Dear Editor,

In this response we reply to the comments by Referee #3 and Referee #4 for the second revision of our manuscript “A global perspective on CO2 satellite observations in high AOD conditions”.

Report #2 by Referee #3

The paper compiles and analyzes a dataset of XCO2 and aerosol observations from MODIS and OCO-2. The authors specifically note a correlation between AOD and XCO2, investigate the effects of relaxing the MODIS AOD threshold to 0.5, and find differences between MODIS and OCO-2 AOD retrievals as well as issues with the OCO-2 quality flag.

Reply: We thank the referee for going through the of the manuscript again after revision, and for the useful comments. We have revised the manuscript accordingly and answer the questions below.

General comments:

The paper is very long, and it is confusing to flip between the text of the paper and figures which only appear in the appendix. I recommend changing this so that all figures which are mentioned in the main text also appear in the main text. This may require making the paper more concise or removing some supplementary figures. I would consider cutting section 4.4, as I think the results there are relatively minor.

Reply: We admit that the manuscript is long and has many figures. We tried to make the main text part of the original manuscript shorter and easier to read by placing some of the figures we found less crucial to the Appendix. The choice of most relevant figures for the main text was not easy, and as the Referee pointed out, this forces the interested reader to flip back and forth between the Appendix and the main text.

We agree that the results in Section 4.4 (Temporal and spatial dependence) are less crucial for the main points in our work, and could be removed to shorten the manuscript. We have now removed Section 4.4 and only briefly summarize the results in section 4.3. We have removed Fig. 8 from the main text and Figs. A9 and A10 from the Appendix. We also removed Tables A2-A3 from the Appendix, and reduced Tables A4 and A5 to the four main subsets of the collocated dataset. We have also removed from the text the references and discussion related to the removed figures and tables.

We have moved Figs. A3, A5 and A8 to the main text. We consider that Figures A1, A2, A4, A6, and A7 are less crucial and are only briefly mentioned in the text and can remain in the Appendix, to not lengthen the main text section. Figures A2 and A6 simply repeat Figures 1 and 5 from the main text with MODIS DB data. Figures A1, A4 and A7 are more technical, and not strictly necessary for following the main text, but contain some detail for the interested reader.

The paper contains several different analyses (in sections 4.1-4.5) which are not necessarily obviously related. For example, I do not understand how the analysis in section 4.5 on coverage improvement resulting from relaxing the AOD threshold relates to the analysis in section 4.1 on the relationship between OCO-2 AOD retrievals and MODIS AOD retrievals. Perhaps the authors can edit the abstract and the conclusion to draw more clear connections between the different analyses, and can discuss how the different findings relate to each other. I think this would enhance the paper and make the impact of the work more clear to readers.

Reply: We thank the Referee for pointing this out, and we now try to explain the connections in more detail both in the abstract and in the conclusions. We have also rewritten the introduction to Section 4 to better explain the connection between the subsections. Furthermore, we have relabeled sections 4.1 and 4.2 to better reflect the contents.

In Sections 4.1 we showed that there is a spatially varying AOD difference between OCO-2 and MODIS, and the difference is largest in areas of constant high aerosol loading. This means that a possible XCO2 bias related to the aerosol treatment in the ACOS retrieval could have different magnitude at different parts of the globe, which might cause uncertainties in emission estimates employing satellite data.

In section 4.2 we show that the OCO-2 quality filtering misses some high AOD cases, which remain in the so-called 'good quality' dataset, and this happens more frequently in areas of high anthropogenic emissions.

In section 4.3 we showed that there is a correlation between AOD and the XCO2 retrievals of OCO-2 (either a real correlation or a retrieval bias). With the TCCON comparison we then showed that this correlation is real, and not due to an aerosol related retrieval artefact. Thus, when the OCO-2 retrievals are limited by an AOD threshold, this causes a sampling bias, where the removal of high AOD areas leads to removal of high XCO2 values. Hence, it is important to estimate in Section 4.5 how much the sampling bias could be mitigated by relaxing the AOD threshold, especially in the urban areas which are critical in monitoring the anthropogenic CO2 emissions.

In section 4.3 we also found that XCO2 is biased low at high AOD. Unfortunately, the connection between the aerosol bias observed in sections 4.1 and 4.2 and the XCO2 bias observed in section 4.3 proved to be complicated, as the other retrieval parameters in the full physics retrieval can compensate for the errors in AOD, and we were unable to draw a direct connection.

(As the changes to the text in the abstract, in the introduction to section 4, and in the conclusions are rather large, we do not repeat them here but refer to the revised track-changed manuscript.)

Report #3 by Referee #4

Review of "A global perspective on CO2 satellite observations in high AOD conditions" by Virtanen et al.

I was requested to review the revised version of the manuscript. In this manuscript, Virtanen et al. present an analysis of the collocated MODIS and OCO-2 satellite observations. The paper addresses an important question within the scope of AMT. One key conclusion is that there is a considerable difference between MODIS and OCO-2 AOD, which affects the XCO2 retrieval. I commend the authors for using several large datasets. I have some major questions related to your analysis. Therefore, the authors should address these questions before the manuscript can be published. For the following comments, please refer the page and line numbers to the track-changed manuscript.

Reply: We thank the Referee for reviewing the revised manuscript and for the valuable comments. We agree that the comments about the quality of MODIS aerosol product as a reference data raised by the Referee are important. After carefully considering the comments, reviewing related literature, and making some additional analyses, we find that we can address these points without performing a full reprocessing of the global five year dataset with the MODIS 3 km aerosol product. We also argue that since the MODIS aerosol product has already been extensively validated in the research literature, there is little benefit of repeating this considerable effort in this work. However, we readily admit that these points need to be better addressed in the manuscript. In the detailed responses below we show our additional analyses and explain how we take these concerns better into account in the revised version of the manuscript.

Major comments:

1. If I understand it correctly, the authors perform the comparison between the MODIS Aqua 10 km product and OCO-2 AOD 1 km x 2 km product. In section 4.1, the authors note that it may affect the comparison because of the different spatial resolutions but does not consider how the spatial representativeness issue contributes to the large difference between the two products. And hence I doubt that the ACOS retrieval is solely responsible for the difference.

I suggest that you may minimize the representativeness issue by using the MODIS Aqua 3 km product (MYD04_3K) and possibly 2x2 OCO-2 AOD pixels (~2 km x ~4km).

Reply:
We argue that the spatial variability of ambient aerosol fields is typically relatively smooth on the length scales considered, so that the choice of MODIS aerosol product should have minimal impact on the global scale. There is a wealth of literature on comparison of the MODIS 10 km and 3 km products (e.g. Gupta et al. (2018), He et al. (2017)). The general conclusion from this literature is that the two products perform rather similarly, on the global scale, in comparison with AERONET. This supports our view that the comparison to the finer spatial scale OCO-2 product should not depend much on the MODIS product resolution on the global scale.

In our analysis we heavily aggregate the individual collocated data points. In the AOD matrix plots we aggregate the data to AOD bins of width 0.02, with dozens of data points in each bin. In the global maps, we aggregate the data to 1x1 degree grid cells with thousands of data points in each grid cell. We argue that this aggregation further reduces any effect the MODIS product resolution might have on the results.

Furthermore, as we emphasize in the manuscript, OCO-2 does not provide a specific aerosol product. The total AOD value given in the OCO-2 product is just one of the dozens of parameters in the full-physics retrieval, and the values are largely dependent on the large-scale aerosol climatology values used as priors. Again, we do not expect to get a one-to-one correspondence with high spatial resolution reference aerosol data but are interested in the larger scale statistics.

We certainly acknowledge that the MODIS 3 km product can be useful for individual overpasses and e.g. for thick aerosol plumes. But as mentioned in the manuscript, detailed case studies are beyond the scope of this work, and we believe that the effect of the MODIS aerosol product resolution and sampling methods would better fit such work. We think that the cited literature supports this conclusion, in that the differences between the two MODIS products show up on a smaller scale, but are smoothed out in larger data sets. Similarly, we expect that averaging the OCO-2 data to 2x2 pixels would make a little difference in the end.

In order to verify our understanding that the choice of the MODIS product makes little difference on the global scale, we have processed a limited number of MODIS 3 km aerosol data (MYD04_3K). The land-global comparison of MODIS 3 km product with OCO-2 for the year 2018 shows minor differences compared to the similar comparison using the 10 km product (see Fig. R1 below). Hence, we are confident that our original analysis sufficiently covers the appropriate MODIS information.

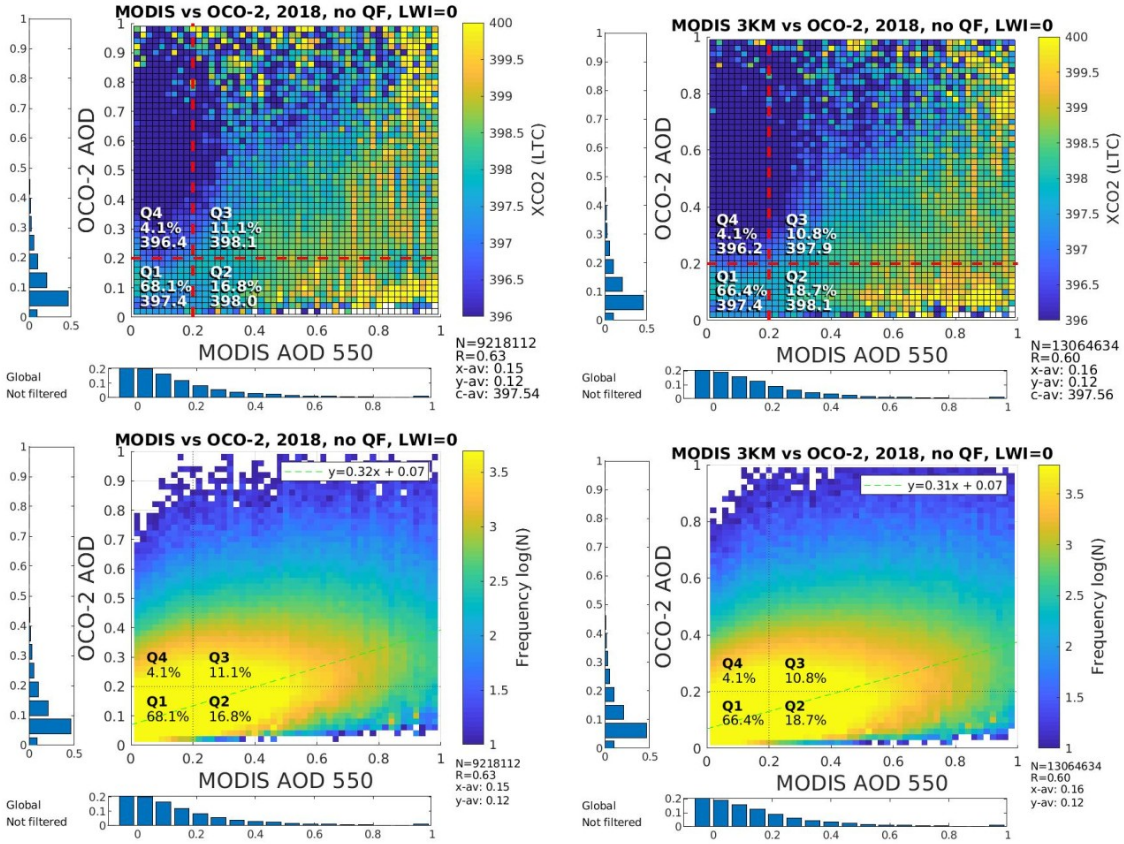

Figure R1. Difference between MODIS 10 km (left) and 3 km (right) aerosol products in comparison with OCO-2 for one-year land-global dataset. The 3 km product provides more matches with OCO-2 but the main conclusions do not change.

Reprocessing the full land-global 5-year dataset with the MODIS 3 km product (which is larger in size compared to the 10 km product due to the increased number of data points) would require considerable resources, and we expect that the resulting difference compared to the existing work would be minimal.
We have now further elaborated our choice of the reference satellite aerosol dataset in the manuscript (Section 2.2), and the benefits of using higher resolution products in more detailed case studies. We also briefly discuss the results of the additional analyses using the 3 km product in Section 4.1. (Due to the request by other Referee to make the manuscript more concise, we do not include additional figures in the manuscript).

Added text, L89 (line numbers refer to the new track-changed manuscript):

> "Collocation with the higher spatial resolution MODIS 3 km aerosol product (MYD04_3K; Remer et al. (2013)) was tested for one year (2018). The results did not differ significantly from the corresponding subset when using the 10 km DT product (not shown). Due to the considerably larger computational burden of the 3 km data, the full dataset was processed only with the 10 km product. Previous studies have shown that the 3 km and 10 km products perform very similarly on the global scale (Gupta et al., 2018; He et al., 2017). For more detailed case studies the use of MODIS 3 km product could be beneficial, but that is beyond to scope of this exercise."

Added text, L220:

> "A limited collocation test made with MODIS 3 km aerosol product for the year 2018 shows slightly enhanced coverage but otherwise very similar AOD patterns as the 10 km DT product."

2. The MODIS retrieval can also contribute to the AOD difference between MODIS and OCO-2. While the comparison between OCO-2 and MODIS AODs is very detailed, it is not very useful information unless you know which one is better with respect to AERONET (sort of a ground truth). I agree that the MODIS Angstrom exponent can be uncertain. But alternatively you can 1) compare MODIS AOD at 550 nm with AERONET and 2) compare OCO-2 AOD at 755 with AERONET using AERONET Angstrom exponents. You have compared OCO-2 with AERONET in Figure A5. I suggest that you should do it again using MODIS.

Reply:
We acknowledge that the MODIS aerosol product has uncertainties which affect the comparison. However, we find that the MODIS product is probably the most extensively validated satellite aerosol product there is [e.g. Levy et al. (2013), Sayer et al. (2014), Wei et al. (2019)], and we see little benefit in repeating the validation against AERONET in this manuscript. Clearly, we have failed to discuss the MODIS aerosol product quality in the manuscript, and this has now been improved in the latest revision.

The validation studies typically show a correlation coefficient R ~ 0.9 and small bias for MODIS against AERONET, while for OCO-2 we find R ~ 0.7 and a considerable low bias at large AERONET AOD values. We also find a striking similarity in the comparisons of OCO-2 AOD data to AERONET and to MODIS data, in that OCO-2 tends to have a low bias at high AOD values. This supports the view that the MODIS product can be used as a global reference for AOD. Here we want to once again emphasize that we are not attempting to validate the OCO-2 total AOD component as if it was a dedicated aerosol product; the OCO-2 total AOD is one of the many retrieval parameters of the full-physics retrieval, and we are mainly interested in the large scale statistics when comparing with the MODIS aerosol product.

As pointed out by the Referee, we evaluate OCO-2 aerosol data against AERONET. To further address the Referee's concerns, we have performed a separate validation of the MODIS part of the collocated dataset against AERONET (see Fig. R2 below). This differs from the typical validation in that the sampling is not optimal for MODIS but limited to the pixels collocated with OCO-2. As expected, this sampling leads to slightly reduced validation metrics against AERONET (R ~ 0.8, small bias), but still better than for OCO-2. We note that while repeating the validation of the full MODIS dataset would be laborious, using the collocated dataset for this limited exercise was straightforward and better fits the scope of this work.

We now discuss the MODIS aerosol product quality and suitability for global AOD reference for this study in more detail in the manuscript. We also draw attention to the similarity of the OCO-2 comparison against AERONET and against MODIS, and discuss the results of the limited validation of the MODIS component of the collocated dataset (Section 4.1). Since we were requested to make the manuscript more concise, we will not use additional figures in the manuscript.

Added text L256:

> "The MODIS aerosol products have been extensively validated, with a typical correlation coefficient R∼0.9 against AERONET (Levy et al., 2013; Sayer et al., 2014; Wei et al., 2019). We do not repeat the MODIS AOD product validation against AERONET in this work, but we have compared the MODIS part of the collocated OCO-2/MODIS dataset to AERONET with similar sampling as used for OCO-2. This differs from the typical validation in that the sampling is not optimal for MODIS but limited to the pixels collocated with OCO-2. As expected, this sampling leads to slightly reduced validation metrics against AERONET (R∼0.8, small bias), but the metrics are still better than for OCO-2. Hence, we

are confident that although MODIS AOD product certainly has higher uncertainty than AERONET, it helps to the extend the evaluation of OCO-2 AOD to the global scale."

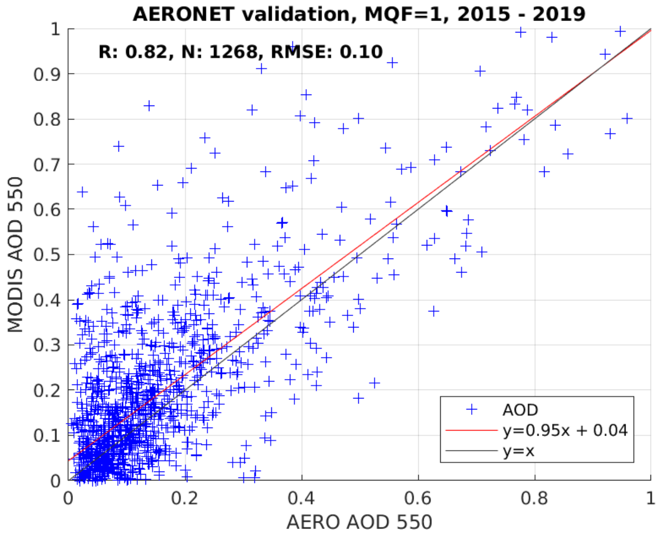

Figure R2. Comparison of the MODIS AOD from the collocated MODIS/OCO-2 dataset to AERONET at 550 nm. Collocated MODIS data is limited to the narrow OCO-2 swath. Collocation distance is 0.125° around the AERONET site and time window is ±1 h.

Regarding the wavelengths used in the comparison with AERONET, we used the approach of averaging two wavelengths for simplicity in Fig. A5. As discussed in the revised manuscript and in our reply to Referee #2 in the first revision of the manuscript (https://doi.org/10.5194/amt-2024-77-AC2), we have repeated the AERONET comparison with a subset of data using the Ångström exponent from AERONET, and showed that this has a small effect on the results. This is also discussed in the revised manuscript (Section 4.1, L251 in the new track-change manuscript). In Fig. R2 we used AERONET Ångström exponent to scale the AERONET measurements to 550 nm.

3. MODIS quality flag can have values from 0 to 3 (best quality). While you remove the poor quality pixels (flag = 0), does it matter if you only compare with pixels with best quality.

Reply:

We see that the discussion concerning the MODIS quality flag was quite brief in Section 2.2, and we have now extended it. The choice of quality flag is a trade-off between data quality and coverage (Fig. R3 below). While we concentrated on the OCO-2 quality flag, we have also tested the effect of using different MODIS quality flags in the filtering. We find that applying more strict filtering on the MODIS data reduces the number of data, reduces the average MODIS AOD by 0.02, and slightly improves the validation against AERONET, but does not significantly affect the comparison between OCO-2 and MODIS too much (see Figs. R3 and R4 below).

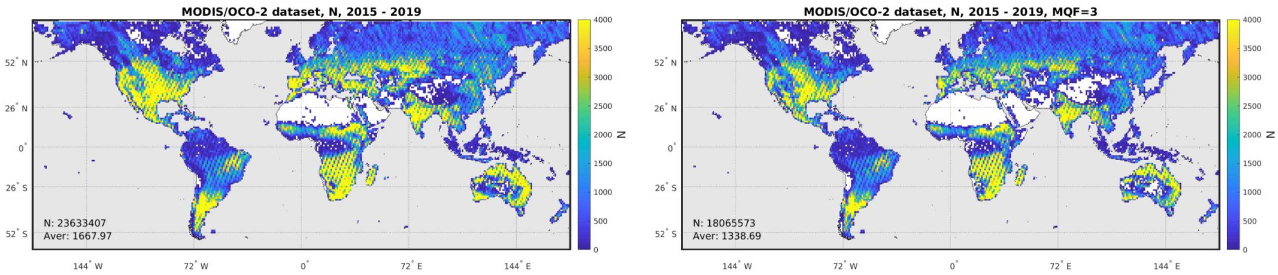

Figure R3. Effect of MODIS quality flag on the coverage of the collocated dataset. Left: only worst pixels (qf=0) removed. Right: only best pixels (qf=3) kept. OCO-2 quality filtering has been applied.

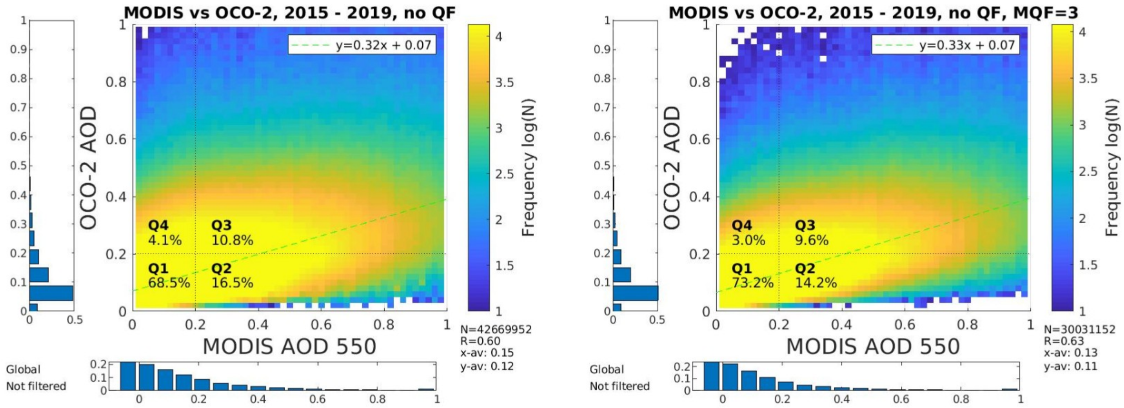

Figure R4. The effect of MODIS quality flag on the collocated data. On left, we have removed only the lowest quality MODIS pixels (qf=0), and on the right hand panel we have kept only the best quality MODIS pixels (qf=3). OCO-2 quality filtering has not been applied.

Since we are interested in the performance of the OCO-2 quality filtering, we did not want to limit the MODIS data too much. Also, we found that more strict quality filtering tends to remove a larger fraction of pixels from the urban areas, which are of specific interest in this study. We now discuss the effect of MODIS quality flag in more detail in the manuscript (Section 2.2).

Added text L101:

> "We also tested using more stringent quality filtering, keeping only the best quality MODIS data (quality flag 3). Although this reduced the number of matches with OCO-2 by nearly 30% and reduced the global average AOD by 0.02, it did not affect the conclusions of our work."

Minor comments:

P4, L106: What does 0.01-0.02 represent? Absolute deviation or standard deviation or something else?

Reply:

According to Eck et al. 1999 this is the estimate total absolute uncertainty in AOD (dimensionless). This is now clarified in the manuscript.

Revised text L115:

> "The AERONET sunphotometer measurements are routinely used as reference measurements for satellite aerosol retrievals due to their high accuracy (∼ absolute error in AOD of the order 0.01-0.02, Eck et al. (1999); Sinyuk et al. (2020))."

P5, L132: What is GGG2020? Please define or explain it.
Reply:
GGG2020 is the name of the latest version of the retrieval algorithm used in the TCCON retrievals (Laughner et al., 2024, https://doi.org/10.5194/essd-16-2197-2024). We have added the explanation and reference to the manuscript.
Revised text L137:

> "The effect of different prior profiles in OCO-2 v10 and TCCON retrieval algorithm version GGG2020 (Laughner et al., 2024) was taken into account by adjusting the OCO-2 XCO2 value, following Mendonca et al. (2021)."

Figure 2: x-av, y-av -> Should it be x-axis and y-axis?
Reply:
In Figs. 2 (as well as in Figs 5, 6, A4, A6, and A8) the text insets on the bottom right ‘x-av’ and ‘y-av’ refer to the average value of the quantities shown on the x and y-axis, respectively (here ‘av’ is short for average). This is explained in the caption of Fig. 2.

P21, L470: seasons -> season
Reply:
This was a typo. The entire Section 4.4 has now been removed as requested by Referee #3.