We would like to express appreciation to the reviewers for their insights and detailed review as well as for the suggested references. Our responses (in blue) for each comment (in black) are provided below.

Authors' response to RC2

Overview

The paper compares AOD products from the GEMS, AMI, and GOCI-II instruments aboard the GEO-KOMPSAT and two fusion products of the single instrument retrievals with AOD AERONET observations and with MODIS DT data. The fusion products are optimised to match the AERONET observations using a deep neural network (DNN) or a maximum likelihood estimate (MLE). The error analysis is detailed and distinguishes between different AOD loads, NDVI, observation time and aerosol types. The authors find that the GOCI-II retrievals have the lowest error of the one-instrument retrievals and that the fused products using DNN has overall the smallest errors.

General remarks

The paper is very detailed and provides a lot of quantitative information about the error of the evaluated AOD products. But, the paper should provide more scientific information to help the reader to understand or interpret short-comings of the products or the choices for the fusion approach. For example, the choice of NDVI (and not other candidate parameters) as predictor of error needs to be discussed in more detail. Likewise, the determination of the aerosol type should be better explained.

Thanks for your comments. We understand that the choice of variables for uncertainty stratification is not appropriately explained. Thus, reasoning of the choices is added in the Section 3.2,

"As shown in Fig. 5-7, retrieval error does not increase (or decrease) linearly. Therefore, merging AOD datasets using the same RMSE value for all pixels is not desirable. The statistical fusion method linearizes the error characteristics by categorizing potential error sources such as AOD values, aerosol types, NDVI values, and observation times. The potential error source variables are selected based on previous studies with the following logistics. First, AOD value itself and aerosol type is selected because as aerosol loading increases, aerosol model assumption affects retrieval performance. Complex aerosol mixture at high aerosol loading leads to high uncertainty and aerosol retrieval algorithms have distinct aerosol model assumptions. NDVI is selected as possible error source to represent surface condition. Different surface types have different surface reflectance and surface types differentiate by vegetation amount and types (Hsu et al., 2013). Observation time difference in GEO measurements leads to distinct optical path of observed radiance. Therefore, GEO satellite AOD products have diurnal error variations (Lim et al., 2019; Zhang et al., 2020; Fu et al., 2023; Cho et al., 2024). To deal with the uncertainty from this, observation time is selected as the possible error source.".

The two fusion products have smaller error against AERONET observations than the single-instrument retrievals. This is perhaps not surprising because the fusion approaches were designed to match the AERONET observations and a prior bias correct of the single-instruments retrieval was performed. Such a correction procedure could also be applied to the individual satellite data sets. So, it remains unclear if the added benefit of the fusion approach is the AERONET-based error correction or the synergistic benefits of the MLE or DNN based methods to merge the products.

Fig. 1 and Fig. 2 are showing validation results of original AOD products for before and after bias correction, and for fused products. Both figures clearly show that the AERONET-based bias correction improves AOD. But after fusion of bias corrected AOD products, the quality of AOD improves even more. Also, tables for validation results are shown in Table 1 and Table 2. Simplified version of the tables is also added in the manuscript as “Table 5 and Table 6”.

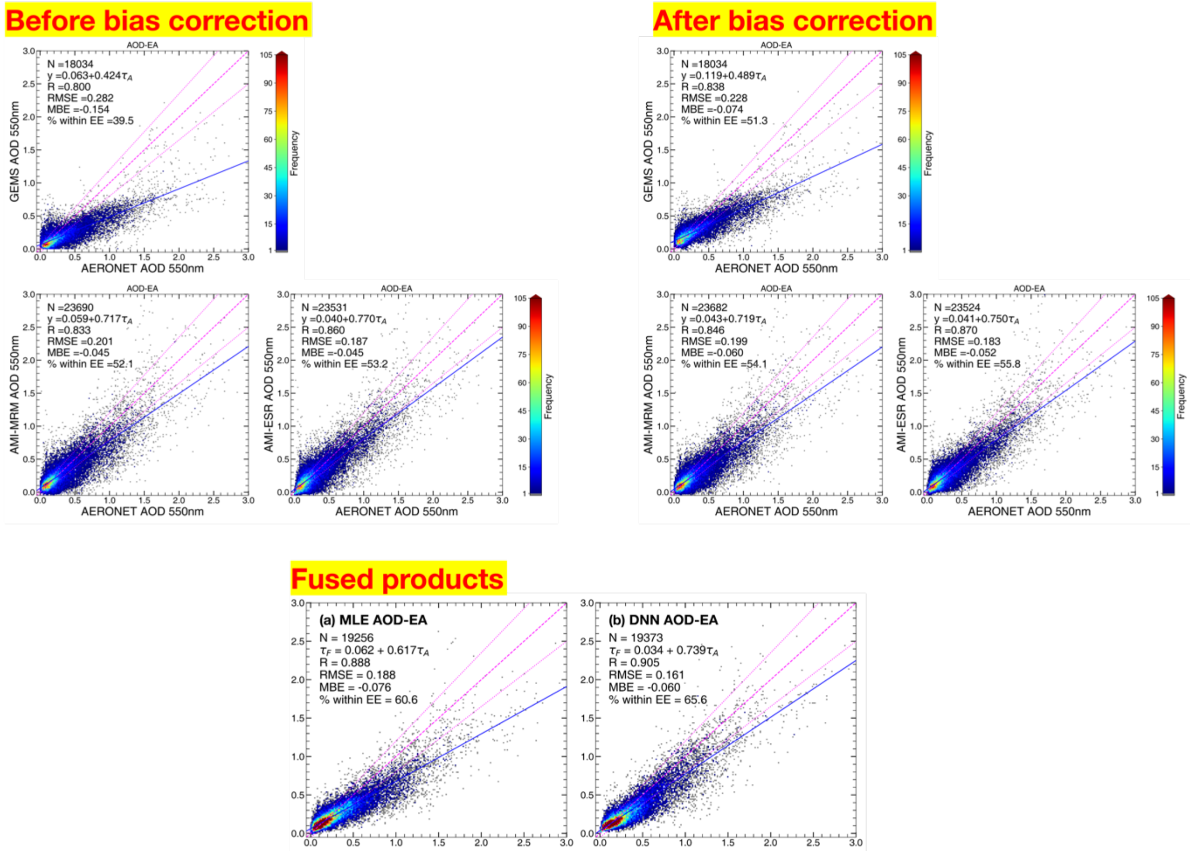

Figure RC2.1 Validation results of AOD products within -EA region. Results of original AOD products before, after bias correction, and fused AOD products are shown. AOD products from November 2022 to April 2023 are used. The number of collocated points (N), linear regression equations, Pearson’s correlation coefficient (R), root mean squared errors (RMSE), mean bias errors (MBE), and percentage within the expected error envelope (% within EE; EE: ±(0.05+0.15 τ_A)) is shown. Dashed line and dotted lines indicate one-to-one line and expected error envelopes. Blue line indicates linear regression line of the satellite AOD and AERONET AOD.

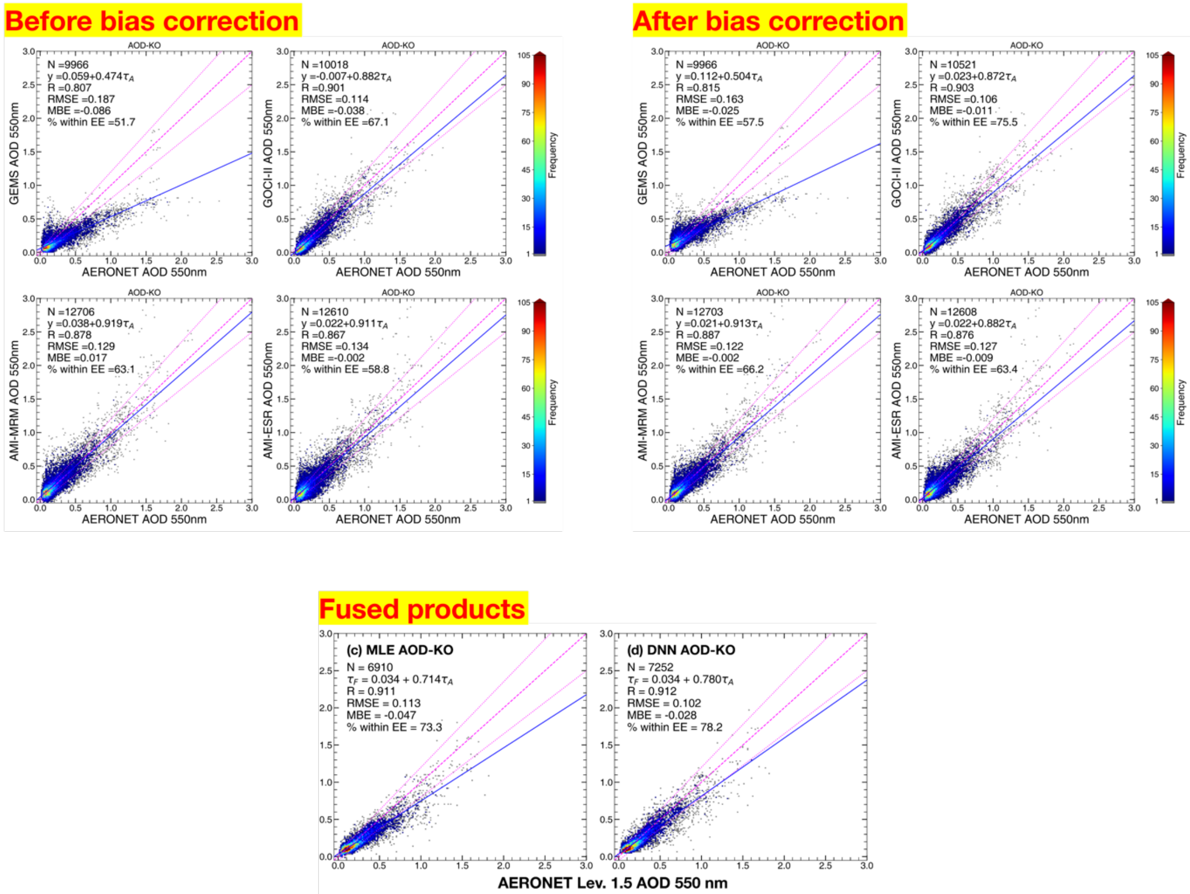

Figure RC2.2 As for Fig. RC2.1, but for -KO region.

Table 1: Validation statistics of original AOD products and fused AOD products in -EA region. Validation period is from November 2022 to April 2023.

	Before bias correction			Fused AOD	
	GEMS	AMI-MRM	AMI-ESR	MLE AOD	DNN AOD
R	0.800	0.834	0.860	0.888	0.905
RMSE	0.287	0.201	0.187	-0.188	0.161
MBE	-0.154	-0.045	-0.045	-0.076	-0.060
% within EE	39.5	52.1	53.3	60.6	65.6
	After bias correction				
R	0.838	0.846	0.870		
RMSE	0.228	0.199	0.183		
MBE	-0.074	-0.060	-0.052		
% within EE	51.3	54.1	55.8		

Table 2: Validation statistics of original AOD products and fused AOD products in -KO region.

	Before bias correction				Fused AOD	
	GEMS	AMI-MRM	AMI-ESR	GOCI-II	MLE AOD	DNN AOD
R	0.807	0.878	0.867	0.901	0.911	0.912
RMSE	0.187	0.129	0.129	0.114	0.113	0.102
MLE	-0.086	0.017	-0.002	-0.038	-0.047	-0.028
% within EE	51.7	63.1	58.8	67.1	73.3	78.2

	After bias correction				
R	0.815	0.887	0.876	0.903	
RMSE	0.163	0.122	0.127	0.106	
MLE	-0.025	-0.025	-0.009	-0.011	
% within EE	57.5	66.2	63.4	75.5	

The paper uses a lot of acronyms for different versions of the retrievals and it is difficult for the reader to follow. For improved readability I suggest 1) to spell out more of the acronyms in the figure captions, 2) to add a table that that summarises the data sets and 3) to add a table that summarised the error measures (bias, RMSE etc) for all considered single or fused products to give a better overview of the accuracy.

For 1), we revised figure captions. For 2), additional information of each aerosol products are added in Table 1. For 3), tables comparing the original AOD products and the fused AOD products are added as "Table 5 and Table 6". For the consistency of validation standard, the same level and period of AERONET dataset that is used for Fig. 8 is also used. Also, validation is done for two separate regions for -EA and -KO.

Specific comments:

L 140 Please provide more detail on the aerosol type classification. Why is GEMS not affected by the misclassification of the type?

Details of type classification of AMI and GOCI-II are further explained at the end of the Section 2.2.1 as "YAER algorithm first retrieves AODs at all wavelengths within UV-NIR range and converted to 550 nm for all aerosol types. Then, aerosol type that shows minimum variance at 550 nm are selected aerosol type for the corresponding inversion pixel.".

We understand the misleading description. All the aerosol retrieval algorithms are affected by misclassification. However, the wide wavelength coverage of AMI and GOCI-II are "more" sensitive to errors from aerosol size, while GEMS are "more" sensitive to errors from misclassification of absorbing/scattering aerosol types. So what we meant was that GEMS is less sensitive to error from misclassification of aerosol types that are in different size. Misleading expressions are revised in the manuscript.

L 159 How is the aerosol type derived?

Following explanation is added to the paragraph:

"Aerosol type is selected with UV aerosol index (UVAI) and VIS aerosol index (VISAI). The algorithm assigns NA type to pixels with low UVAI values. The other pixels are separated into highly absorbing fine (HAF) type and DU type according to the VISAI values."

L 164 Please comment on the differences and biases between AERONET version 3 level 2 and level 1.5

The comment on the differences on the AERONET data levels is added as

"The AERONET level 1.0 data are unscreened measurement data. The cloud and pointing error screening is applied to level 1.0 data to produce a level 1.5 dataset. The level 1.5 data series are raised to level 2.0 (quality-assured) series after final calibration values are applied and manual data inspection is completed."

L 178 Please motivate better the choice of NDVI. Fig 5, 6 and 7 (b) do not show a distinct relation between NDVI and error.

The motivation for the choice of NDVI and the other error sources are described in detail in the Section 3.2.

L 225 What is the procedure if an instrument product has no data?

As indicated on the table 3, fusion AOD is produced in accordance with the data availability.

L 226 All retrievals come from the same satellite. So, index i should represent the instrument or product.

Done. Thank you.

L 235 From which instrument was the aerosol type obtained?

Each aerosol product has their own aerosol types as side products.

L 245 This type classification should be explained earlier.

Type classification is explained in the Section 2.2.1 and 2.2.2.

L 265-280 It remains unclear what type of cloud masking was applied for the different products and if all problems related to cloud masking could be resolved. Please provide a summary in this section.

The paragraph is moved to Section 3.1 so that the readers can understand that the cloud masking procedures listed in Table 3 and that the cloud masking is applied to both GEMS and GOCI-II.

L 287 Please provide more detail on EE. What is its purpose? What is tau. Why does it make sense to use the MODIT DT approach here.

It is used as a common metric to evaluate multiple aerosol optical depth products at once. Also, the expected error envelope was calculated analytically by Levy et al. (2013). Authors clarified that the EE envelope is borrowed from the reference as "The expected error envelope (EE envelope $\pm(0.05 + 0.15AOD)$.) of AOD was established by Levy et al. (2013)"

Also, tau is replaced to "AOD".

L 343 Why were only the fused data processed and evaluated for the two different domains EA and KO? Which was the domain for the single instrument data?

The reason of the use of two different domains is explained as "The GOCI-II field of regard focusing on KO was smaller than those covering EA, so the fused AOD utilizing GOCI-II AOD was confined within the domain. Therefore, two groups of fused AOD products were generated: one involving the entire EA domain (AOD-EA), and the other focusing exclusively within KO (AOD-KO), which is the domain covered by GOCI-II.". The evaluation of original data at two domains are added in the Table 5 and Table 6.

L354 "The statistical fusion approach thus effectively accommodated nonlinearity in retrieval uncertainty, despite possibly not capturing all complexity in the data." Please explain better what you mean. The fused data have the advantage of being optimized to match the AERONET data.

Revised as "By merging the original AOD dataset according to retrieval error compared to AERONET in different retrieval conditions (NDVI, observation time, aerosol loading and type), the statistical fusion approach thus effectively accommodated nonlinearity in retrieval uncertainty, despite possibly not capturing all complexity in the data.". Also, additional explanation is added in Section 3.2 "Statistical aerosol fusion: MLE AOD" as "As shown in Fig. 5-7, retrieval error does not increase (or decrease) linearly. Therefore, merging AOD datasets using the same RMSE value for all pixels is not desirable. The statistical fusion method linearizes the error characteristics by categorizing potential error sources such as AOD values, NDVI values, and observation times."

L 361-415 This section is perhaps to detailed and complicated. It would be sufficient to simply compare the AOD products and MODIS DT against AERONET and compare the errors for the different situations or locations. It remains slightly unclear if the new fusion products have smaller errors than the MODIS DT retrievals over the study area.

We agree that the including MODIS DT to compare other products is confusing. Fig. 9 has been simplified to plot 1σ error vs. AOD. The overall description of section 4.2.2 is much simplified.

L 363 please explain the retrieval error. Is that the "theoretical" retrieval error provided by the retrieval algorithm or the error of the product against AERONET. How is the theoretical retrieval error of the fused data set calculated.

Retrieval error is calculated as 68th percentile of AOD error of the product against AERONET. Description is added as "(68th percentile of absolute AOD error against AERONET, $|\Delta_S|_{68}$; 1σ of gaussian distribution)". Also, according to Sayer et al. (2020), the "retrieval error", which is 68th percentile, corresponds to 1σ of gaussian error distribution. Thus, the retrieval error can be regarded as theoretical retrieval error.

Figures:

Please include the statistical error measure of Fig 4 and Fig 8 in a table.

Table 5 and Table 6 are added in the manuscript.

It is not obvious what Fig 9 shows.

Fig. 9 is changed. Please refer to the response on L361-415.

Tables:

See general comment.