Referee #1

The Virtanen et al manuscript entitled "A global perspective on CO_2 satellite observations in high AOD conditions" tries to evaluate the OCO-2 product from the point of view of aerosols by the combined analysis of the OCO-2 retrieved XCO_2 and AOD, collocated MODIS AOD, and TCCON retrieved XCO_2. This manuscript found a systematic difference between AOD retrieved by OCO-2 and MODIS, which can impact the retrieval of XCO_2. The dependence of the OCO-2 retrieved XCO2 on the AOD difference was also found. But toward the future CO2M mission, in my opinion, besides simple data coverage variation with different AOD threshold, you should also discuss the quality control with the AOD threshold of 0.5.

Reply: We thank the reviewer for the insightful comments which helped us to improve the manuscript. We have revised the manuscript considering the comments and the questions presented.

Major comments:

You only analyzed the data coverage variation after adjusting the AOD threshold from 0.2 to 0.5. According to your analysis of the relationship between OCO-2 retrieved XCO_2 and AOD, the accuracy of the former significantly depend on the uncertainty of the later. You should also discuss how to conduct quality filter, or what the data quality will be for CO2M, which will use 0.5 as AOD threshold.

Reply: Clearly the discussion in section 4.5 was not sufficiently clear, as pointed out by all Referees. We have revised this section to better describe the objective of this part of the work and to clarify the issues raised. Please find a more detailed response in the specific comments below.

The specific comments are listed below:

P2L24: Change the order of the sentences "An essential monitoring component will be the Copernicus Anthropogenic CO_2 Monitoring Mission (Meijer et al., 2023)." And "While ground-based greenhouse gas measurements are mainly 25 available in developed countries – with limited coverage and representativeness – satellite-based XCO_2 information will be irreplaceable in areas where ground-based measurements are not made."

Reply: We agree that this makes text more fluent and have changed the order.

P2L33: Check the format of the citations "(e.g., (Houweling et al., 2015; Crowell et al., 2019))", should they be "first name et al. (year)"?

Reply: We agree.

P3L68: "In the collocated database we include only a limited selection of OCO-2 data fields, but the sounding ID in the combined daily files is equivalent with the original OCO-2 lite files, allowing addition of more data fields in an effective manner." It makes me confused. Please further explain it. What is "sounding ID"? Why can you get "more data fields"?

Reply: Initially, only a few selected variables from the OCO-2 data were included in the collocated database (e.g., XCO2, XCO2_quality_flag), to save space. Later we wanted to test possible effects of additional variables, so the database was constructed in such a way that this could be effectively done. Sounding ID is a unique identifier for each observation.

We now realize that this is a technical detail on a level that is probably not relevant for readers and might just cause confusion. We have removed this part from the text.

P3L70: "The aerosol parameters of the ACOS algorithm include five scatterers, two cloud types (water and ice), two tropospheric aerosol types and a stratospheric aerosol type (sulfate)." This may should be revised as "The aerosol products of the ACOS algorithm include parameters from five scatterers". Is "five scatterers" refer to "two cloud types (water and ice), two tropospheric aerosol types and a stratospheric aerosol type (sulfate)" If so, this may should be revised as "five scatterers, which are two cloud types (water and ice), two tropospheric aerosol types and a stratospheric aerosol type (sulfate)."

Reply: This sentence was poorly formulated, as noted also by Referee #2. We have clarified the sentence, and it now reads:

"The aerosol parameters of the ACOS algorithm include five scatterers, which are two cloud types (water and ice), two tropospheric aerosol types and a stratospheric aerosol type (sulfate)."

P3L72: There are five aerosol types mentioned in the brackets. Why did you write "Two representative types of tropospheric aerosols"?

Reply: We have reformulated this as:

"Two most representative types of tropospheric aerosols out of five possible types (dust, sea salt, sulfate aerosol, organic carbon, and black carbon) are drawn from a climatology based on location and time (Crisp et al., 2021)."

P3L74: "Atmospheric Carbon Observations from Space (ACOS)". Only use abbreviation is ok.

Reply: We now see that ACOS was defined already on line 43 and have removed the second explanation of the abbreviation.

P3L73: "From the large number of data products…".

Reply: We have replaced "quantities" with "data products".

P4L86: Add last accessed date after the website.

Reply: Done.

P4L106: What are "data fields which are not relevant for this study"?

Reply: This is also a technical detail related to the reduction of the size of the collocated database. As this was explained already on line 88 in section 2.2, we have now removed this part from line 106. The removed MODIS data fields include e.g. viewing geometry and algorithm performance parameters, which were not considered in this study.

P4L120: Why did you use latitude and longitude as the collocation criteria? Using this criterion, the distance between OCO-2 and TCCON is different at different latitudes.

Reply: It is true that latitude affects the sampling in this method. Initially, this was a crude sampling done as a first step when constructing a database, and the adapted approach is similar to earlier studies focusing on the validation of satellite data (e.g. Wunch et al., 2017). The original geolocation information was included in the database. The idea was that different refined sampling criteria (within the larger crudely sampled area) could be tested easily later. When testing smaller sampling areas around TCCON sites, it turned out that the finer sampling had little effect when considering the full global dataset including all 26 TCCON sites. Hence, we kept the original crude collocation criteria. For more detailed studies the sampling should be revisited.

P4L128: 0.1◦ is also the latitude and longitude threshold? Please describe it. Why did you not use the same spatial and temporal criteria with TCCON?

Reply: In the case of AERONET, the sampling criterion was a circle of 0.1 deg around the AERONET site, assuming pseudo-cartesian lat/lon coordinates. We have now clarified this in the text.

The practical reason for different sampling for AERONET and TCCON is that the data were taken from existing collocation databases originally made for other purposes. For AERONET sampling we can afford a smaller sampling area due to the abundance of the data, while for TCCON we wanted to maximize the number of matches. In addition, the atmospheric lifetime of the species affects the choice of the co-location method. Different sampling criteria were tested also for AERONET, and we found that the results were not too sensitive to the sampling distance. Please see also the reply to comments for Referee #2 for more details.

P4L129: "The OCO-2 observations are not spatially averaged." Not necessary for AERONET part.

Reply: This sentence refers to the comparison of OCO-2 total AOD component to AERONET AOD. Typically, in satellite vs AERONET comparison the satellite data around the AERONET site is spatially averaged and the AERONET data is temporally averaged around the satellite overpass time, to reduce the random noise (see e.g. Virtanen et al. 2018).

We have revised this sentence as: "The OCO-2 AOD observations are not spatially averaged for the comparison."

Reference: Virtanen, T. H., Kolmonen, P., Sogacheva, L., Rodríguez, E., Saponaro, G., and de Leeuw, G.: Collocation mismatch uncertainties in satellite aerosol retrieval validation, Atmos. Meas. Tech., 11, 925–938, https://doi.org/10.5194/amt-11-925-2018, 2018.

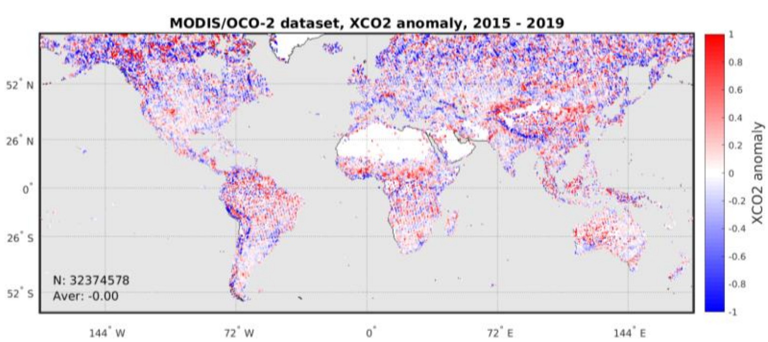
P6L148: What is the specific definition of XCO_2 anomaly? How to calculate XCO_2 anomaly from the median XCO2? The anomaly was calculated from all data, or for every year? After the calculation, have you removed the anomaly or how to use the anomaly? And you said "This is an alternative way to de-trend the data", when did you use the LTC method and when did you use the XCO_2 anomaly? Please supplement this section with more information.

Reply: The first sentence in section 3.6, "The OCO-2 XCO 2 anomaly is calculated for each good quality OCO-2 pixel in the collocated dataset as the difference from the median XCO2 value calculated within 500 km for the corresponding OCO-2 orbit." defines the XCO2 anomaly. It is calculated for each good quality pixel, as the difference of XCO2 value from the spatially and temporally varying median. This reference median for each pixel is calculated from the good quality pixels on the same OCO-2 orbit, within 500 km from the pixel considered.

Since the values used to calculate the median are from the same orbit (within few minutes), using this anomaly is expected to remove both the trend and seasonal effects. The idea is that the yearly increase in CO2 and the seasonal variation are large spatial scale effects which are captured by the 500 km portion of an orbit. When the median value is subtracted, the remaining 'anomaly' part is assumed to contain information on local sources and sinks, while the trend and seasonal effects are (presumably) removed. The XCO2 anomaly is used as such (i.e. we do not attempt to 'remove' the anomaly from XCO2 data).

The XCO2 anomaly is used as a proxy of local CO2 emission and compared with AOD data, to study the covariance of aerosols and CO2 emissions. Since the anomaly data looks rather noisy when aggregated to global maps for five years, we do not show it as a map on the manuscript (see Fig. 1 a) below). Instead, XCO2 anomaly statistics are shown in various tables in the manuscript (e.g. Table 2, Table A2, Table A3). In Table 2, the XCO2 anomaly is larger in AOD quarter Q2 than in Q1, indicating a correlation between local XCO2 enhancement (emissions) and MODIS AOD. Figure 1 b) below illustrates this. The XCO2 anomaly analysis supports our findings in the manuscript and shows essentially the same results as obtained with 'usual' XCO2 data, so they are not shown in more detail in the manuscript.

We have slightly extended the description of XCO2 anomaly in section 3.6 to clarify these points.



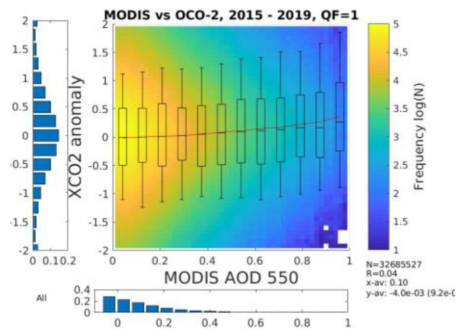


Figure 1. a) Local XCO2 anomaly data [ppm] for five years aggregated to 0.5 deg grid cells. b) Correlation observed between XCO2 anomaly and AOD.

P6L167: The dust loads may be identified by the areas, but "biomass burning aerosols increase AOD in central Africa and South-East Asia" lack of evidence.

Reply: We think that the high AOD values in these areas are often contributed to biomass burning, see e.g. van der Werf et al. (2010), Myhre et al. (2009), or Kinne (2019). As this is not in the focus of the manuscript, we have removed the reference to biomass burning aerosols.

References:

Van der Werf et al., 2010, www.atmos-chem-phys.net/10/11707/2010/

Myhre et al., 2009, www.atmos-chem-phys.net/9/1365/2009/

Kinne, 2019, https://doi.org/10.1080/16000889.2019.1623639

P6L172: I think the largest differences appear in Central Asia and South Asia.

Reply: Our definition of 'South-East Asia', as shown in Fig. A7, is rather loosely defined. It covers areas usually understood as South Asia, East Asia and parts of South-East Asia.

We have reformulated this as "The largest differences in AOD appear to be concentrated largely in the high AOD areas in parts of Asia, where OCO-2 AOD is lower than MODIS AOD."

P6L173: "These positive difference values are related to the MODIS DT algorithm permitting small negative AOD values." Any references?

Reply: The negative MODIS AOD values are discussed in Sayer et al. (2014). This reference was given after a couple of sentences on line 176. We have moved the citation to line 173.

P7 Fig.1: The colorbar of panel (b) should be adjusted because 0 and invalid data are both white.

Reply: We have duly adjusted the color for missing data in Fig. 1 b).

P8L187: You should point out earlier.

Reply: The discussion on the wavelength issue is lengthy, which is why it was delayed until this point. We added a note on it at the beginning of section 4.1, with reference to later discussion.

P8L200: You acquired 770 nm AOD by averaging 675 nm AOD and 870 nm AOD. Can you really achieve that by the simple average? Please give more evidence or reference.

Reply: A simple average is not the optimal way to scale the wavelengths but should be acceptable in this case. Using scaling with Ångström exponent with a smaller subset of data did not change the results dramatically. Please see comments to Referee #2 for more detailed discussion.

We have added a note on this to section 3.3.

P8L210: How did you define urban areas?

Reply: The urban areas are discussed from line 329 on, where references are given. We have added a note on this earlier to L210.

P9L217: Please point out the spatial resolution of OCO-2 again.

Reply: We have added the OCO-2 resolution (approx $1x2km^2$) to the text here.

P10L250: How did you handle with the AOD exactly equal to 0.2?

Reply: There were practically no data points with AOD exactly at 0.2. They are included in the larger AOD bin. We have changed the text to replace "AOD<0.2" by "AOD<=0.2".

P13Figure 4 (b): You should distinguish 0 values and invalid values. Both of them are white.

Reply: We have duly changed the color for missing data.

P13L299: Table A3 or Figure A3?

Reply: The reference was incomplete. We have changed this to "(see Table A3)".

P14L316: What does "the measured CO_2 absorption is divided into too short distance" mean?

Reply: The measured radiance at CO2 absorption channels carries information of the total column load of CO2 in the observed light path. The inversion of measured radiance signal to CO2 column load (XCO2) requires information on the light path length. The light path length is affected by aerosols: more aerosol means effectively more scattering, and a longer light path. If the retrieved AOD is incorrect, the light path length is also incorrect, and this directly affects the retrieved XCO2 value. We have revised the text to state this more explicitly.

P15L340: Again, if you aquire OCO-2 XCO_2 anomaly by calculating the difference between the OCO-2 XCO_2 data and the median value along 500 km orbit, how can the anomaly be used to de-seasonalize and de-trend the data? Please describe more details about this.

Reply: The XCO2 anomaly values are used as such, instead of using them to 'correct' the XCO2 values. It is assumed that the large scale seasonal and trend effects are removed when subtracting the 500 km median, and that the anomaly contains information on the XCO2 variability in the local spatial scale, including local sources (and sinks). Please see the extensive reply to the previous similar comment above (reply to comment P6L148).

P15Table2: "TCCON(1)" should be revised as "OCO-2(1)"?

Reply: The naming refers to different subsets of OCO-2 data, hence 'global', 'urban' and 'TCCON'. To distinguish between the two data sources (OCO-2 and TCCON) for the collocated OCO-2 & TCCON dataset, we have used the numbers in parenthesis, (1) and (2). We have revised the Table caption to clarify this.

P16L365: "OCO-2 seems to slightly overestimate XCO2 for low AOD values, and underestimate at high AOD values."

-For panel (a) of Figure A8, you used MODIS AOD and XCO2 difference (deviation of OCO-2 from TCCON). MODIS AOD and TCCON XCO2 can be considered as references, so you can definitely get the statement about overestimate or underestimate of OCO2 XCO2 at different AOD.

Reply: We have revised the text: "OCO-2 slightly overestimates XCO2 for low AOD values, and underestimates at high AOD values."

-For panel (c), it seems that the XCO2 difference is close to zero when the AOD difference is in the range of 0 to 0.1, and when the AOD difference becomes higher or lower, the XCO2 difference will be minus, which means the underestimate of OCO2 XCO2. It seems to be not totally consistent with the statement of Figure 5 "If the aerosol load is underestimated in the retrieval (Q2), the light path is also underestimated, and the measured CO2 absorption is divided into too short distance, leading to overestimation of XCO2. Similarly, if AOD is overestimated, the light path is also overestimated, causing underestimation of XCO2." Do you have any comments on it?

Reply: The discussion at Figure 5 was meant as the first impression and possible explanation of the results seen in Fig. 5 a), accompanied by the disclaimer that it is not easy to distinguish the two effects of co-emission of aerosols and CO2 and the aerosol bias on XCO2 retrieval. The discussion of the light path is a possible explanation, which turned out not to be consistent with the results obtained with collocated TCCON data, as correctly pointed out by the reviewer.

We find that it is difficult to draw detailed conclusions from Fig. A8 c), because there are two dependencies mixed in the plot. First, the AOD difference between OCO-2 and MODIS depends on the MODIS AOD in a nontrivial way as shown in Fig. 2, with OCO-2 low bias at one end and high bias at the other. Second, the XCO2 bias depends also on MODIS AOD. We find that Fig. A8 a) best describes what we can say about the XCO2 bias, but nevertheless we wanted to show Fig. A8 c) for completeness.

It is certainly important to state these points explicitly in the manuscript. We have revised the discussion to section 4.3 to achieve this.

P18L388: The XCO2 and AOD also show negative correlation in North America and Europe. Do you have any comments on this?

Reply: Here we wanted to highlight only the most obvious differences between the areas. In Fig. 8 c) the negative slope of XCO2 as function of AOD is most obvious for North Asia, but the linear fit slopes are negative also for Europe and South America (for North America the slope is slightly positive). Looking at Fig. 4 b), we see that in North America there are areas of negative and positive correlation, but the data is noisy. For South America the correlation is weaker, and close to zero in large areas.

Actually, most of the areas have negative correlation (when looking at the matrix plots, not shown in the manuscript), with the exception of Africa and 'South-East' Asia.

We have reformulated the text.

P19L395: "Satellite XCO_2 retrievals are known to have higher uncertainty in high aerosol conditions." Where is this statement summarized from? From Figure A8 (a) and (b), it can be seen that the OCO-2 XCO2 also has deviation from TCCON XCO2 at low AOD.

Reply: This sentence refers to earlier research in the field; we have added references to Connor et al. (2016) and O'Dell et al. (2018). Figure A8 a) and b) are plotted using quality filtered OCO-2 data, where high AOD cases (OCO-2 AOD over 0.2) have already been removed. Unfiltered OCO-2 data would show a larger spread (see Fig. 2 below). However, it is true that there is considerable spread in the data also at the low AOD region.

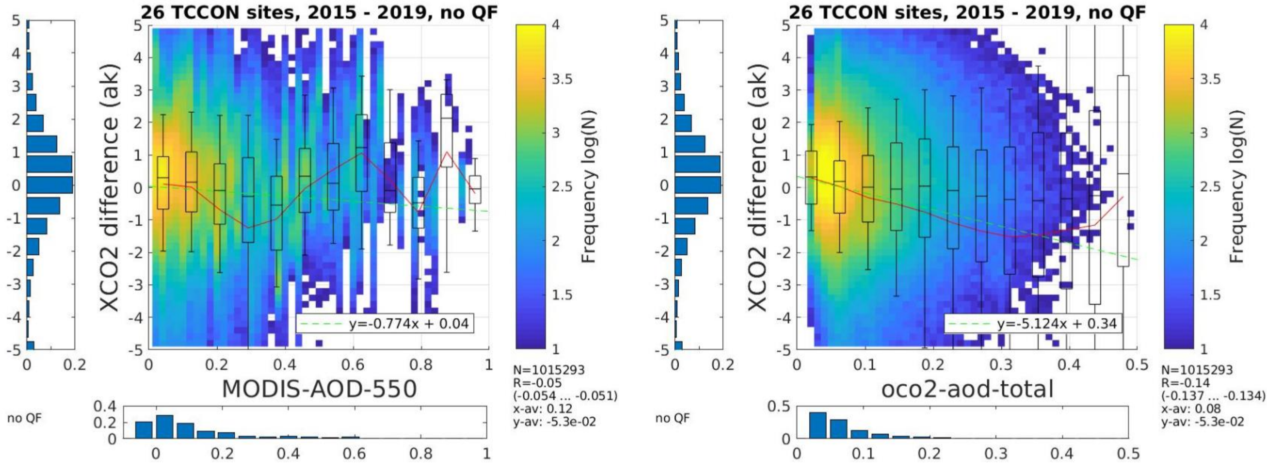


Figure 2. Same as Fig. A8 a) and b) in the manuscript, but without OCO-2 quality filtering.

P21L427: For future CO2M, you found the large increase of data coverage after adjusting AOD threshold from 0.2 to 0.5. After adjusting, there will be more underestimated and overestimated data when the AOD is below 0.5, i.e., the data in Q1 area you defined will include data in Q2 and Q4 areas. You should discuss how to conduct quality filter for CO2M, or the data quality with the AOD threshold of 0.5.

Clearly the discussion in section 4.5 was not sufficiently clear, as pointed out by all referees. The idea here is that the coming CO2M mission will have a dedicated aerosol instrument - Multi-Angle

Polarimeter (MAP) – which is expected to allow XCO2 retrieval at heavier aerosol conditions, with AOD threshold of 0.5.

Here we use the collocated, quality filtered OCO-2/MODIS dataset as a proxy for CO2M data. This dataset includes high MODIS AOD pixels, although the OCO-2 quality filter including an AOD threshold of 0.2 has been applied. We assume that the OCO-2 quality filtering assures that the XCO2 data is of good quality even for higher MODIS AOD cases, as CO2M data is expected to be up to AOD of 0.5. We further assume that the MODIS AOD in the collocated dataset is representative of ‘true’ AOD and can be used to study the AOD threshold, even though the OCO-2 quality filtering has removed a large part of the original pixels.

With this collocated data set, we can test what is the effect of relaxing MODIS AOD threshold from 0.2 to 0.5. Note that this does not mean that we extend the OCO-2 coverage; the MODIS AOD threshold used here is an additional constraint on the quality filtered OCO-2 data.

In the manuscript we did not address the effect of different MODIS AOD thresholds on the XCO2 retrieval quality. Figure 3 below shows such assessment: We find that in this case the MODIS AOD threshold did not have any significant effect on the XCO2 retrieval quality. But as mentioned, in these cases we use the quality filtered OCO-2 data, which should be of sufficiently good quality. Having said that, we need to clarify that the purpose of this study was not to improve the capability of OCO-2 to deal with higher aerosol loads, but to find how increasing the AOD threshold affects the coverage.

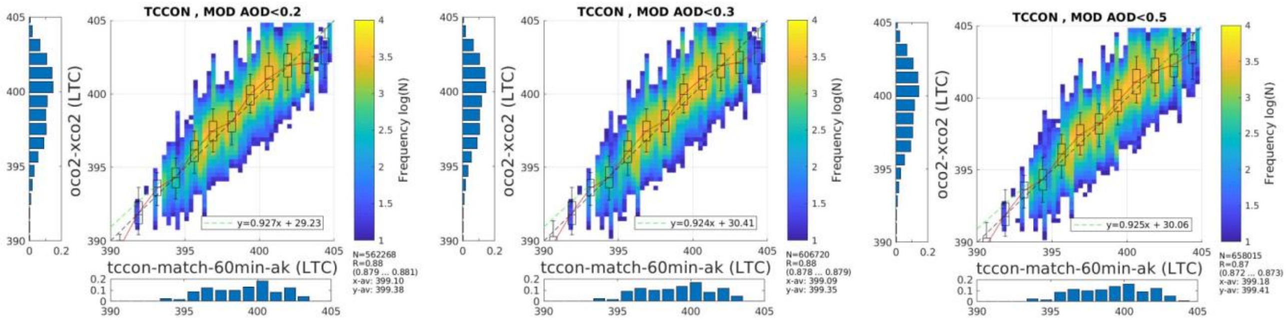


Figure 3. Effect of MODIS AOD threshold on XCO2 comparison between OCO-2 and TCCON. The OCO-2 quality flag has been applied.

We do not propose replacing the OCO-2 quality filtering with a single MODIS AOD threshold. The OCO-2 quality filter is based on 32 different tests, of which the total AOD threshold of 0.2 is only one component. Replacing this with a MODIS AOD threshold, even as stringent as 0.2, would result in reduced XCO2 quality, as shown in Figure 4 below.

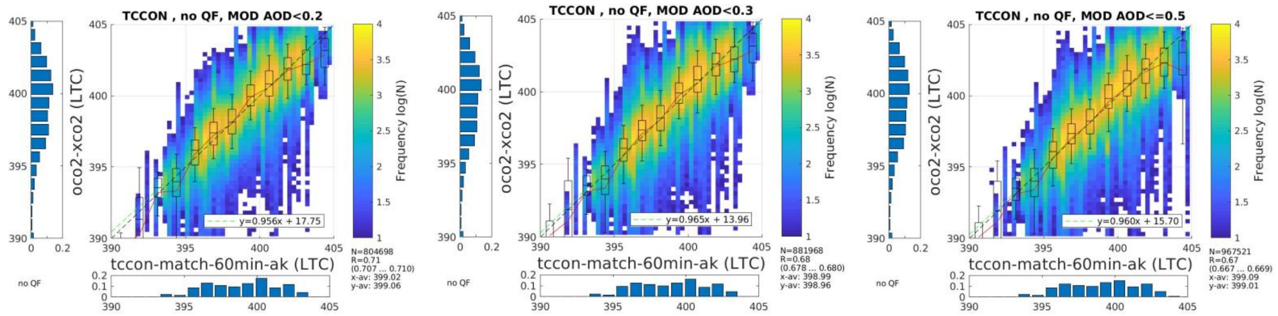


Figure 4. Effect of MODIS AOD threshold on XCO2 comparison between OCO-2 and TCCON, when the OCO-2 quality flag has not been applied.

In conclusion, here we treat the quality filtered OCO-2 data as a proxy of the coming CO2M data, which can be further filtered by using MODIS AOD thresholds. Our main finding, then, is that if CO2M can handle AODs up to 0.5, this will increase coverage, particularly in urban areas, compared to a case where AOD only up to 0.2 could be allowed.

We have clarified the text in section 4.5 to bring these points explicitly out.