

We would like to thank reviewer-1 for the interest shown in our paper and the valuable feedbacks that will improve the quality of the paper.

1. *“Line 59: think this should be 2000, not 2001?”*:

We agree the starting year for MODIS data is 2000

2. *“Line 65: a citation for the GCOS requirements should be added. I think the numbers for some geophysical quantities change periodically”*:

This was taken from the GCOS implementation needs GCOS-200, 2016 report. We added the reference.

3. *“Table 1: the MODIS DT ocean uncertainty is missing from the table. I believe it is $0.04+10\%*AERONET$ in Collection 6 onwards (based on Levy et al 2013 paper cited in the manuscript). “ :*

The MODIS DT uncertainty for C6.1 is included in Table 1. It was taken from the updated information given in the official DT website (<https://darktarget.gsfc.nasa.gov/validation/results>)

4. *“Section 2.1.1: might be good to specify here again that this is the NOAA VIIRS product. There are NASA DB and DT VIIRS products as well. I see people cite the NASA papers using the NOAA products sometimes, and vice versa, so doesn't hurt to add a sentence saying directly that NASA products exist but are not used here. I don't know the NRT status of the NASA VIIRS ones at present, and see it makes sense to use the NOAA ones given they are NRT (and the resolution is nice too).”*

As suggested by reviewer 1, we added a sentence in paragraph 2.1.1 to clearly state that the VIIRS product used in this work is the EPS dataset produced by NOAA in NRT. We acknowledge that VIIRS AOD datasets are also produced in NRT by NASA (Sayer et al., 2017; Hsu et al., 2019; Sawyer et al., 2020). However, the NASA dataset only include retrievals from S-NPP while the NOAA EPS product includes both S-NPP and NOAA20 retrievals. For this reason, it was decided to implement and test the NOAA product in CAMS. We may envisage to test the NASA products in the future (the next processing should include NOAA20).

5. *“Line 162: I think this is the first time the acronym TOA is used, I know what it means but it should be defined”*:

We defined TOA here.

6. *“Line 169: is there a paper or tech report citation about the issues with PMAp and SLSTR over land?”*:

We used the validation reports of PMAp (EUMETSAT, 2021a) and SLSTR (EUMETSAT, 2021b) to justify their lack of accuracy over lands.

7. *“Line 170: is this really how DB and DT are combined here? If so, why not just use the merged product provided within the files? It is not the same as gap filling DB with*

DT, there is some averaging and QA comparison done too. See the Levy (2013) paper mentioned earlier”.

The combined DT-DB product has been available from NASA since collection 6.0. It was not available at the time of the operational implementation of MODIS DB in CAMS. It was thus decided to use best quality DT retrieval and to gap-fill it with best quality DB retrieval. We agree that this would be slightly different than using the combined product produced by NASA which consists in selecting the DB or DT retrieval which has the best QA value or averaging both if their QA are equal. We added a sentence to justify our choice.

8. *“Line 222: for non-modeler readers, it would be useful to state what the TL511 model resolution corresponds to in km or degrees. Is this the same as the 40 km resolution mentioned on line 230 or something different?”*

TL511 is equivalent to a grid size of about 40 km. We added it at line 222.

9. *“Paragraph beginning line 248: if I understand correctly, the first guess departure will be useful for the absolute evaluation of the satellite products if the (un-assimilated) model itself is somewhat skillful. If the model is not good at a certain place/time then you wouldn't necessarily know whether the difference is due to model or observation errors, and conversely if the model were perfect you could use it perfectly to diagnose observation errors (but then assimilation itself wouldn't bring a benefit). Is that right? I suggest adding a sentence or two here for non-modeler readers to explain more why this is a useful metric and what the caveats/assumptions are. Presumably the fact that first guess departure is based on the model field including MODIS/PMAp assimilation from previous time steps, makes up for some potential errors in the model (assuming in that case that MODIS/PMAp in the previous time step were good).”*

The first guess departure represents the differences between the observation and the model equivalent of the observed variable at the time and location of the observation (observation space). It is the result of both the observation and the model errors (bias and random error). The mean and the standard deviation of the first guess departure represent the systematic and random, respectively, components of the difference between the observation and the model. Model errors mainly arise from uncertainties in process representation, parameters and forcing variables. Observation errors include retrieval error (e.g. errors in radiance measurement and retrieval algorithm) and representation errors (e.g. observation operator used to convert the model variable into its observation equivalent, spatial and temporal interpolation to map the model variable in the observation space). Data assimilation systems are designed to correct small changes and random errors. The first guess departure needs to keep reasonable small values to mitigate the impact of non-linearities. Any bias in the observation may lead to larger errors in the analysis and can result in inconsistencies between distinct satellite observations which may fight against each other when they are assimilated. The use of the first guess departure is twofold: i) check that the mean departure between each type of observation and the model is reasonably small and not impacted by any biases in the observation and ii) evaluate the retrievals relatively to the model to identify possible spatial and temporal inconsistencies between AOD satellite products that would impact the assimilation of multi-satellite AODs. This requires that the model is skillful to some extent and lowly biased compared to the observation which a reasonable assumption given that the short-range forecast used to compute the first guess departure is simulated from optimised

initial conditions produced by the data assimilation system. We agree with the reviewer's remark that this is not a pure model-observation comparison since the simulated values are influenced by the previous assimilation cycles. Besides, lower mean and SD of first guess departure are expected for MODIS which is assimilated and influence the last analysis cycle. The characteristics and the role of the first guess departure and the lower first guess departure expected for assimilated observations are provided in the last paragraph of Section 3.3 of the submitted version. As suggested by reviewer-1, we added two sentences on the assumptions associated with the use of the first guess departure for the relative evaluation of multi-satellite AOD retrievals.

10. *"Figure 1, 2, 5, 6: the paper says the analysis is only for data with latitudes smaller than 70 degrees. The maps include data above 70 degrees (except for MODIS which seems to have a cutoff). So it's not clear if that data is used in the discussion of this figure or not. If data above 70 degrees are not used, I would suggest not plotting it in the maps."*

We agree and we modified the plot to display data within 70S-70N

11. *"Mapped figures: these still say SDD for standard deviation, not SD like in the rest of the paper. I also wonder if it's possible to put the mean and SD on the same line as the sensor name, with the two-line plot titles there's a lot of space between panels which makes it a bit harder to visually compare than if they were closer together."*

We replaced SDD by SD and we adjusted the title as suggested.

12. *"Line 374: I wonder if here (or earlier) you could introduce an acronym FGD or a symbol for "first guess departure". The phrase appears a lot in this section, it would be easier if there were some shorthand for this term. I counted 22 uses on page 22 alone, and about a dozen on the corresponding land section (page 24)."*

"We use FGD as acronym for first guess departure.

13. *"Line 376: if I understand correctly, negative first guess departure means the satellite is lower than the model field. Is that right, or do I have it backwards? For a non-modeler reader, as this is the first example given in the paper, it would be good to state this clearly to make sure people don't get the conclusions backwards. If that is correct then would it imply the model AOD is higher over ocean (since we know most satellites are also too high) – possibly because of the assimilation of biased MODIS and PMAp observations in the previous time step?"*

The first guess departure is the difference between the observation and the model. Negative first guess departure values indicate that the satellite AOD is lower than the value of the model. We explicitly state this in the text to avoid any misinterpretation. The first guess departure of VIIRS and SLSTR are frequently negative for the oceanic background aerosol. The larger value of the model can be related to i) positive bias in the model for sea salt and ii) the assimilation of TERRA/MODIS which is known to be positively biased ocean. This is consistent with the lower value of VIIRS AOD than MODIS AOD over ocean.

14. *“Figures 7, 16, D5, D9: it would be useful to add the horizontal line $y=0$ here, as a reference for zero mean departure.”*

We added horizontal lines on Figures 7, 16, D5, D9.

15. *“Section 4.2: how are the unphysical negative AOD retrievals in the Dark Target land product handled here? Are they set to 0, set to invalid data, or something else? From the Levy (2013) paper again, it happens about 20% of the time over land (see e.g. Figure 10 of that paper). As a non-assimilation reader it would also be useful to know how these are handled in the assimilation process because of course a negative aerosol mass would not make sense. This should be explained as it’s an important long-term issue with that data product which is relevant for assimilation.”*

MODIS DT retrieval allows negative AOD retrieval to avoid artificial bias in long time series. To avoid unphysical AOD values, it was decided to set the negative AOD values from the DT retrieval algorithm to zero in the pre-processing of the MODIS observation (Benedetti et al., 2009)

16. *“Line 618: the Schugens reference should be Schutgens. “*

We corrected the reference

17. *“Line 620: Sea Surface Temperature does not need to be capitalized.”*

We corrected sea surface temperature

18. *“Lines 644-646: This comment about resolution seems speculative and unsupported to me. At those transport distances, I don’t see why MODIS would not see the transported smoke at 10 km retrievals but VIIRS would at 750 m. From looking at imagery of the event, the source plumes fairly quickly spread out to more diffuse areas tens to hundreds of km wide. I think it is more likely that the differences in the model-scale aggregate are influenced by different populations of high vs. low AOD retrievals (real or artefact) in these two products, i.e. a pixel screening issue, not a retrieval resolution issue. If the authors want to make a claim here, evidence should be shown to back it up. For example showing examples of the source L2 data and of the L3 data for such a transported smoke case would quickly show what pixels from each sensor are available and what the retrievals look like.”*

The differences between MODIS and VIIRS AOD at the model spatial resolution with respect to the detection of the transported Australian smoke plume in the South Pacific are likely related to differences in spatial representativity between MODIS and VIIRS generated by differences in cloud screening. A possible reason for the differences in cloud contamination between VIIRS and MODIS is the use of a smoke detection test in the VIIRS product which should reduce the commission errors between smoke and cloud pixels. The consequence for the data assimilation system is that the smoke plume cannot be resolved when assimilating only MODIS data.

19. *“Line 675: again, this seems speculative and needs to be better supported with evidence or deleted. The water-leaving signal is not that large or variable at SLSTR wavelengths (green and longer), so the lack of a blue band would not be so important here. The MODIS DT ocean retrieval does not use the blue band either so it would not explain the MODIS-SLSTR differences. I don’t know if the VIIRS one does, but either way, I don’t think it dynamically accounts for pigment variations. So the sensors are all basically using the same spectral information, i.e. green to swIR wavelengths. I think that differences in the Southern Ocean are more likely due to different tolerances of cloud contamination and 3D effects, which may influence SLSTR differently because of the dual view and resulting parallax difference in location for clouds. See Toth (2013) for a discussion of MODIS Aqua in the Southern Ocean: <https://doi.org/10.1002/jgrd.50311>”*

We agree that the NASA MODIS DT and the NOAA EPS VIIRS AOD products are not relying on the blue band to retrieve AOD over ocean. The blue band is used in the internal cloud detection schemes of both NASA MODIS and NOAA VIIRS products and for the NOAA product in the heavy aerosol identification test. We agree that the uncertainties in the retrievals and the differences between satellite products in the South ocean are likely related to cloud contamination. This is discussed in Section 5.7 where Toth et al., 2013 reference is already quoted. We thus decided to remove the statement on line 675 identified by the reviewer.

20. *“Line 689: Sirish reference should be Uprety (Sirish is the given name, Uprety the surname).”*

We corrected the reference

21. *“Line 710: I don’t think “exploits” is the right word here. Rather, the MODIS retrieval LUT contains nodes at those wind speeds.”*

We replaced “exploits” by the reviewer suggestion.

22. *“Line 748: that Sayer (2018) reference is to the NASA VIIRS data products, not the NOAA VIIRS data products used in this study, so may not be directly relevant to this point (other than to show another algorithm as a point of comparison). It was not clear to me reading whether this was intended to be an example of surface influence or an attempt to explain the results of the present study.”*

The Sayer et al., 2018 reference to the NASA VIIRS product was used as a point of comparison to illustrate the impact of uncertainties in aerosol and surface reflectance models on AOD retrieval. We agree that this can be confusing since the NASA VIIRS product is different than the NOAA VIIRS product investigated in this work. We decided to remove this statement and to keep the Tao et al., 2017 reference to support the underestimation of the MODIS DB over desert regions.

23. *“Section 6: as written, I did not find this useful as there was a lot of repetition with the immediately preceding Section 5. I suggest this is shortened and rewritten to focus more on what the abstract says the paper is about: evaluating the SLSTR and VIIRS data sets. I understand the actual assimilation will be analyzed in a follow up paper. But I think that the Conclusion here should maybe present a few brief expectations of how useful the data may/may not be, instead of repeating the previous discussion. For*

example, my guess is that the SLSTR product might not be useful to assimilate over ocean as it seems to be unusually low compared to the other ones. VIIRS on the other hand might be ok from NOAA20 but not SNPP, because SNPP seems to have radiometric calibration issues. So my (non modeler) takeaway from the results presented is that neither of these products are likely to help CAMS much, at least so long as the current MODIS products remain available. Is that a fair assessment, and if not, why not? This is the sort of content I think the conclusions should be giving, i.e. don't repeat the results of the analysis but more talk about what they mean. There is some of that in the current Conclusions but not much."

We shortened the conclusion and provided recommendations for the assimilation of the investigated products based on the outcomes of this work: The assimilation of the SLSTR collection 1 product is not envisaged due to the differences in spatial representativity which are related to the stringent cloud filtering applied to the SLSTR radiances used to retrieve AOD. EUMETSAT is currently preparing a collection 3 based on a new cloud mask that will be evaluated in a future work. This paper highlights the overall good consistency between the NOAA VIIRS product and the NASA MODIS product. The consistency between the NASA MODIS and the NOAA EPS VIIRS AOD products reported in this paper shows that the assimilation of VIIRS will ensure the continuity of the CAMS data assimilation system and it will strengthen the resilience against possible future failure of MODIS. This work shows that the NOAA VIIRS product will enhance the spatial coverage of AOD observations and will provide a more accurate detection of smoke plumes. However, the conclusions reported in this paper are not sufficient to automatically include the additional AOD observations into the CAMS system and further assimilations tests are planned and will be reported on in a follow-up paper. For example, there is a need to understand how the differences between MODIS and VIIRS over ocean and land will impact the analysis. While the magnitude of the mean deviation between the products is smaller over ocean than over land, given the low AOD value of the ocean background, a slight difference in AOD between products will have a large impact on data assimilation. Since VIIRS has lower values than MODIS over ocean, its assimilation will probably decrease the analysis increment over ocean which is known to be too high due to the positive offset of TERRA/MODIS. Over land, the larger VIIRS AOD for biomass burning and dust source regions should lead to larger analysis increments that may affect AOD and surface particle matter predictions other these regions. In addition, our study reveals significant departures between products retrieved from the same instrument but from different satellite platforms and this will affect how bias correction is carried out within the system. This work shows that it would be preferable to use NOAA20/VIIRS as an anchor and apply the bias correction to S-NPP/VIIRS which was found to be positively biased over ocean. Our results also highlight the role of geometry in retrieval uncertainties that can lead to systematic differences between products. Adding the scattering angle in the current variational bias correction scheme implemented in the CAMS data assimilation system could help to represent any geometry-dependent biases in the retrieval. Finally, the observation error is an important variable to weight the relative contribution of each satellite observation to the analysis. Further work is required to evaluate the retrieval error associated with each product which could be inflated to better reflect the larger diversity between products reported in the South ocean and over bright land surfaces.

24. *"Appendix A: thank you for moving this section out of the paper into an Appendix, it makes the main paper more readable, and now if someone wants to know more details but not read the algorithm/validation papers this gives a summary."*

25. *“Appendix B: it’s not obvious to me what the blue lines on the plots represent, what is it? I’m not sure they are useful and maybe they can be deleted to reduce clutter. I also think it’s more useful to show the 1:1 line than what I guess is a regression line here (again, the Appendix doesn’t say). That way one can more directly see whether one data set is higher/lower than another by whether they are above/below this line, without having to cross-reference the existing lines to the labels on both axes. The regression lines seem skewed by offsets at low-AOD conditions (as most of the points are there) whereas as a reader I am more interested in whether one is lower or higher than the other across the full range of AODs. That is less clear when showing the regression line than showing the 1:1 line would be.”*

We modified the scatterplots and we replaced the regression line by 1:1 line