Response to Referee #2 comment on "Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000–2022) from the TOAR-II/HEGIFTOM Project" by Van Malderen et al.

Review of manuscript entitled "Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000-2022) from the TOAR-II/HEGIFTOM Project by R. Van Malderen et al.

This manuscript describes the homogenized 2000-2022 tropospheric ozone records (column, free tropospheric and lower tropospheric partial columns) archived as part of the TOAR-II/HEGIFTOM project. Data are from five ground-based networks; measurements from each network included in HEGIFTOM have be reprocessed using standardized procedures and records include requisite uncertainty estimates. The focus of this work is intercomparisons of mean values, seasonal cycles and trends among the many individual instrument records to demonstrate the usefulness and limitations of the data for direct analysis as well as a reference for model and satellite-based tropospheric ozone records.

The manuscript is very well organized, especially given the difficulties intercomparing so many individual records with different long-term and daily sampling, station distribution, instrument type and related vertical resolution, uncertainty, etc. The number of tables and figures are substantial but also very well organized, and appropriate for this type of comprehensive review of a large collaborative project and resulting data archive. The analysis is very thorough, the authors use/compare several statistical approaches, suggest explanations for resulting biases and trends, and include relevant references throughout the manuscript.

I have only a few comments, and otherwise editorial corrections and suggestions. I recommend publication after these minor comments are addressed.

Thanks a lot for your time and effort in reviewing our manuscript. And thank you for complimenting us with the organization of the manuscript, which was indeed a difficult task given the level of comprehensiveness that we aimed for with this "review" paper.

Comments:

Section 4.1.4: The authors present an analysis of the seasonal cycle changes over the record and discuss both the impact of Covid-19 and Bowman et al. (2022) results. Although not relevant to the trends, I think it would be very interesting to also compute the seasonal cycle over several years pre-Covid (say 2014-2019, which is 6 years matching 2000-2005) to see how this differs from the seasonal cycle change including the Covid years. One would assume this pre-Covid time period is a better comparison in reference to the Bowman et al. results.

This is a very good suggestion. We computed the differences in TrOC and FTOC seasonal cycles between 2000-2005 and 2014-2019. These figures are shown here below (Fig. R3) and will be included in the supplementary material as well. Again, no consistent TrOC phase change is found between both periods. The amplitude reduction is more modest (-6%) and less general (for 70% of the sites). The increase of the minimum annual TrOC values (at 65% of the sites) now contributes slightly more than the decrease of the maximum annual TrOC concentrations (at 55% of the sites). Also for the FTOC, the amplitude reduction (-3%, for 65% of the sites) is smaller than for the 2015-2022 period, with equal contributions from increasing minimum and decreasing maximum FT ozone column amounts.

From this analysis, we can conclude that the post-COVID-19 time period is responsible for about half of the amplitude reduction, without a noticeable seasonal cycle phase shift. This amplitude reduction can be mainly ascribed to a decrease of the maximum annual TrOC/FTOC concentrations (for 79%/85% of the sites) during the post-COVID-19 era. This is consistent with other observations of tropospheric ozone reductions during the COVID-19 period in NH Spring/Summer time series, mentioned in Section 4.1.3.

In the manuscript, we added: "To be more directly comparable with the Bowman et al. (2022) results, we also calculated the TrOC and FTOC seasonal cycle characteristics of the pre-COVID period 2014-2019 and compared those again with the 2000-2005 seasonal cycle (see Fig. S4). We found that, between those periods, the amplitude reduction is more modest (-6%) and less general (for 70% of the sites) than between the 2015-2022 and 2000-2005 periods. The increase of the minimum annual TrOC values (at 65% of the sites) contributes slightly more than the decrease of the maximum annual TrOC concentrations (at 55% of the sites). Also for the FTOC, the 2014-2019 amplitude reduction (-3%, for 65% of the sites) is smaller than for the 2015-2022 period, with equal contributions from increasing minimum and decreasing maximum FT ozone column amounts. From this analysis, we can conclude that the post-COVID-19 period is responsible for about half of the amplitude reduction between 2015-2022 and 2000-2005, without a noticeable seasonal cycle phase shift. This post-COVID-19 seasonal cycle amplitude reduction can be mainly ascribed to a decrease of the maximum annual TrOC/FTOC concentrations (for 79%/85% of the sites) during the post-COVID-19 era. This finding is consistent with other observations of tropospheric ozone reductions during the post-COVID-19 period in NH spring/summer time series, mentioned in Section 4.1.3, and reported in Ziemke et al. (2022)."

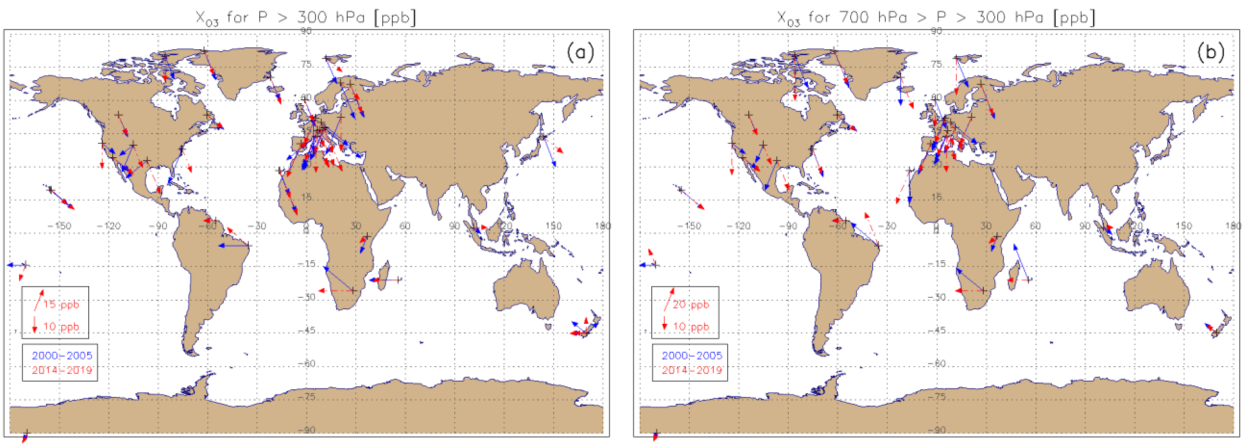

Fig. R3: Illustration of the mean seasonal cycle for the TrOC (a) and FTOC (b) time series for two different periods: 2000-2005 (blue) and 2014-2019 (red).

L551-555: In reference to the Law et al. "dipole effect" did the authors analyze the monthly trends in the Arctic ozonesonde records as done for other stations in Figure 17? It seems they should be able to further comment on the seasonal variation in the trend. I am not familiar with the study but note they say the "dipole effect" in the vertical tropospheric ozone. If by "vertical" it means this feature varies with altitude it may be more difficult to resolve from the partial columns but I was curious if this could be directly checked.

The "dipole effect" of the tropospheric ozone trends in the Arctic ozonesonde records (positive trends in winter and summer, and negative trends in spring and autumn), as discussed by Law et al., has a limited vertical variation: Positive winter (notably Jan.) trends are found up to 400 hPa at most sites (except Resolute and Sodankyla), and at Scoresbysund in early spring. Positive wintertime trends are more evident in the earlier period in the UTLS. Eureka, Resolute, and Sodankyla have periods with negative trends especially during spring and early summer in the lower troposphere (LT). Resolute decreases extend up to 500 hPa in March-April.

To further comment on the seasonal variation of the trends, we followed your suggestion and calculated the MLR monthly TrOC and FTOC trends (method described in Appendix A) for the Arctic sites (see Fig. R4 here below, and Fig. A2 in the manuscript). From these monthly trends, it is obvious why the overall trends are negative, except for Alert. As in Law et al., the largest negative TrOC and FTOC trends are evident in the springtime (MAM) for most of the sites (with 4 sites having trends significantly different from zero during one of those months). Also, for most of the sites, the winter (December, January) shows among the largest trends, but hardly reach positive values. In general, the mentioned dipole effect of the tropospheric ozone trends is not clearly present in the TrOC and FTOC time series considered here, and the different Arctic sites display different patterns of seasonal trends. For instance, Resolute has one of the more pronounced seasonality in the trends, with a peak in negative trends in the Spring (April) and a peak in positive trends in the Autumn (September and October). Possible reasons for the different seasonal signatures of the trends for the same ozonesonde sites between the Law et al. study and ours are the different time periods considered (1993-2019 vs. 2000-2022), the use of operational vs. homogenously reprocessed ozonesonde time series, and the calculation of vertical tropospheric ozone trends vs. partial column tropospheric ozone trends.

This discussion has been included in the Appendix and in the main part of the manuscript, we have added "In the appendix, Fig. A2, we calculated the monthly TrOC and FTOC 2000-2022 trends for the Arctic HEGIFTOM sites and found mostly negative trends, except for Alert, with the largest negative trends in springtime. We refer to the appendix for more details."

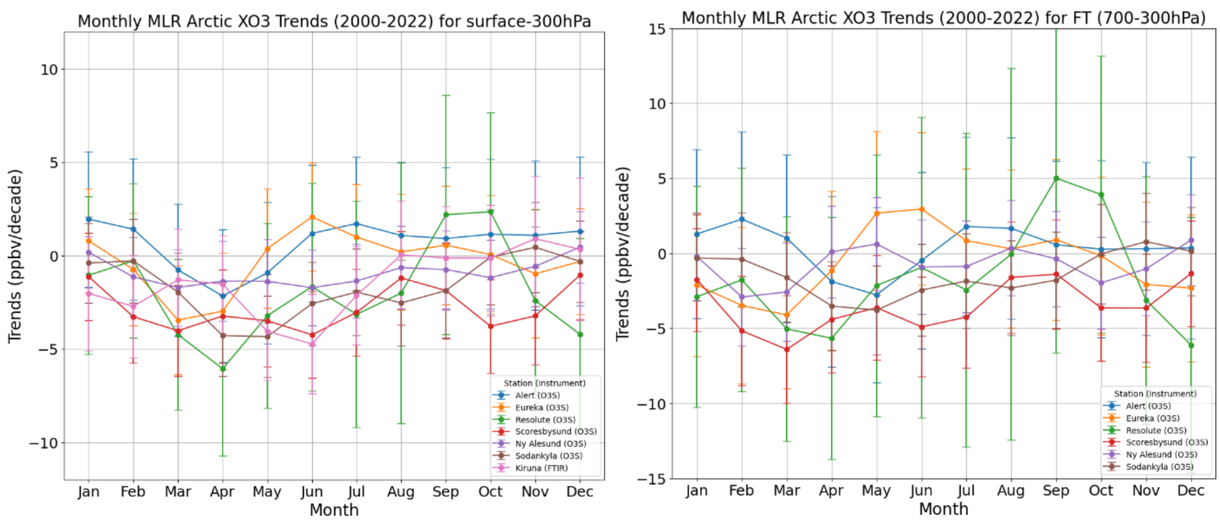

Fig R4: MLR monthly mean TrOC (left) and FTOC (right) trends for the Arctic HEGIFTOM sites.

Section 4.2.2: I did get a bit confused reading this section and trying to understand all the permutations. Most notably, the last two sentences seem to say the same thing. For Natal TrOC and LTOC "have increases" over the period but the FTOC increase in greater. The next example for several European stations says the same but "now with positive LTOC rates" but Natal LTOC was also increasing, so I do not see what distinguishes these cases.

We tried to clarify the structure of this section 4.2.2 by adding sentences like "We first consider the sites with a smaller trend in FTOC relative to TrOC.", "Second, we look at the sites with FTOC increases somewhat greater than TrOC, suggesting imported ozone above the boundary layer. We distinguish two different subsets here: ...", and "Another subset is made up of the European ozonesonde sites OHP, Hohenpeissenberg, and De Bilt, which …"

You are right, the statement about Natal is wrong. The LTOC relative trend is the largest, greater than the TrOC and FTOC relative trends. But in contrast to the other tropical stations, the Natal FTOC relative trend is much larger than the TrOC relative trend, so this is a special case. We will not mention it in the manuscript and removed the sentence from the text. In the next example, several European stations, have positive trends for all partial ozone columns, but the FTOC relative trend is the largest. The statement "but now with positive LTOC rates" was referring to the case before Natal (Irene, Fiji, Samoa, Ascension Island, Hilo, Atlanta, Wallops Island, Trinidad Head, Churchill, Sodankylä, Ny Ålesund, all sites where LTOC is negative) and was therefore somewhat misleading. Thank you for pointing this out.

We changed this sentence to "Another subset is made up of the European ozonesonde sites OHP, Hohenpeissenberg, and De Bilt, which have positive 2000-2022 trends for all partial ozone columns, with the largest relative increase for the FTOC, suggesting that at least some of the column increase is from mid-tropospheric transport."

Also in Line 588, I am not sure what the “least negative relative LTOC trends” are referring to. Is this saying that the FTOC is more sensitive than the LTOC due to mid-tropospheric/lower stratospheric dynamics?

This has been changed into “For the mentioned Arctic sites, whose TrOC, LTOC and FTOC trends are all negative, the larger relative LTOC than TrOC (and hence FTOC) trends (Fig. 12) indicate that the negative free-tropospheric ozone trends, due to mid-tropospheric or low-stratospheric dynamics, are partially compensated by the larger LTOC trends for obtaining the TrOC trends.”

Line 759 (and L520): the authors mention here in the conclusion that multiple observations per day can exist for several instruments (looking back I see this was well covered, I just missed the details as I was reading through). I see the Daily Mean (L2) used in the bias computations. For the QR analysis using all measurements, does this include multiple daily measurements, or are the L2 averages used? Also, when discussing the sample numbers within the month (around L520) does the sample count include multiple profiles in a day? This would make the monthly sampling problem even more of an issue, if say the 12 Umkehr samples per month actually occur on fewer than 12 days. If the sample numbers (SN) are notably different than the number of days typically sampled in the month, that would be useful to note.

It is true that we flip around with the temporal sampling of the time series in the manuscript. All measurements (L1) include multiple measurements a day for some techniques, and all those measurements have been used for the QR L1 trend estimation. They have also been used for the intercomparison analysis (or bias computations) in Sect. 4.1.1 between collocated or neighbouring sites, coincident within 12h (closest measurements). The daily mean values (L2) have not been used in the trend estimation and are only used for illustration purposes in the manuscript (e.g. the daily mean time series plots at collocated and nearby sites, and at sites in a specific region, Figs. 2 and 3). The monthly mean values (L3) are calculated from the daily values L2 and sample numbers within the month are, as a consequence, counted from the available daily values and expressed as number of days. We do not count multiple profiles in a day for these numbers. The monthly mean values (L3) are used in the QR and MLR trend estimations and in the intercomparison analysis between collocated or neighbouring sites. We also derive monthly anomalies from the L3 time series by subtracting the long-term monthly mean for each month. Monthly anomalies are shown in Figs. 14 and 15 and drifts between time series are determined as the linear regression fit slopes through the monthly anomaly time series differences at 2 sites.

As a response to the first reviewer, we also included an extra paragraph describing the impact on the trends of reducing the monthly sampling frequency to exactly 2 days a month. Impacts on both QR and MLR trends are described with clarification on how the sample numbers are exactly determined.

Figures:

The Figures are understandable complex, but I have a couple of suggestions to consider.

Figures 2 and 3, showing the time series and mean value, it is difficult to see the mean value dashed line in the figures, but since they are constant they could be extended into the white space (i.e. to 2025) then the reader will be able to see the individual dashed lines for easier comparison.

Thank you very much for this very valuable suggestion. It is implemented in the figures.

Figure 7 and similar: It is difficult to see the station names, although this may be the best that can be done. Did the authors try listing the station names along the plot axes, for example horizontally on the right hand side in panel a (would have to shift the legend), and vertically along the top axis in panel b? Or possibly where there is overlap, some stations are listed on the right/left or top/bottom so they all can be read.

Thank you for the suggestion. Figure7a (trends by latitude), the hardest to read, has been modified with their color-coded names along the axes. The rest have stayed the same as they were more legible.

Time Series Plots: For stations with extended time gaps I suggest removing the line over the gaps, for example in Figures 3 a+b and 15 a+b. Many of the longer lines connecting points with large gaps distract from the pertinent results of the plot, and the fact of the limited coverage is still clear because the color is missing in the gaps.

Very good suggestion. Done: measurements spanning gaps over more than 4 months are no longer connected by lines.

Minor Editorial Comments/Suggestions

L55: depending on site; (3)

Done.

L71-73: "In the first phase …" This sentence is incomplete, or maybe it is supposed to be a clause, in which case the period before Gaudel et al. (2018) should be a comma.

Done.

L77: remove comma after 2019

Done.

L85: comma after De Maziere et al., 2018)

Done.

L121: and lidar

Done.

L136-137: suggest "… in electrochemical cells (ECC). Known as the ECC sonde, this type is used in the HEGIFTOM analyses…"

Followed suggestion.

L152: suggest "… in a WMO/GAW Report (Report 201 by Smith et al., 2014)."

Followed suggestion.

L159: suggest "(i) Removing all known inhomogeneities … ; (ii) ensuring consistency …; and (iii) providing …"

Followed suggestion.

L230-231: (Bjorklund et al., 2023; Gordon et al., 2022) (correct punctuation after et al)

Done

L234: having -> have

Done.

L235: suggest continuously -> commonly or consistently

Replaced with commonly.

L255: relevant -> relative

Done.

L264-265: the minimum -> a minimum

Done.

L273: 30-m is that 30 meters? If so, 30 m and 2 km.

Done.

L283: introduce XO3 (I didn't see it before)

Done.

L349: within a region (delete -)

Done.

L415: 700>p>300 hPa

Done.

L448: Suggest removing "?" from section title

Followed suggestion.

L463: (Fig. 6b)

Done.

L501: suggest "with the sign of the Umkehr trend at some collocated sites differing from the other instrument(s).

Followed suggestion.

L521: suggest … only 3 airports with sufficient coverage to compute trends, the sample …

Followed suggestion.

L521-522: … most divergent: ATL and DAL have only … (note remove extra period after :)

Done.

L 524: … or Chang et al. (2004) …

Done.

L526: I'm not sure what "On the other hand" is referring to here. It seems the text before is describing the potential complications of the differing monthly sampling and that the different uncertainties by network type similarly make intercomparisons different. If this is correct, I would suggest "In addition, the different techniques … "

Followed suggestion.

L526: … with mean values of: …

Done.

L 527: comma after IAGOS rather than period

Done.

L621: The wording here was just a little confusing to me, maybe "There is a trend reduction for all but one Arctic site (Churchill ozonesondes) and for all but one North American site (IAGOS Dallas)."

Followed suggestion.

L634: suggest "Differences due to instrument technique"

Followed suggestion.

L641: that -> which

Done.

L654: details -> detail

Done.

L691-692: I think the sentence "This adds extra information …" could be removed here.

Done.

L694: suggest "the trend estimates are robust across statistical methods and the DLM results complement the previously reported results."

Followed suggestion.

L696: Figure 15. (remove :)

Extra space removed after : , but according to Copernicus guidelines, we cannot remove : in Figure caption.

Line 700: in function of -> as a function of

Done.

L706: A reference to the ozonesonde "total ozone drop off" might be useful here. If the Hilo tropospheric column is impacted as identified though intercomparison with other instruments at the same station location, do we expect other ozonesonde stations which experienced the drop off to be similarly impacted even though it is difficult to quantify due to lack of co-located data, or is Hilo thought to be a special case?

Thank you for comment. We already included references to Stauffer et al. (2020, 2022). The total ozone column drop-off is mostly found at tropical stations where *it affects only the stratospheric segment of the ozonesonde profile.* There is one clear exception. For the Costa Rica SHADOZ station anomalously low ozone post-2013 extends into the troposphere. Costa Rica data are therefore not included in our paper nor in equatorial SHADOZ trends papers (Thompson et al., 2021; Thompson et al., 2025). Hilo is a unique case. The magnitude of a "tropospheric" drop-off at Hilo is borderline so it is retained in the HEGIFTOM analyses and in Thompson et al. (2025; Supplement). Note that the origin of the drop-off remains not fully clear, although there might be a link with changing ozonesonde pump characteristics around the time of the drop-off occurrence (Nakano and Morofuji, 2023). It is also possible that Hilo ozone has recent low-ozone readings due to an interference in the ozone measurement from volcanic SO_2.

Thompson, A. M., Stauffer, R. M., Kollonige, D. E., Ziemke, J. R., Cazorla, M., Wolff, P., and Sauvage, B.: Tropical Ozone Trends (1998 to 2023): A Synthesis from SHADOZ, IAGOS and OMI/MLS Observations, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-3761, 2025

Nakano, T. and Morofuji, T.: Development of an automated pump-efficiency measuring system for ozonesondes utilizing an airbag-type flowmeter, Atmos. Meas. Tech., 16, 1583–1595, https://doi.org/10.5194/amt-16-1583-2023, 2023.

The text now reads: "After 2014, significant discrepancies are found with significantly positive Umkehr trends estimates and negative ozonesonde trend estimates. The latter may be related to a small total column ozone drop-off in the Hilo ozonesonde dataset (Stauffer et al., 2022) that is negligible for the other tropical HEGIFTOM data. Hilo is the only station in the analysis where some of the negative trend could also derive from an artifact tropospheric ozone loss caused by SO_2 interferences from Hawaiian volcanic activity in recent years."

Line 734: In Appendix A …

Done.

Line 750-751: suggest slight rewording so that the main point is that the data are available rather than the data include uncertainties, maybe "The HEGIFTOM data and associated uncertainties, covering more than 350 individual datasets, are available via http: … "

Followed suggestion.

L782: the word sparse confused me at first because it made me think of the sparse data issue (probably just me) but maybe consider using sporadic or intermittent, such as "to highlight intermittent periods over which the trend is significant, where trends estimated with the traditional QR and MLR methods to not show any significance."

Followed suggestion.

L805: "in the sense North America … " (remove "the")

Done.

L815: remove "will"

Done.

L846: change period to comma after "stations"

Done.