Reviewer 2:

This study presents a daily AOD data set over Europe over the period 2003-2020, which was derived by post-processing the current satellite and reanalysis products, based on Machine Learning method. The accuracy of the total AOD in this dataset has been greatly improved. At the same time, the dataset can provide additional fine/coarse AOD data, which are also relatively reliable and will be very helpful for particulate matter (PM) prediction. The dataset will be interesting for the scientific community. Therefore, I have some comments before it could be accepted for publication.

Response: Thank you for taking the time to review our study and provide your useful feedback. We appreciate your interest in our work and are happy to hear your thoughts and address any concerns you may have regarding the dataset we have presented. We believe that our dataset has the potential to be a valuable resource for the scientific community and look forward to discussing it with you further.

Major comments:

1. For the Route in the absence of satellite data, the spatial resolution of all input reanalysis of AOD data (e.g. MERRA-2, CAMS) is relatively coarse lower than 0.1 degrees, it is not appropriate to increase the spatial resolution of final AOD product to 0.1 degrees through interpolation, as simple interpolation cannot increase the AOD variation in spatial details. I think the spatial resolution of the final AOD product should not be higher than the maximum spatial resolution of one of input reanalysis data.

Response: Thank you for your review and for bringing up these concerns. We apologize for any confusion regarding the resolution of our inputs. Previously, we have not clearly described the variable inputs from ERA5 (0.25 degrees) and ERA5_land (0.1 degrees). We have made revisions to the description of the variable inputs from ERA5 and ERA5_land (Line 138-160), as well as added resolution information in Table S2. And some surface-related variables provided by ERA5-Land, like surface solar radiation downwards (MSDWSWRF), surface humidity (RH), surface wind speed (U10,V10, WINDSPEED), surface pressure (SP) are also top 20 contributed features in our final 0.1 degree resolution product (Figure R1).

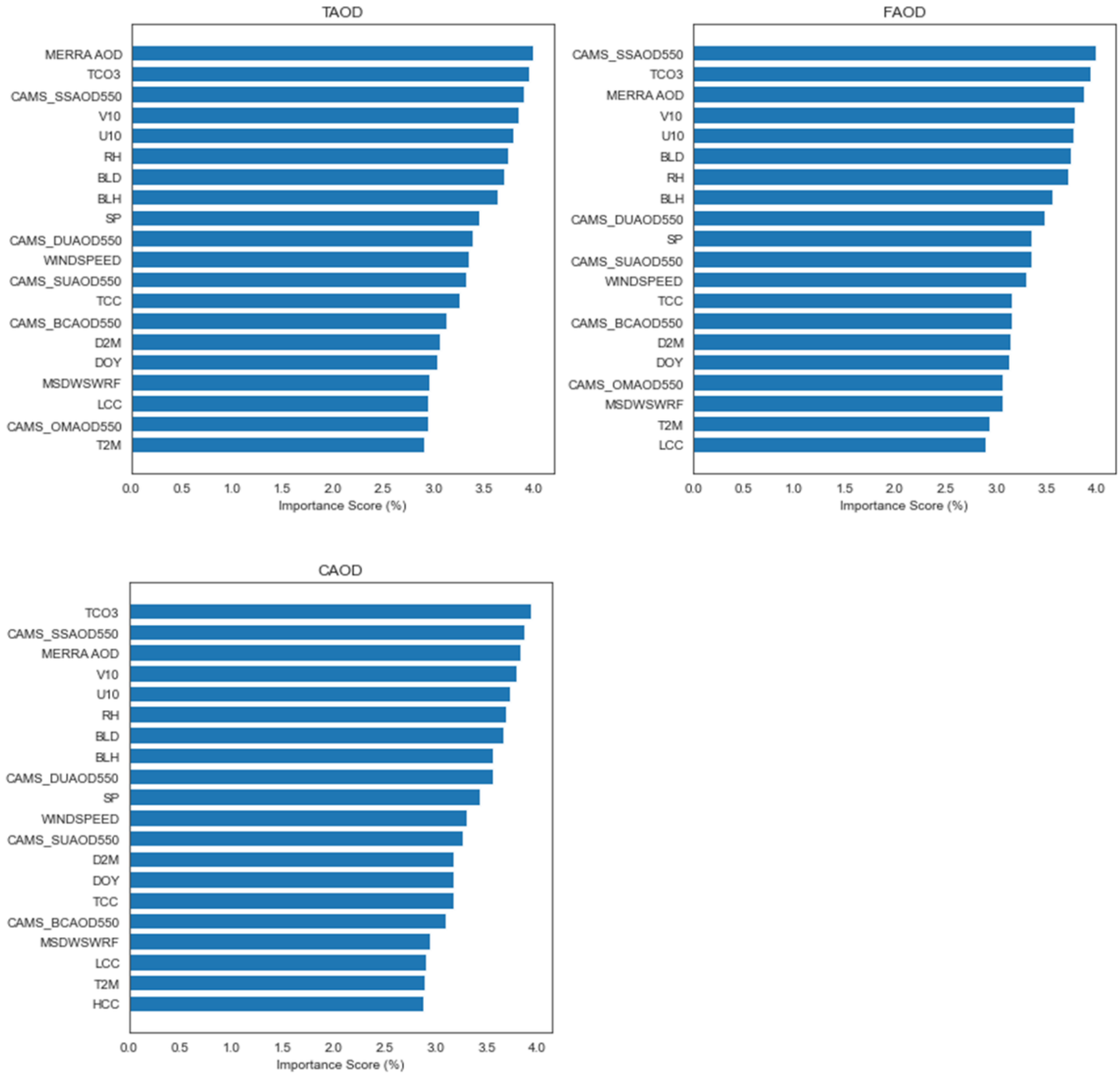

Figure R1. the top 20 important feature plots of AOD, fAOD and cAOD model, importance scores here representing the proportion of model contribution for each feature. The full name of variable is as following：

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

2. For the correction of total AOD, it can be understood that the information of AOD mainly comes from the AOD data of reanalysis product. But for obtaining fine AOD and coarse AOD, this study should clarify which input data plays a dominant role.

Response: Thank you for your question. We understand your concern about the dominant role of input data in obtaining fine AOD and coarse AOD. To address this, we have added a feature importance score plot as Figure S1 (the same as Figure R1 here) in the appendix. The plot lists the top 20 variables for the AOD, fAOD, and cAOD model, respectively. Interestingly, the top 10 inputs for these models are quite similar, but with a slightly different order. The top 10 inputs for all models are MERRA-2 AOD (MERRA AOD), total column of ozone from ERA5 (TCO3), CAMS sea salt AOD and dust AOD (CAMS_SSAOD550 and CAMS_DUAOD550), u and v component of wind (U10 and V10) from ERA5_land, boundary layer dissipation (BLD) and height (BLH) from ERA5, humidity (RH), and surface pressure (SP) from ERA5_land.

We found that the contributions of each variable in the top 20 were quite similar, ranging from around 2.8% to 4%. Therefore, we cannot identify a single dominant variable in the model.

3. I'm also curious, what would happen for QML AOD if two reanalysis datasets MERRA-2 and CAMS were not used as input data simultaneously?

Response: Thank you for your question. To address this concern, we conducted a sensitivity analysis in which we excluded both MERRA-2 and CAMS from our input datasets. Figure R2 showed a decrease in model performance for AOD, fAOD, and cAOD in the test sites, with the correlation coefficients decreasing from 0.71 to 0.47, from 0.69 to 0.45, and from 0.70 to 0.43, respectively. It suggested that air quality reanalysis data indeed contributes around one third to the information of the model, while other sources of data, such as meteorological data from ERA5 or surface data from ERA5 land, also contribute to some degree to the model. Figure R2 (a1-c1) also suggested that excluding both MERRA-2 and CAMS as input datasets could lead to the underestimation of the higher AOD values.

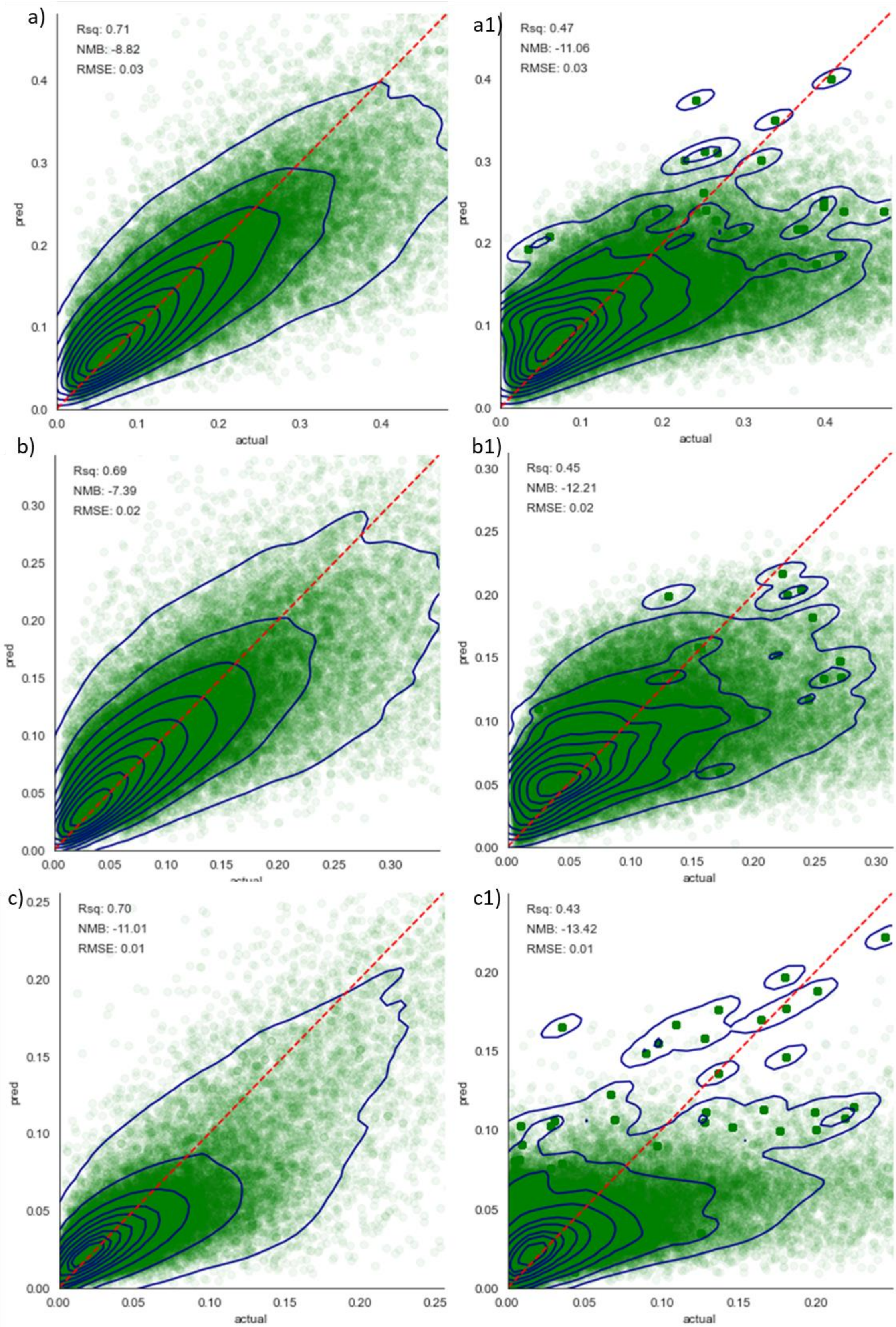

Figure R2. Comparison of the original models (a-c) and the model without MERRA-2 and CAMS data (a1-c1): AOD (a), fAOD (b) and cAOD (c).

Minor comments:

1. In section 2, this manuscript should introduce the basic information of PM data, as it was used in subsequent experiments.

Response: Done, We added in Line 161-167.

2. Line 105, how about fAOD and cAOD at 550nm was interpolated?

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:

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

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:

$$fAOD_{550} = AOD_{550} * FMF_{550} \quad (3)$$
$$cAOD_{550} = AOD_{550} * (1 - FMF_{550}) \quad (4)$$

"

Line 108, I believe the MODIS MAIAC data that the manuscript used is Collection 6 (C6), not v6.1, as the C6.1 product (MCD19A2) has not yet completed production.

Response: Thanks for correction. we revised the sentence according to the comment of the referee

3. Line 155, how is the MODIS 1km AOD product made to 0.1 degrees?

Response: We used the area-weighted averages within 1km grids. The 1km pixels within the 10km grid cells are averaged using a weighted average approach based on the fraction of the 1km pixel that falls within the 10km grid cell.

4. Line 269, the description is not clear about"Sat scenario"and "Non-Sat scenario", what do these two words mean? How to distinguish"Sat scenario"and "Non-Sat scenario"?

Response: We revised the text to clarify the issue (Line 294-300): "To account for the large fraction of missing values in satellite MAIAC AOD data (64%), we divided the validation results into two subgroups based on the availability of satellite MAIAC AOD data. The first subgroup, referred to as the "Sat scenario", included validation dates and sites where satellite MAIAC AOD data were available. The second subgroup, referred to as the "Non-Sat scenario", included validation dates and sites where satellite MAIAC AOD data were not available. This division was made in order to ensure comparability among the different products and to provide insights into the factors that limit the performance of the models."

5. Line 391, how was EE=±0.025 ±20 %/40 % determined? I think most literature uses 0.05 instead of 0.025.

Response: Thank you for bringing this to my attention. For total AOD, we also applied the standard EE=±0.05 ±20 %/40 % as most previous studies. The ±0.05 here represents expected error which includes measurement errors and other uncertainties that may come from instrument calibration or atmospheric conditions. Since the 0.05 value is around the 10th percentile of total AOD, and around ±5% for those larger AOD values (AOD >1), this EE can work well in most percentiles. However, fAOD and cAOD represent a subset of total AOD, and are thus generally less variable than total AOD. The 0.05 value is around the 40th percentile of fAOD, and it seems to be too big for small percentiles, while 0.025 is around the 10th percentile of fAOD. Thus, we used a narrower EE, like ±0.025 ±20 %/40 %, to ensure that the validation results are reliable in small percentiles, while still providing adequate evaluation in larger AOD values. In summary, the determination of the specific values for the EE used in this study was based on a consideration of the variability and distribution of AOD/fAOD/cAOD values in our dataset, and the will to balance the need for a reliable evaluation in small percentiles while still providing adequate evaluation in larger AOD values.