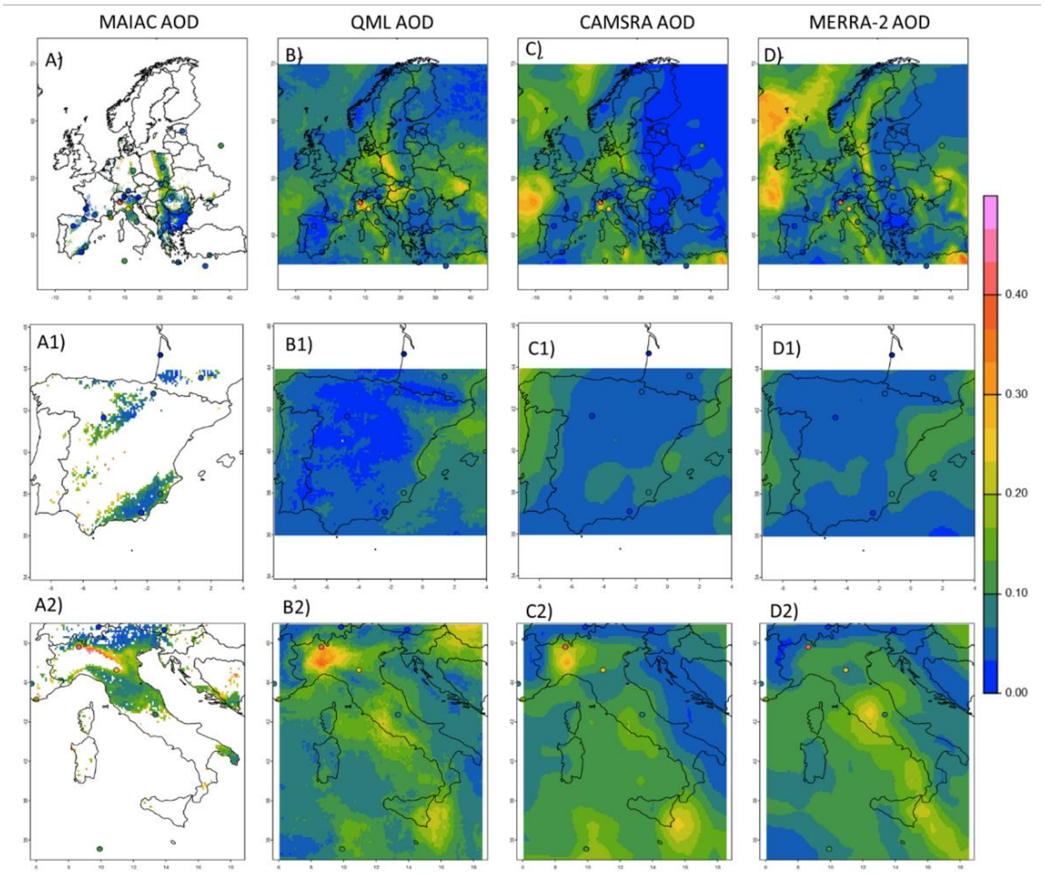

Figure R3. the different AOD in Europe (A-D), Spain (A1-D1) and Italy (A2-D2) in 2016-01-01.

3. Based on the input variables listed in Table S2, it appears that only the CAMS reanalysis data provides information related to aerosol size. The study seems just used the lightGBM algorithm to correct the CAMS-based fAOD and cAOD using meteorological data.

Response: Thank you for raising this question. Note that Table S2 has been updated to show the variables that come from ERA5 (0.25 degrees) and ERA5-Land (0.1 degrees). It is important to note that we do not use any explicit size information from CAMSRA. We use the separate contributions to AOD at 550 nm of "black carbon aerosol," "dust aerosol," "organic matter aerosol," "sea salt aerosol," and "sulphate aerosol, which provide size information implicitly.

We developed fAOD and cAOD products specifically to better model the particulate matter of different diameter ranges (e.g., PM2.5 and PM10), which is useful for epidemiological studies. Our sensitivity analysis indeed shows that the correlation between PM2.5 and QML fAOD is stronger than with other CAMSRA composition products (Table R1), and similar results are seen with PM10 and PMcoarse.

Table R1. The spearman correlation between different-size ground-level particulate matter (PM10, PM2.5, PMcoarse) with QML AOD products and CAMSRA composition aerosol products.

Correlation	QML AOD	QML fAOD	QML cAOD	CAMSRA Sea Salt	CAMSRA Sulphate	CAMSRA Organic Matter	CAMSRA Dust	CAMSRA Black Carbon
PM2.5	0.40	**0.45**	0.02	-0.31	0.16	0.19	0.11	0.12
PM10	**0.41**	0.37	0.12	-0.29	0.17	0.17	0.16	0.14
PM coarse	0.15	0.06	**0.21**	0.13	0.12	0.07	0.19	0.12

The CAMSRA compositional data contributes to the final model. However, their contribution is not more significant than other data sources. The importance score results from the lightGBM models in Figure R4 (or Figure A1 in supplementary) reveal that the contributions of each variable in the top 20 are relatively similar, ranging from around 2.8% to 4%. It is important to recognize that meteorological factors derived from ERA5 or ERA5_land, such as boundary layer dissipation (BLD) and height (BLH), humidity (RH), surface pressure (SP), and wind speed, also play a significant role in the modeling of fAOD (fine mode aerosol optical depth) and cAOD (coarse mode aerosol optical depth).

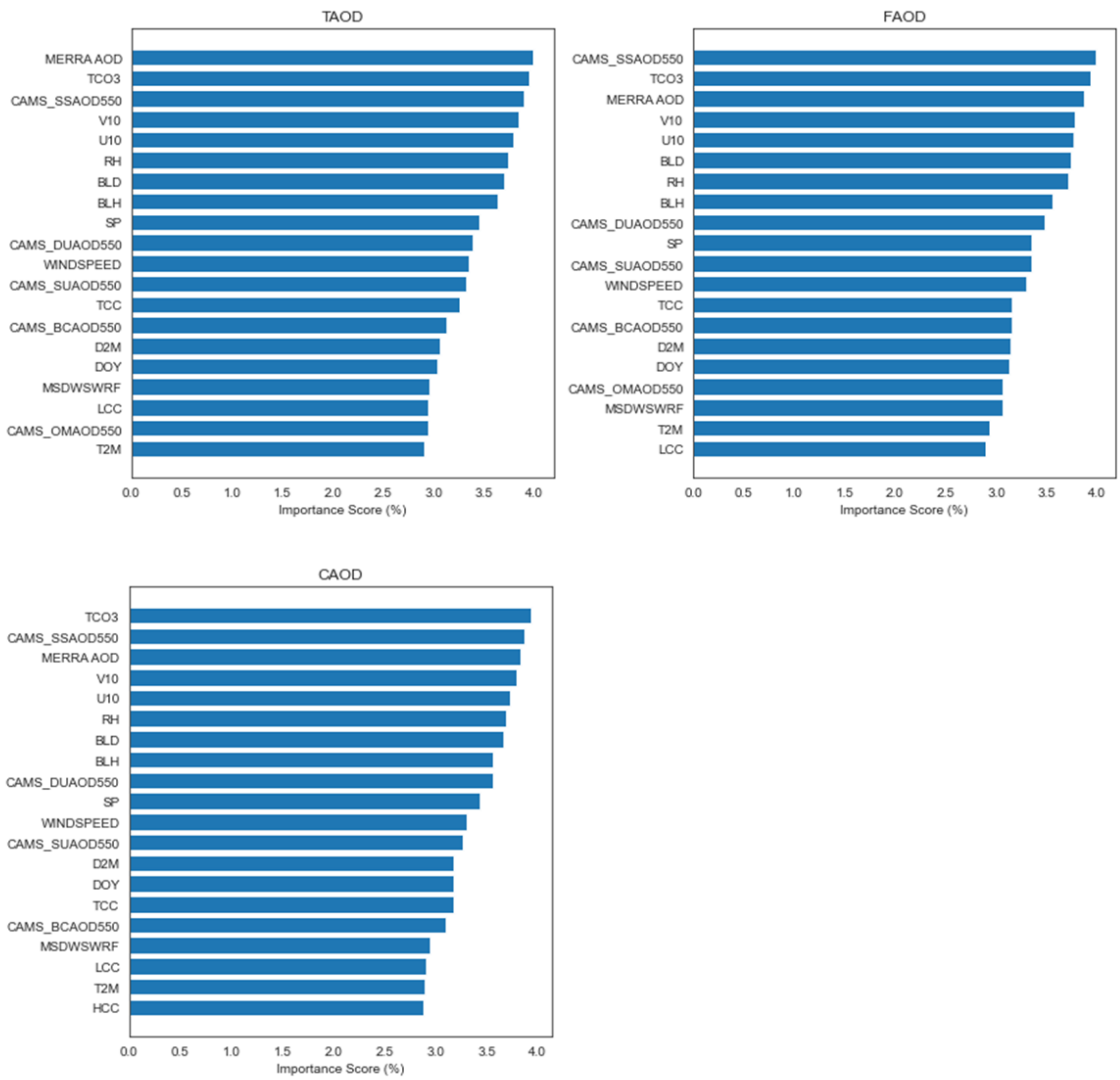

Figure R4. the top 20 important feature plots of AOD, fAOD and cAOD model, importance scores here representing the proportion of model contribution for each feature. The full name of variable is as following:

Short name	Source	Long name
MERRA_AOD	MERRA-2	MERRA2 aerosol optical depth 550nm
TCO3	ERA5	total column ozone
CAMS_SSAOD550	CAMSRA	sea salt aerosol optical depth 550nm
U10	ERA5_land	10m u component of wind
V10	ERA5_land	10m v component of wind
RH	ERA5_land	Surface relatively humidity
BLD	ERA5	boundary layer dissipation
BLH	ERA5	boundary layer height
SP	ERA5_land	surface pressure
CAMS_DUAOD550	CAMSRA	dust aerosol optical depth 550nm
WINDSPEED	ERA5_land	10m v component of wind
CAMS_SUAOD550	CAMSRA	sulphate aerosol optical depth 550nm
TCC	ERA5	total cloud cover
LCC	ERA5	low cloud cover
CAMS_BCAOD550	CAMSRA	black carbon aerosol optical depth 550nm
D2M	ERA5	2m dewpoint temperature
YEAR	Time	year
DOY	Time	day of year
MSDWSWRF	ERA5_land	Surface solar radiation downwards
CAMS_OMAOD550	CAMSRA	organic matter aerosol optical depth 550nm

Short name	Source	Long name
T2M	ERA5	2m temperature
HCC	ERA5	high cloud cover

4. It is unclear how well the developed fAOD and cAOD models perform at locations where no AERONET data is available. It is also unclear whether the study used completely independent ground-based data to test the results, such as a test site that was not used in the training process. If the Table S3 intends to show this validation, but the R2 of fAOD decreased significantly from 0.68 to 0.56 in M3, suggesting that the model may have a severe issue with overfitting.

Response: Thank you for raising this point. We acknowledge that the description of techniques to improve the models was misleading or confusing the readers. We have revised the manuscript to clarify any misunderstanding (Line 180-187). We have done two validation processes: On the one hand, we randomly selected 70% of the sites as training data for the quantile lightGBM models, additional 20% of the sites were used to optimize the model, and the rest 10% sites were completely independent test data. On the other hand, we used 5-fold cross-validation to repeat the first process, in order to test the stability of all model configurations.

The R2=0.56 mentioned in the comment corresponds to the subgroup of top 1% farthest sites, with distances to their neighbors over 463km, which were mostly located at the edge of our domain. Considering your concerns regarding the performance of the model in regions with sparsely distributed stations, we divided the top 20% of sites that were farthest from their nearest neighbors, requiring distances of at least 130.5km, and trained the model using the remaining sites. Then we compared these results to random split 20% validation datasets. We have done these sensitive comparisons for tAOD, fAOD and cAOD in Figures R4, R5 and R6, respectively. It shows that the results of our models among top 20% of farthest sites is relatively similar or a bit lower than the results in 20% random sites. It indicates models are quite stable without any severe overfitting.

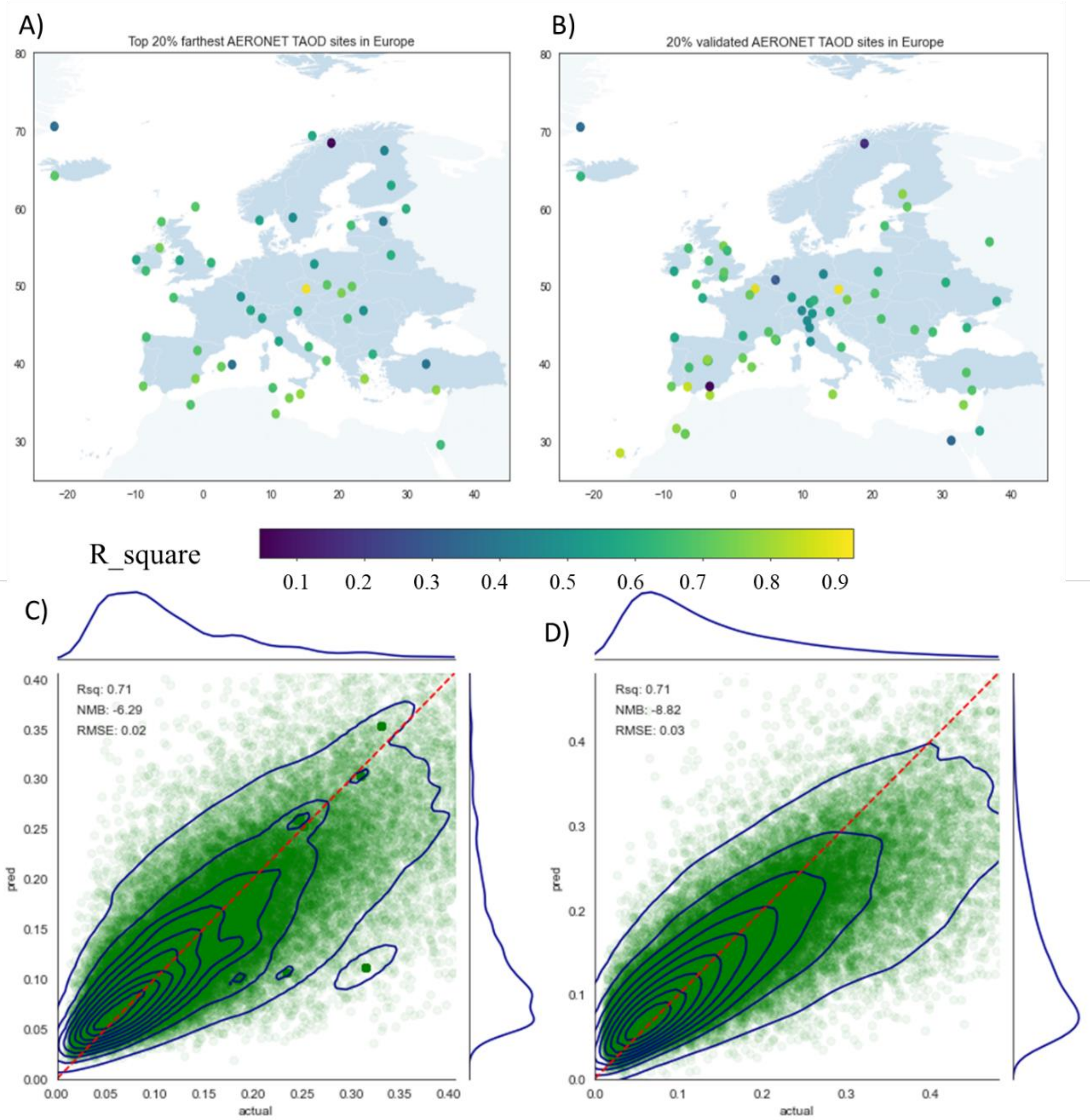

Figure R4. the out of sample R-square of AOD model in top 20% of farthest sites (A) and 20% random validation sites (B), their corresponding scatter plots (C and D).

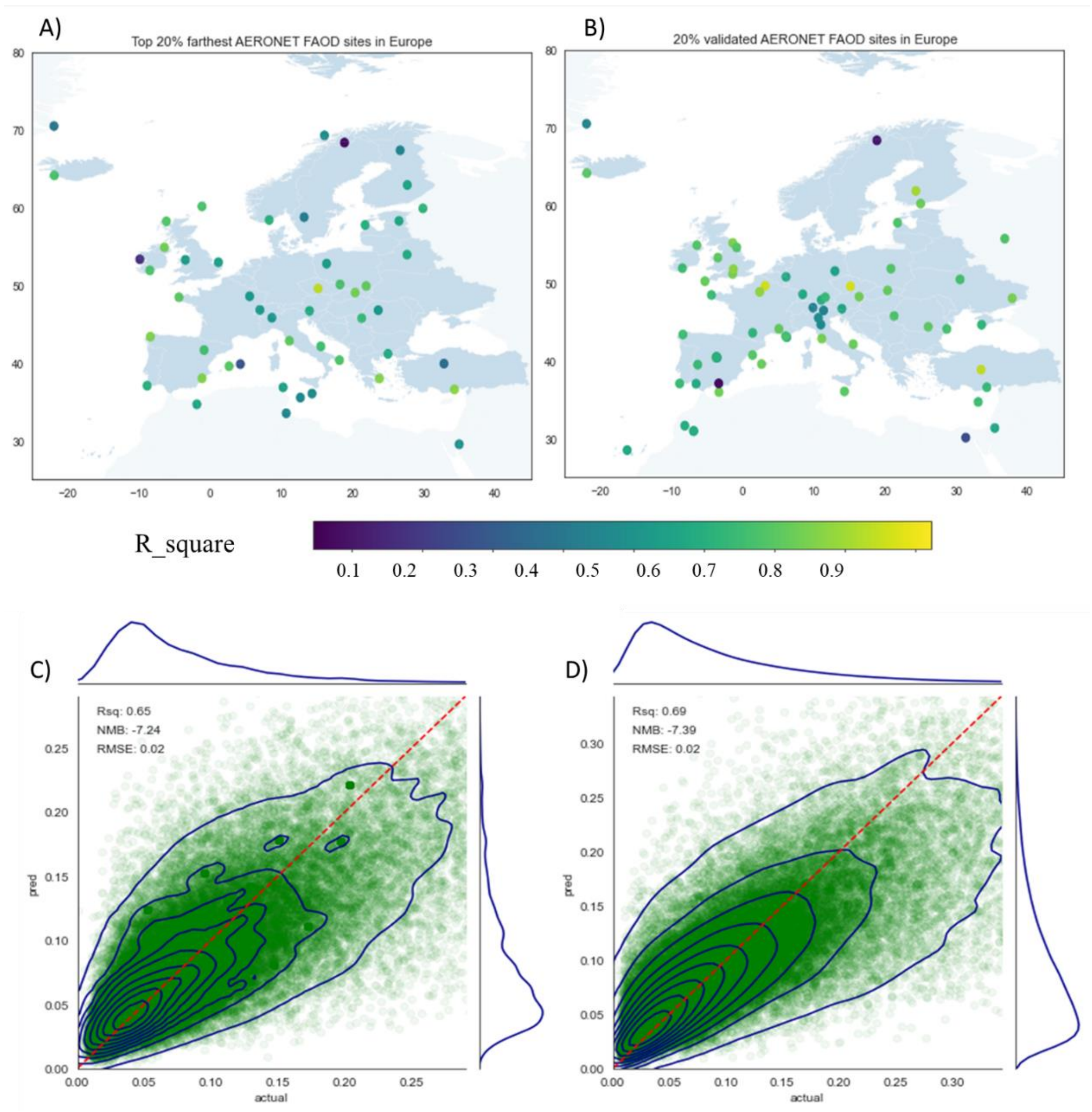

Figure R5. the out of sample R-square of fAOD model in top 20% of farthest sites (A) and 20% random validation sites (B), their corresponding scatter plots (C and D).

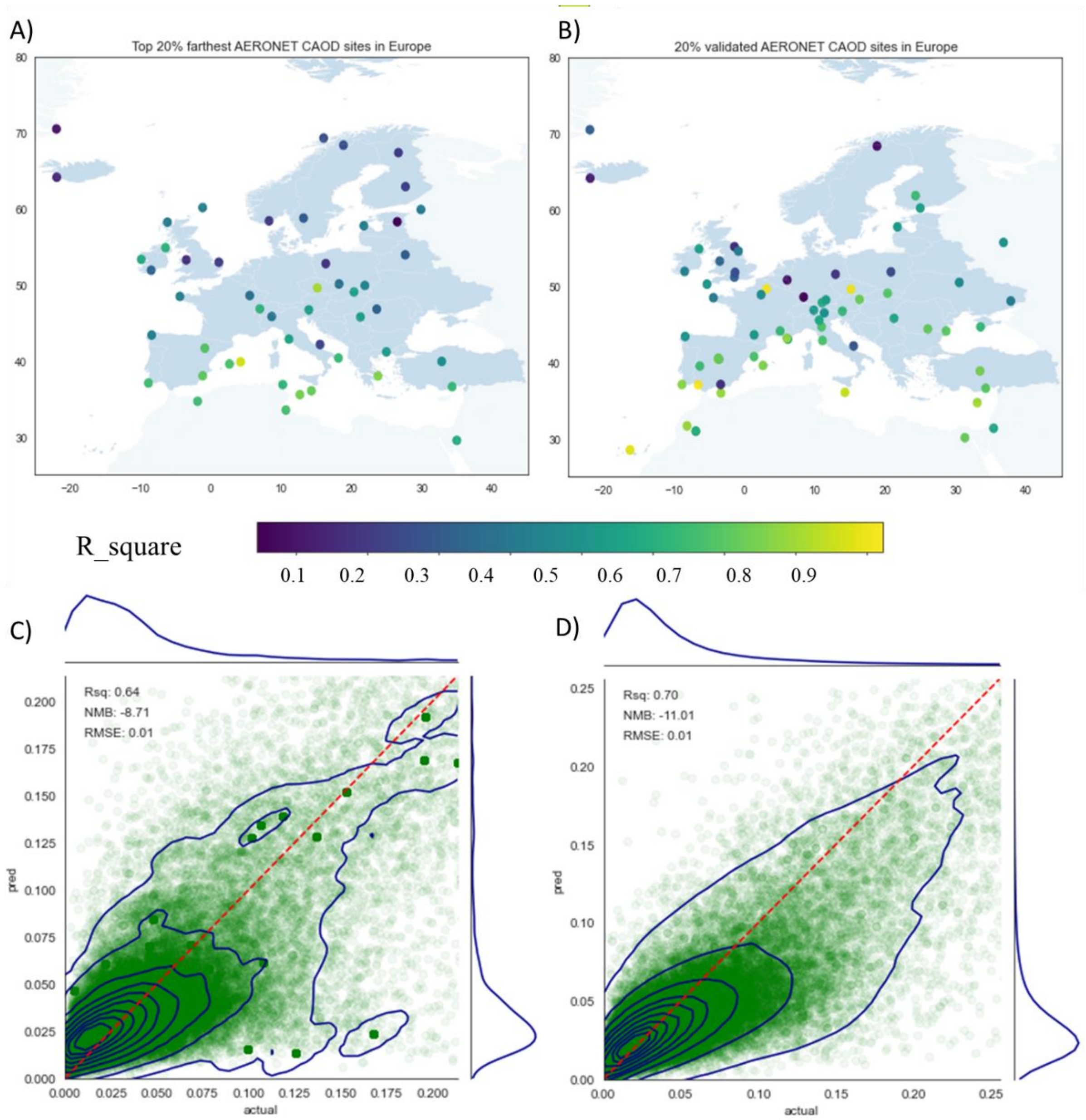

Figure R6. the performance of cAOD model in top 20% of farthest sites (left) and 20% random validation sites (right)

5. During the lightGBM-based training for fAOD and cAOD, the AERONET only provides data for fAOD and cAOD at 500nm. However, it is unclear how the model was trained to calculate fAOD and cAOD at 550nm, which is a crucial issue that the paper did not address.

Response: Thank you for bringing up this concern, we added a more detailed description (Line 104-115) on the procedure we followed to obtain the fAOD and cAOD at 550nm.

"To be comparable with the satellite and reanalysis data, the AERONET AOD data at 550 nm (AOD_{550}) was interpolated from the AOD_{500} (Gupta et al., 2020; Duarte and Duarte, 2020). The equation (1) used for this interpolation is as follows:

$$AOD_{550} = AOD_{500} * \left(\frac{550}{500}\right)^{-\propto^t} \quad (1)$$

where $\propto^t$ is the AERONET AOD Ångström exponent at 500nm, which is obtained from AERONET spectral deconvolution algorithm (SDA) output. Before obtaining the $fAOD_{550}$ and $cAOD_{550}$, we first transformed the Fine mode fraction at 550 nm (FMF_{550}) from the 500 nm (FMF_{500}) using the equation (2):

$$FMF_{550} = \frac{fAOD_{500} * (\frac{550}{500})^{-\propto^f}}{AOD_{500} * (\frac{550}{500})^{-\propto^t}} = FMF_{500} * (\frac{550}{500})^{\propto^t - \propto^f} \quad (2)$$

where $\propto^f$ is the AERONET fAOD Ångström exponent at 500nm. All of these parameters are available from AERONET SDA products. Finally, we obtained $fAOD_{550}$ and $cAOD_{550}$ by following the formula:

$$fAOD_{550} = AOD_{550} * FMF_{550} \quad (3)$$

$$cAOD_{550} = AOD_{550} * (1 - FMF_{550}) \quad (4)$$

”

Specific concerns:

1. In Figure 1, it is not clear how to use Boruta to select the variables.

We revised the description on the use of Boruta in Line 180-187. “For each iteration, the Boruta algorithm generates a new set of shadow variables by randomly permutating the values of each potential variable, and trains a random forest classifier on the original and shadow features. The importance score of each original feature is compared to the maximum importance score of its corresponding shadow features. If the original feature has an importance score that is significantly higher than the maximum importance of its corresponding shadow features, it is considered important. Then Boruta marks the important features and removes the shadow features associated with them, and repeats these steps until a predefined number of iterations (e.g., 50 iterations in our study) have been reached.”

2. The caption of the Figure 3 says “Spatial and temporal distribution of the median value of AERONet (a) AOD, (b) fAOD and (c) cAOD data”. It makes me confused how (a), (b) and (c) reveal the temporal information.

Revised, thanks: “Spatial distribution of the median value of AERONet (a) AOD, (b) fAOD and (c) cAOD data”.

3. AERONET in the figure caption is “AERONet”, but in the text is “AERONET”.

Revised.

4. Typing errors: P10, L285, (Levy et al., 2010; Xiao et al., 2016; Yan et al., 2022)).

Revised.