

Referee report to the “Five years of Sentinel-5p TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks” manuscript by Arno Keppens et al.

The manuscript describes the retrieval algorithm used for the operational retrieval of ozone vertical profiles from Sentinel-5p TROPOMI measurements. Sensitivity of the retrievals and uncertainties of the results are well described. The retrieved ozone profiles are validated using the ozone sonde and lidar measurements. The topic of the manuscript is of a high scientific importance for ozone data users and is well suited for publication in AMT. In the first part of the manuscript authors describe the instrument, algorithm, data processing chain, quality assurance methods and reference data in much detail, very concise and clear. This part of the manuscript is written almost perfectly. Unfortunately, in Section 4, which describes the retrieval sensitivity and validation results, authors decided to save the space and joined tens of panels with thousands of curves into single plots. This makes the plots absolutely unreadable (Figs 9, 10, 11 and A2). Although the text is still good written, the reader cannot follow the discussion and verify the conclusions of the authors because no information can be read from the plots. In my opinion, the authors have to reconsider the way how they present their results in Section 4 to make the manuscript suitable for the publication. It should be analyzed which information is important and how to present it without making the plots looking like squeezed together unresolved color spots. A rotation of the plots by 90 degrees is also not a really good idea. This makes the reader terribly difficult to look at the plots when reading the text.

Detailed comments

- The paragraph mentioning previous retrievals (lines 505–516, Section 4.7) should be moved to the introduction.
- Figure 1: the right box shows an endless loop between the radiative transfer and optimal estimation.
- Line 125: for the pixel size after August 6, 2019 please indicate which size is cross-track and which along-track.
- Line 147: Usage of the CAMS data, with assimilated MLS profiles, and scaling them to OMPS total columns is expected to give a very good approximation for the ozone profile. In this respect author should put a bit more focus to show that TROPOMI measurements increase the information content in comparison to a priori. Maybe it is already shown in plots in Section 4, they are, however, completely unreadable for me.
- Line 163: “computed by combining the black and light gray points of the same year” - There are no black points in the plot, I see two sets of grey points. It is unclear why there are two curves and how they are combined to get red points.

- Section 2.2.2: it is unclear how the cloud fraction is included into the forward model.
- Line 179: “For all state vector elements, the OE method requires an a-priori value and its error estimates.” - I think “a priori values and their error estimates” would be more correct
- Line 199: “when the measurement errors are assumed to be uncorrelated” - it is unclear if it is the case in this study
- Eq. (2): Can this filter result in skipping real ozone profiles, which are strongly different from a priori, e.g. unexpected ozone loss?
- Line 258: “The full width at half maximum (FWHM) of the AK corresponding to a given altitude is selected in this work as an indicator of the effective vertical resolution of the retrieved profile at this altitude” - Please add here a note, that this approach ignores the displacement of the AK maximum (which you treat then independently as offset).
- Line 258: “the true, physical resolution” - if I understand the meaning of the sentence correctly, the comma needs to be deleted.
- Line 337: “retrieval differences and vertical sampling and smoothing differences” - the wording seems to be incorrect
- Figure 6: A relative difference should be provided in addition.
- Line 359: “cost function > 200” - This criterion was not properly described and justified in Section 2.2
- Figure 9 is unreadable. The panels are too small. There are too many panels in the plot. A rotation by 90 degrees makes the plot even less readable.
- Line 388: “Yearly drifts are added to the temporal dependence plots in Figure 9.” - the meaning of this statement is unclear. Are you correcting some data for the drift?
- Figure 10 is not referenced, it is unclear what is the difference between Fig 9 and 10. Figure 10 is unreadable similar to Figure 9.
- Figure 11: The scale of the figure does not enable the reader to estimate the values. The figure cannot definitely be read by people with color vision deficiencies. There are too many panels in one plot. The white dashed line is almost invisible. It cannot be read from the plot if the retrieval compares better than a priori.
- Section 4.4: please state clearly whether the comparisons are done convolving the reference data with the averaging kernels of TROPOMI retrieval.
- Line 431: “... has a mean bias below $\pm 5-10\%$ in the troposphere...” - It is absolutely impossible to see if the mean bias decreases in comparison to the a priori

- Lines 447-450: It looks like this text refers to Figure 10. It is unclear why it is placed after the discussion of Fig. 11.
- Figure A2 needs to be moved to the main text as it is discussed here.
- Line 466: "... on average the observed differences confirm..." - with exception of the stratospheric lidars this is valid only above 20 km. Thus, this statement is not suitable on general.
- Line 466: Again, Figure A2 is discussed in the main text but placed in the Appendix
- Line 481: "All requirements are summarized in Table 2, with the compliance of the operational TROPOMI ozone profile product added" - It is unclear how the values for the table are obtained. From the ozonesonde comparison, for example, it is difficult to understand why authors claim that the accuracy between 12 and 18 km is below 5%.
- Line 486: "This can be seen from Figure 10, with the black lines (average differences) being within the grey areas (SRD requirements)." - I cannot see anything from Figure 10 but this is most probably because of the quality of the figure.
- Line 501: "... typically amounts to about 5%, meaning ..." - From the color scale used in the plot it is difficult to read whether the values in maxima exceed 5%.