The study by Arosio et al. uses multiple satellite products of tropospheric column O3, derived from limb and nadir sounders, to investigate long-term trends in tropospheric ozone. Overall, this is a nice study and provides useful updates on tropospheric ozone trends, especially as part of TOAR-2. My main question is on Section 5 and Appendix A where the authors use a proxy of any OMI averaging kernel (AK) to investigate the impact of vertical sensitivity on comparisons to other datasets (i.e. HEGIFTOM) and long-term trends. I also have several minor comments listed below. Once these have been addressed, the manuscript is suitable for publication in AMT.

We thank the reviewer for the time spent on our manuscript and for his/her feedback. The comments are addressed below after each paragraph, and marked in blue; the suggested references were added to the paper. The manuscript has been accordingly modified. Line numbers in the answers refer to the updated manuscript.

Major Comments:

The authors attempt to investigate the impact of satellite averaging kernels on tropospheric column ozone trends by applying an approximation of an OMI AK to the ozonesonde data. So, firstly, the function in Equation 3 is based on what? Just the approximate shape of an OMI AK or a peer-reviewed study? Depending on the satellite product used (i.e. UV-Vis vs. IR or DOAS vs. optimal estimation), the shape of the AK profile can change substantially. So, would it not be worth trying to simulate the impact of multiple types of AKs? Also, to support your choice of Eqn 3, could you plot some actual OMI ozone Aks?

The reviewer addresses an important discussion topic and a weak point in the manuscript. Although interesting, the application of instrument-specific AK goes beyond the scope of this manuscript. We decided to restructure this part of the paper by moving the discussion related to the weighting of the ozonesonde profiles to resemble the sensitivity of nadir instruments to Appendix A. As a consequence, we left the drift assessment in Sect. 5 with the original TrOC time series provided by HEGIFTOM. In the Supplements we included a plot (corresponding to Fig. 1 below) of a typical AK from the OMI instrument, and added the Sofieva et al. (2022) reference. We slightly changed the former Eq. 3 (current Eq. A1) to better approximate the AK shape. This equation is a crude approximation of the typical OMI vertical sensitivity.

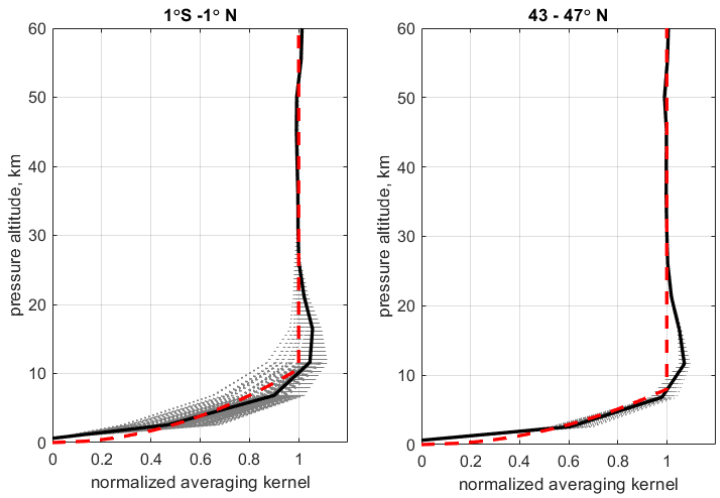

Figure 1: OMI individual averaging kernels normalized to 1 above 25 km (grey lines) for clear-sky conditions on 1 July 2008, the mean averaging kernels (black lines) and their crude approximation (red dashed lines). Left: latitudes [1°S, 1°N], Right: latitudes [43°N, 47°N].

Minor Comments:

Line 11: State which time-period the Southeast Asia trend is.

Sure, we stated that the trend is for the 2005-2021 period.

Line 17: Should the x in NOx be subscript?

Right, thanks.

Line 43-44: You state that ozone trends in Europe and North America has stabilised. This has been supported by several recent TOAR-2 studies (e.g. Pope et al., (2023 & 2024)), so might be worth mentioning them to support this statement.

Thanks, we included the suggested references in the introduction, line 46.

Line 50-52: Gaudel et al. (2018) identified large scale discrepancies between satellite product tropospheric ozone trends and you state that more work is needed to try and reconcile these. However, some TOAR-2 studies have attempted this, would be good to cite those (e.g. Gaudel et al., 2024 and Pope et al., 2024).

Thanks for the additional information, we included this in the manuscript, lines 55-56.

Figure 2: Could you add the global average trop O3 column average for each panel? While you can see the differences by eye, having a metric (e.g. average +/- standard deviation) above each map would add a useful overview for the reader.

We added this information only for the top panels, as the datasets in this plot have been debiased to the same multi-instrument mean. As a consequence, we provide in the figure a mean value with standard deviation for each season.

Equation 1: It is not clear what this is. Can this be defined?

We tried to clarify it by removing the subscript j and leaving only t, and modifying the explanation. The sentence reads now: "where Nm is the number of available monthly mean values TrOC(t) for each specific month of the year m, e.g. January, in each time series."

Line 136: What does m represent? It is not overly clear.

With m we indicate the generic month of the year, e.g. January.

Line 136-138: What are the offsets based on? Is this relative to the ozonesondes or the satellite ensemble average? Same for "after debiasing" in the Figure 3 caption.

Thanks, we specified in the text that the offsets are relative to the satellite average ensemble in both cases.

Line 144: "in spring time"…is this the boreal or austral spring?

We changed the text to "austral winter (JJA)".

Line 148: To make this clearer, could you add an example. E.g. (e.g. "where m indicates the month (e.g. January) of the year and tm all months m (e.g. all Januarys) in the time series").

Right, we included such an example in the description of Eqs. 1 and 2.

Figure 6e: What is causing the large standard deviation at approx. 20-30N?

We think the reviewer is here referring to the low standard deviation values found after the TPH correction in the northern sub-tropics. We verified that in this latitude band the TPH correction actually performs well, bringing together the mean values of the data sets. This indicates that at least in this region the biases between data are mostly caused by TPH discrepancies. We slightly changed the text at lines ~202 to take into account this comment.

Line 213: What do you mean by "after the harmonization of their time series by the HEGIFTOM"?

Right, this was a confusing statement and it was removed. We meant that the ozonesondes were harmonized by the HEGIFTOM working group.

Line 214: What data source is the thermal tropopause level based on?

We included in the text that the TPH was calculated, by the HEGIFTOM working group, using the temperature profiles from ozonesondes and the WMO thermal tropopause definition.

Equation 3: Add a reference, if appropriate, for this choice of function to represent the AK.

This was moved to Appendix A. We included in the Supplements (Fig. S9) a plot of the typical averaging kernels from OMI and the Sofieva et al. (2022) paper reference.

Line 243: Excess space between "last" and "10 years".

Sorry, we don't see that, probably a matter of spacing within the line.

Lines 244-246: I did not follow this text. This point probably needs more explanation.

The reviewer is right, we moved this section to Appendix A to better explain the point and expanded the discussion at line 257-258.

Lines 251-253: Would be useful to add the time-periods the authors derived trends for.

Thanks for the tip, we included the time frame of each study and expanded this paragraph.

Line 284-285: "The positive trend from SCIA-OMPS in the Amazon is likely related to artefacts in the datasets at the very beginning and end of the time periods". Can you expand on this. Not clear what you mean be artefacts and why this would drive an unrealistic trend.

We removed the sentence, but we meant that the larger anomalies towards the end of the SCIA+OMPS time series can influence the trend estimations.

Line 291: You discuss "insignificant trends at the 95% confidence level". I might be wrong, but I believe TOAR-2 are trying to move away from such definitions.

Thanks for this. It is true and also following Owen Cooper's comment we moved away from this formulation and introduced the p-values of the trends, together with their confidence level.

Line 293: If the time-series are influenced by "positive drift", since you can quantify this, could it not be removed to leave the real trends?

This is a good point, however we find such a subtraction not straightforward. First, a thorough drift investigation would be required to better target specific regions. This could provide a better characterization of the drift but is beyond the scope of this manuscript. In addition, the presence of

artifacts in the tropics in the drift analysis (see lines 257-258) gives less confidence in such a direct subtraction of the drift values for trends in specific regions.

Line 343: Why chose the OMI-LIMB product for this example?

This was chosen as it is one of the time series providing the longest record. OMI-MLS or GTO-LIMB could replace it without changing the conclusions of the paragraph.

References:

Gaudel, A., Bourgeois, I., Li, M., Chang, K.-L., Ziemke, J., Sauvage, B., Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Smith, N., Hubert, D., Keppens, A., Cuesta, J., Heue, K.-P., Veefkind, P., Aikin, K., Peischl, J., Thompson, C. R., Ryerson, T. B., Frost, G. J., McDonald, B. C., and Cooper, O. R.: Tropical tropospheric ozone distribution and trends from in situ and satellite data, Atmos. Chem. Phys., 24, 9975–10000, https://doi.org/10.5194/acp-24-9975-2024, 2024.

Pope, R. J., Kerridge, B. J., Siddans, R., Latter, B. G., Chipperfield, M. P., Feng, W., Pimlott, M. A., Dhomse, S. S., Retscher, C., and Rigby, R.: Investigation of spatial and temporal variability in lower tropospheric ozone from RAL Space UV–Vis satellite products, Atmos. Chem. Phys., 23, 14933–14947, https://doi.org/10.5194/acp-23-14933-2023, 2023.

Pope, R. J., O'Connor, F. M., Dalvi, M., Kerridge, B. J., Siddans, R., Latter, B. G., Barret, B., Le Flochmoen, E., Boynard, A., Chipperfield, M. P., Feng, W., Pimlott, M. A., Dhomse, S. S., Retscher, C., Wespes, C., and Rigby, R.: Investigation of the impact of satellite vertical sensitivity on long-term retrieved lower-tropospheric ozone trends, Atmos. Chem. Phys., 24, 9177–9195, https://doi.org/10.5194/acp-24-9177-2024, 2024.

Sofieva, V. F., Hänninen, R., Sofiev, M., Szeląg, 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.