

General comments: This work provides an intercomparison of tropospheric ozone column datasets from combined nadir and limb satellite observations. Although this intercomparison is of interest to the community and deserves publication into the TOAR-II Special Issue, its presentation in terms of scientific clarity and focus could be substantially improved. Providing a more consistent story that is less broad would increase readability and significance. “The overall goal ... to assess the consistency between the datasets and explore possible strategies to reconcile the differences between them” as phrased in the abstract does not seem to be fully (quantitatively) addressed, or the information is too scattered to be properly captured.

We thank the reviewer for the time spent on our manuscript and for his/her feedback. We reviewed the manuscript in light of this general comment, by making the overall structure more linear and consistent. For example, we reduced the content of the supplements, removed the analysis of global trends in Sect. 6 and focused the trend analysis only on quantile regression (QR). Specific comments are addressed below after each paragraph, in blue text; the suggested references were added to the paper. The manuscript has been accordingly modified. Line numbers in the answers refer to the updated manuscript.

Specific comments:

- Line 22: “due to the overlap in signals” is too vague.
We agree, this part of the sentence was deleted.
- Line 26: It would be appropriate to refer at least to ESA’s operational TROPOMI nadir ozone product (possibly next to scientific products).
Thanks, the reference to Keppens et al. (2024) was added.
- Line 56: Does “bias” here refer to bias from different TPH definitions or in general?
It refers to the bias due to TPH definitions, it was clarified in the text: “a method to correct the tropopause definition-related bias between time series is presented and assessed.” (line 60)
- Line 65: Does “its profile” refer to sensitivity or to stratospheric ozone?
It refers to the stratospheric ozone profile, it was re-worded in the text: “as the ozone profile generally starts to increase below the typical thermal tropopause” (line 70)
- Lines 95-96: Was this drift correction done by the authors, or by the data providers? Please elaborate, possibly with reference(s).
The correction was implemented by the data provider. This sentence was changed in the text: “To take into account an identified drift in OMI time series, this dataset was corrected by the data provider by adding a drift at a post-processing step, as described in Gaudel et al. (2024).”
- Line 98: “WMO thermal definition from NCEP reanalysis” requires explanation and references.
We better defined the “WMO thermal definition” of TPH at this point and included a reference for it. (Line 103)
- Figure 1: Ordering the plots north to south would seem more logical, but is not mandatory. What happened to the 20-40 °N band plot?
Thanks, it indeed makes more sense. We re-ordered the plots and slightly changed the latitude bands according to the TOAR II recommendations (the missing 20-40°N band is now included with the 20-30°N panel).
- Line 120 and following: Which global mean has been used as a reference for de-biasing?
We used the multi-instrument mean. This was added in the text, line 128.

- Note sure whether the more qualitative Figure 2 adds a lot to 1 and 3?
This is a good point, but we find interesting to show the TrOC patterns among the data sets and for different seasons to assess not only their latitude-averaged values but also their geographical ozone distributions.
- Line 164: Explain and explicitly indicate the “Quasi Biennial Oscillation signature”
The reviewer is right, as it is not visible from this plot. We provide here a time series in the tropics, where the correlation of the satellite measurements with the QBO signal is remarkably high, as an example. This correlation was found to be higher for OMI-LIMB and GTO-LIMB than for OMI-MLS. This plot has been added in Fig. S7 (modified).

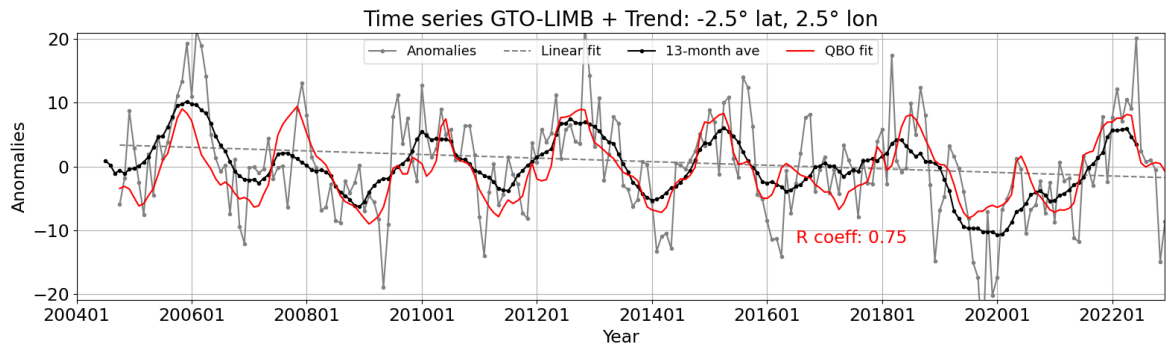


Figure 1: GTO-LIMB time series in one lat-lon bin in the tropics, with its 13-month running average. The red line is the QBO contribution after fitting it using a MLR. The correlation coefficient is calculated between the fitted QBO and the 13-month average line.

- Lines 168-169: “plays an important role in the biases between them” – How do you know, and doesn’t this at least partially contradict conclusions made later, e.g. on line 178?
Thanks for pointing this out, we replaced it with “it may play an important role...”
- Line 183: “As reference TPH, the ERA5 dataset was selected.” This must refer to a different dataset than the monthly gridded ozone profiles mentioned in the previous sentence.
It refers to the same ERA5 data set: in the previous sentence we describe the usage of ERA5 ozone profiles to compute the ozone column gaps due to the difference in TPH between the satellite dataset and ERA5, which is the reference value.
- Line 186: “We subtract the mean column gaps...” Is this the global mean, or determined for each latitude-longitude bin?
This is determined for each latitude-longitude bin.
- Figure 7: More of interest than the trends themselves seem to be the differences in the trends due to different TPH definitions? The latter could be given more attention.
This is an interesting point of discussion, we expanded the discussion in the manuscript, lines ~213-217. However, we think that the more relevant part is the TPH trend themselves, as they can have a direct impact on the TrOC trend from these datasets. The differences between the TPH are indeed related to different TPH definition, sampling of the data set or different reanalysis.
- Sections 5 and 6 refer to figures and tables that are distributed over the main text, appendices, and supplement, which hampers a fluent appreciation of the research and results by the reader.
Thanks for pointing this out, we agree with the reviewer and implemented some changes in Sections 5 and 6 to reduce references to Appendix and Supplements (which were also reduced).

- Line 220: Possibly briefly explain the difference between drift and trend studies? It doesn't really help calling the drift plots in Figure 8 a "trend in DU per decade of the differences" and "trend values (drifts)"

We introduce the concept of drift at line 238. The drift assessment involved computing a linear trend of the differences. In the caption of Fig. 8 we now say "Drift in DU/dec of the difference time series between satellite and HEGIFTOM sonde anomalies".

- Equation (3): Where does this come from? Is this a fully arbitrary choice, or is this based on common approaches in the literature?

Thanks, we re-worked this section of the manuscript, moving this vertical weighting to Appendix A and added a reference (Sofieva et al. 2022) for the OMI AK.

- Line 231: Does "collocated" mean containing the station location?

Yes, we replace "collocated" with "containing".

- Line 241: "with negative trends until around 2014 and positive trends in the last 10 years" Despite talking about drifts here (also see above), this is really not clear from Figure 8, especially as this figure only contains linear fitting to the full time series.

Thanks, this has been changed, saying that the positive drift affecting most datasets in the tropics is larger during the last 10 years. However, the similarity between the SAT-HEGIFTOM (colored lines) and -HEGIFTOM (black lines) anomalies indicates that the patterns in sondes are not captured by the satellite datasets.

- Line 252: "trends of +(1-4)%/decade" over which period?

We expanded this paragraph with a more detailed description of some studies at the beginning of Sect. 6.

- Figure 9: Do the regions defined here correspond to those agreed upon within the TOAR-II initiative?

The TOAR II guidelines provide several options for the choice of regions, such as basic lat/long boxes around continents, HTAP regions, GBD regions and IPCC regions. We rather investigated the correlation of the seasonality between the data sets to define the specific regions, similarly to what was proposed by Van Malderen (2025).

- Line 268: Can you provide info and references on the multivariate regression model that is being used?

We decided to remove the comparison between MLR and QR models and focused only on QR including proxies. This brought to a simplification of Sect. 6.

- Despite the anticipated focus on geographical regions in Section 6, Figure 11 extends to the global anyways. Possibly, limit this work to one of both only?

The reviewer is right, we decided to remove this figure from the manuscript, to focus only on specific regions, and just point out in the manuscript the presence of larger differences when looking at the global picture.

- It would be very insightful to add an estimate (zonally) of the fraction of the trend that could be explained from the TPH trends in Figure 7?

Thanks, in the trend discussion we introduced such a estimation, which helps the assessment of the confidence of the trend values.

- Line 298: "reconcile the discrepancies ... rather than highlight the inherent differences" sounds conflicting. This requires some explanation of the distinction between both terms...

With this sentence we meant to say that we didn't focus only on the discrepancies between the datasets, which have been presented for example in Figs.1 and 4, but rather on ways to

find similarities between the time series, in terms of climatological patterns (Figs. 2 and 3) and trends (Fig. 10). We modified this sentence by deleting “reconcile the discrepancies”, to sound less contradicting.

- Line 319-320: “Our analysis shows that the homogenization is a crucial step before using the datasets for global trends studies.” This is not what seems to come out of this work. Do you mean a global bias correction, a TPH correction, calculation of anomalies, or the ozonesonde homogenization for comparison? Both what this sentence is referring to as well as how this can be concluded from the analyses is very unclear, and makes this conclusion inappropriate. We agree that this sentence is confusing and sounds not well supported. We modified it as follows: “Our analysis shows that a better understanding of drifts and biases is a crucial step before using the datasets for global trends studies”.

Technical corrections:

- Often inconsistencies occur between singular versus plural nouns and verbs. Please verify throughout the text.
Thanks, we performed a review of the English.
- Line 9: “morphology” does not seem to be the right term here, or should be explained.
We replaced “morphology” with “distribution patterns”.
- Line 91: Rephrase “a DU dataset”
“a DU dataset” → “a dataset in DU”.
- Line 105: Add monthly temporal resolution?
We added this information at line 110.
- Lines 108-117: “overall” is used quite often, sometimes inappropriately, i.e., conflicting with other parts of the sentence(s), e.g. “zonal averages” in line 108.
Thanks, we reviewed the usage of “overall”.
- Caption of Figure 4: “De-seasonalized anomaly time series...” ?
We removed “anomaly”.
- The color scale of Figure 6 (b) is hardly visible in print.
Thanks, the scale has been improved.

Additional references

Arosio, C., Rozanov, A., Malinina, E., Weber, M., & Burrows, J. P. (2019). Merging of ozone profiles from SCIAMACHY, OMPS and SAGE II observations to study stratospheric ozone changes. *Atmospheric Measurement Techniques*, 12(4), 2423-2444.

Keppens, A., Di Pede, S., Hubert, D., Lambert, J. C., Veefkind, P., Sneep, M., ... & Zehner, C. (2024). 5 years of Sentinel-5P TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks. *Atmospheric Measurement Techniques*, 17(13), 3969-3993.

Sofieva, V. F., Hänninen, R., Sofiev, M., Szelağ, M., Lee, H. S., Tamminen, J., & Retscher, C. (2022). Synergy of using nadir and limb instruments for tropospheric ozone monitoring (SUNLIT). *Atmospheric Measurement Techniques*, 15(10), 3193-3212.

Van Malderen, R., Thompson, A. M., Kollonige, D. E., Stauffer, R. M., Smit, H. G., Barras, E. M., ... & Sussmann, R. (2025). Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000–2022) from the TOAR-II/HEGIFTOM Project. *Atmospheric Chemistry and Physics*.