

We thank the editor and additional reviewer for the detailed evaluation of the paper. We have gone through and addressed their comments as described below. All editor/reviewer comments are presented in italic font while the author responses are displayed in standard font.

Your revised manuscript addresses the main points raised by the reviewers, and is therefore acceptable for publication in AMT. However, the manuscript is very lengthy, and I find it tough to go through, even while being familiar with the topic. Therefore I strongly recommend that you shorten the manuscript wherever possible. There is quite a bit of repetition in the text, for example when you extensively quote the outcome (including uncertainties) for all the comparisons, when these can also be seen from the tables or figures. The text reaches a total of almost 1000 lines, and I think this can and should be reduced by at least 20%, so that readers can more quickly digest the essence of your work. The work basically does not justify 1000 lines. One of the reviewers, while positive about your revision recommended a useful opportunity to make the paper more concise: "I'd suggest the authors combine the discussions of biases and uncertainty, so that readers could have a wholistic picture about the performance of each satellite retrieval".

The authors have done our best to reduce the length of the text. From the title to the end of the Data Availability section was 988 lines and is now 801 lines. This is a 19% reduction in length. We have tried to reduce the repetition in the text as much as possible and have combined the systematic bias and uncertainty discussion into single sections.

One other reviewer had two useful comments, which should also be addressed:

(1) The abstract focuses very much on absolute bias numbers while the analysis (e.g. the regression slopes far from unity) indicate that these have limited meaning (many of the effects being multiplicative). Perhaps consider reporting %-biases or slopes (instead or additionally)?

We now discuss the systematic bias and uncertainty value primarily using normalized mean bias percentage and linear regression slopes in the abstract.

(2) Section 2.4: Agreed on the importance of the assumed vertical profile in the satellite retrievals, and the value of homogenizing these, but what about the profile used in the retrievals of the GeoTASO and GCAS tropospheric columns? In 3.3, you write "GeoTASO and GCAS retrievals were not reprocessed in order to have a consistent reference data set for satellite evaluation." I do not understand that argument. Why would reprocessing them mean they are no longer consistent (with what)? There would of course be a concern about independence, but I believe it would have been a meaningful exercise nonetheless, unless I miss a strong argument against it. Actually, unless I missed it, little detail is given about the profile that underpins the airborne retrievals. Some text on this would be appreciated.

The text in this section has been updated to “GeoTASO and GCAS retrievals were not reprocessed in order to have a consistent reference data set for the evaluation of the standard and reprocessed satellite retrievals. While reprocessing the airborne data with the higher spatial resolution model output would in itself be interesting (as done in Judd et al., 2020), the direct evaluation of the improvements in the reprocessed satellite data compared to the standard

retrieval would not be possible.”. We wanted the GeoTASO and GCAS data to remain constant during the evaluation of both the standard and reprocessed satellite products to determine the potential improvements in the reprocessed satellite data. Thus, we refer to this as keeping the GeoTASO/GCAS consistent.

Specific comments from the editor:

Throughout the manuscript you use two words for many concepts, where only one should be used. Example: on line 30 you state "clean/background". Just use "clean". There many such cases.

We now use the term “clean” throughout the manuscript.

L34: this sentence should be phrased such it makes clear that the number is a campaign average NO2 uncertainty. Now it could be interpreted as a generic finding, which is not the case. Same for the HCHO uncertainty on L42-43.

We have identified these uncertainty values as campaign-averaged in the abstract.

The abstract is too long - I suggest to shorten it by 30%.

The abstract has been reduced by ~35%.

L141: the equator overpass time of OMI is 13:30 hrs.

The equatorial overpass time for Aura-OMI is 13:45

(<https://earth.esa.int/eogateway/missions/aura/description>, <https://aura.gsfc.nasa.gov/scinst.html>, <https://earth.esa.int/eogateway/missions/aura>).

L167: please double check that the NASA NO2 algorithm uses the OMI O2-O2 cloud parameters. They have been using the Raman cloud product from OMI (Joiner-papers) in the past, and I doubt that this has changed. Regardless I think the O2-O2 product citation should be Acarreta (2002) or Veefkind (2016), not Vassilkov.

Originally, the O2-O2 algorithm was developed by KNMI, and the representative papers are in fact Acarreta et al. (2004) and Veefkind et al. (2016) (KNMI product is named OMCLDO2). Vasilkov et al. (2018) produced an OMI cloud product using the O2-O2 algorithm and named it OMCLDO2N (or OMCLDO2N). The previous version of the NASA NO2 product (V3.1) used OMCLDO2, but then they replaced it with OMCLDO2N (NASA version 4.0 used in our study). In short, NASA produced OMCLDO2N for use in OMI (Vasilkov et al., 2018), which is broadly similar to the OMCLDO2 but has a number of differences.

We have added the reference to Acarreta et al. (2004) and Veefkind et al. (2016) when we first discuss the O2-O2 cloud parameter.

L224: it is useful to remind the scientific community that v2.3 of TROPOMI NO2 has indeed improved compared to previous versions. The main reason for this is -better cloud pressures via FRESCO+ wide- explained and demonstrated in Riess et al. (2022), and it would be appropriate to cite that study here (along with van Geffen et al. (2022)) here.

See our response to the comment about L442-444. The study by Riess et al. (2022) is now referenced.

L341: qa_values should not be zero or < threshold, but > threshold. Please correct

The wording for this discussion has been updated accordingly.

L360-367: these lines can be removed

This opening paragraph has been removed.

L442-444: here the main reason why v2.3 has improved could be brought up as explained in Riess et al. (2022)

The text has been updated to read: “It should be noted that the TROPOMI low bias in tropospheric column NO₂ is improved with the newer retrieval algorithm used in this study compared to early versions of the data product (e.g., v1.2.2 had a campaign-averaged median low bias of $-1.3 \pm 4.0 \times 10^{15}$ molecules cm⁻²) primarily due to better cloud pressure input data (FRESCO+ wide) now used in TROPOMI retrievals (Riess et al., 2022)”.

L445: that NO2 (and all spaceborne tropospheric DOAS retrievals) have a magnitude dependence is well-known since the early works of Martin et al (2002) and Boersma et al. (2004). Those studies showed that the uncertainties in the AMF drive this magnitude dependency.

This has now been acknowledged with the following text: “This magnitude dependence has been shown to be driven by uncertainties AMF values used in the retrievals (Martin et al., 2002; Boersma et al., 2004).”.

Caption Figure 6: is this Figure showing the FNR derived from the NASA scaled or NASA standard products? This should be clarified in the caption.

This figure caption now states “Campaign-averaged relative error in FNR products from standard NASA OMI, QA4ECV OMI, and TROPOMI retrievals ...”

References

Riess, T. C. V. W., Boersma, K. F., van Vliet, J., Peters, W., Sneep, M., Eskes, H., and van Geffen, J.: Improved monitoring of shipping NO₂ with TROPOMI: decreasing NO_x emissions in European seas during the COVID-19 pandemic, Atmos. Meas. Tech., 15, 1415–1438, <https://doi.org/10.5194/amt-15-1415-2022>, 2022.

Please go through the manuscript one more time, address the above comments, and make the paper more concise, and then resubmit. Thanks for an informative paper.