

Reply on CC1

The topic of merging ozone data records is of great interest and is indeed a great method to improve sampling and reduce trend uncertainties. The scope of this paper however is mostly to look at the biases and drifts between 2 measurements directly and less so about the individual trends. However, co-author Richard Querel is planning on performing an analysis exactly on this topic of merging the ground-based ozone measurements available at Lauder. This could indeed help resolve discrepancies found with in Pope et al. (2023).

Here, we chose to include a comment on the benefit of using the merged product in future trend studies in the conclusions:

“The good agreement of the three measurements in the troposphere (concerning no significant bias or drift) show that these are reliable to use for trend studies within HEGIFTOM. Future studies can take advantage of this by merging the FTIR, Umkehr and ozonesonde measurements in order to provide more accurate trends thanks to the higher sampling (Chang et al., 2024). However, because no strong correlation is found between Umkehr and FTIR in the troposphere and the Umkehr DOFS or consistently low here, one has to be careful in the inclusion of the Umkehr tropospheric column into the merged product.”

Information has been added according to the TOAR-II recommendations:

-Drifts are still shown in %/decade, but now a figure is added (Figure 6) showing those same values in the averaged mole fraction per decade.

-In the abstract it has been specified that it concerns 21st century trends

-A completely new statistical analysis is believed to be out of the scope for the paper in its current state. In section 4.2 (and briefly in the abstract), we have added a statement referring to the language used, especially the term ‘statistical significance’ such as it is used through the rest of the paper afterwards.

Additional comments have all been accounted for in the manuscript.

Reply on RC1

Thank you for all your comments and feedback on the manuscript. We have implemented all your comments in the paper, or when applicable explain here our response. The comments from both referees resulted in significant changes to improve the text accordingly. Most significant changes have occurred in: the introduction to provide additional motivation for the study; the discussion of the results, which have been expanded to provide more scientific explanations; the conclusions to clearly explain the impact of the results on TOAR-II and

LOTUS and the implications for the use of all included instruments; a new appendix per suggestion of an editor comment; and lastly the abstract has been updated to match all the changes. Here I give a short explanation to each of your comments:

- *Table 1 is not clear about the total ozone column measurements (TCO). TCO is provided by FTIR, Umkehr, Dobson and UV2. This does not appear in the table. The table should include a part dedicated to TCO measurements and another one to ozone profiles measurements. In the latter case, it should also include the altitude range of the ozone profiles measurements. Such an information is lacking in section 2 describing the various ground-based instruments.*

The table has been rearranged and a column is added to show which techniques provide total column measurements. Additionally, the information of the vertical extent has been mentioned now explicitly in the sections detailing each measurement technique.

- *References to the ozone time series obtained specifically at Lauder is lacking. For instance, the reference for the intercomparison campaigns of ozone profilers should be McDermid et al., 1998 <https://doi.org/10.1029/98JD02706> and a reference to the RIVM ozone lidar instrument should be included, e.g. Swart, Daan P. J., et al., RIVM's Stratospheric Ozone Lidar for NDSC Station Lauder: System Description and First Results, 17th International Laser Radar Conference, Sendai, Japan, 405-408, 1994.*

Both these intercomparison studies are now included in the introduction where they are mentioned as past intercomparisons of ozone profilers performed at Lauder. We have added the sentence: “Our study continues intercomparison studies performed at Lauder before 2000 such as by McDermid et al (1998) who look at several ozone profilers (lidar, microwave radiometer, and ozonesonde) and Swart et al. (1995) who focus on RIVM (Rijksinstituut voor Volksgezondheid en Milieu) lidar.”

- *Lidar ozone profiles are not highly resolved in the upper stratosphere, see Leblanc et al., <https://doi.org/10.5194/amt-9-4029-2016>, 2016. This should be mentioned in section 2.4*

A comment about the lidar resolution is added to section 2.4 together with the mentioned reference: “The lidar measurements are well resolved in altitude with the resolution standardized within NDACC according to Leblanc et al. (2016). This resolution ranges from a few hundred meter at 10 km to several kilometers in the upper stratosphere at 50 km.”

- *Figure 1 shows partial ozone columns and not total ozone columns for the instruments detecting ozone in specific altitude ranges (e.g. ozonesondes, lidar, microwave spectrometer). This is manifest in the range of partial columns shown and the seasonal variation.*

The caption and y-axis label have been changed to clarify that the time series show integrated ozone column. It is also clarified that this shows partial ozone column for ozonesonde, lidar and microwave radiometer and that it shows total column for FTIR, Umkehr, Dobson, and UV2.

- *Partial columns definition and DOFS: the article is quite explicit on DOFS for FTIR measurements but much less for the quantification of Umkehr and MWR DOFS in the altitude layers selected for partial columns evaluation. More information is needed on these DOFS computation.*

The information on the DOFS for both MWR and Umkehr have been added to their sections 2.2 and 2.3 respectively. Additionally, in section 3.1, both instruments now have the DOFS mentioned as calculated for each partial column.

- *Equation 4: How is handled the discontinuity in the ozone profile (upper range of the profile for the sondes; lower and upper range for the lidar) when the smoothing is applied to these measurements?*

We disregard profiles from MWR, lidar, or ozonesonde if they stop in the middle of a partial column such that no discontinuity needs to be accounted for in the intercomparison. This assures us that all measurements from these three instruments fully cover the altitude extend over which the intercomparison occurs. This clarification has also been added to the text in section 3.1.

- *Time coincidence: More information is needed on the number of measurements made per day made by FTIR and MWR measurements. As for FTIR, various MWR measurements call fall within the time window of other observation. In that case the MWR measurements averaged in the same way? Appendix B does not really answer this question.*

Often both FTIR and MWR have multiple measurements made per day and can fall within

the same time window. The coincidences are constructed by looking at each MWR measurement and finding all the FTIR measurements within the defined window to average out and compare to. If both measurement techniques have multiple measurements in the same day that fall within the time window (for example, one or more FTIR measurements are taken within three hours from two different MWR measurements), then the FTIR measurements that fall in this overlap are used for comparison in both observation pairs. This explanation is added to the section 3.3 for clarity.

- *Bias and dispersion analysis: no reference is given for the scaled MAD computation and more specifically for the 1.4826 factor. More traditionally, standard deviation and standard error (e.g. the standard deviation of the mean) are used to evaluate the significance of the bias between 2 times series. The standard error is then compared to the combined uncertainty averaged over the whole record. How does the methodology used here compares to such traditional method?*

This scaling factor makes the MAD representative as a deviation from the median, similarly as the standard deviation is to the average, in the case of a normal distribution (see Rousseeuw & Croux, 1993). We are not dealing with perfect Gaussian distributions, but the factor still creates a reasonable value for the scatter. The scaled MAD is thus similar to using the standard deviation but is more robust in the sense that it will be less affected by outliers. The mentioned reference has been added to the text in section 4.1.

- *The correlation between 2 ozone time series is heavily driven by the seasonal variation. What would be the correlation for the deseasonalized time series?*

To analyze the correlation without influence of the seasonality, we have added to the table the correlation between monthly anomalies of the time series and discuss these also in the results for each instrument comparison for each partial column.

- *Section 4.1.1: it is not clear if the Dobson TCO corresponds to that computed from the Umkehr retrieval or to the Dobson TCO itself.*

The text had a mistake here and it is clarified which results belong to the Umkehr TCO and which to the Dobson TCO.

- *Section 4.1.5: There is no mention of issues linked to ozone diurnal variation for the*

comparison in the upper stratosphere, see Sauvageat et al. <https://doi.org/10.5194/acp-23-7321-2023>, 2023. Impact of ozone diurnal variation on comparison results should thus be discussed.

The largest diurnal variation is found near the stratopause according to Sauvageat et al. (2023). Our definition of upper stratosphere only reaches up to 42 km. We still estimate the diurnal variation, though, by calculating the short-term variability of the MWR measurements in our upper stratospheric column. We find a maximum value of 18% for the variation in a day, and the variation within the 3-hour window that we consider is very small with a mean value of 1.1%. This is now added to the text in Appendix B.

- Equation 13: how the N parameter corresponding to the number of degrees of freedom is computed?

The degrees of freedom N, used to calculate the drift error is found from the length of the time series in monthly means that is used to calculate the drift. A small clarification has been added to the text in section 4.2.

- *Section 4.3: Trend results are interesting and explanations on trend differences due to instrumental drift are convincing. However, if trends are computed for coincident measurements only, the authors should indicate the reduced number of observations used to derive the trends for each data record.*

Because indeed the number of observations used in the intercomparison study has an important effect on the results, we have now added the number of observation pairs that are used for each comparison in table 3.

Reply on RC2

Thank you for all your comments and feedback on the manuscript. We have implemented all your comments in the paper, or when applicable explain here our response. The comments from both referees resulted in significant changes to improve the text accordingly. Most significant changes have occurred in: the introduction to provide additional motivation for the study; the discussion of the results, which have been expanded to provide more scientific explanations; the conclusions to clearly explain the impact of the results on TOAR-II and LOTUS and the implications for the use of all included instruments; a new appendix per

suggestion of an editor comment; and lastly the abstract has been updated to match all the changes. Here I give a short explanation to each of your comments:

Summary of Paper

There are five ground-based (GB) instruments at Lauder, New Zealand, that have measured total column ozone and/or partial columns throughout the interval 2000 to 2022. They do not appear to all give the same values in total ozone or in various segments: troposphere, lower stratosphere, middle stratosphere, upper stratosphere. Accordingly, computed trends over the 23-year period differ, especially in the lower stratosphere (LS) where LOTUS has concentrated its efforts. A major goal of this paper, as expressed in the Abstract, is to determine why LOTUS trends are not similar among the techniques. The second goal is to determine “quality and relevance” for TOAR II trends, two criteria that are not well-defined.

This paper makes comparisons of the ozone amounts systematically within the segments, using FTIR as the primary reference. Of the four independent measurement types considered, three (Lidar, Microwave, Umkehr) all display significant drift relative to FTIR in one or more stratospheric segments; ozonesondes do not (Table C1). Certain discontinuities near the end of the record contribute to these drifts and the divergence of trends (Appendix D). A reprocessing (modified FTIR retrieval) improves some of the drifts. In summary, the paper contains worthy analyses, carefully carried out.

However, there are two reasons why the paper is not ready for publication. First, after all the tables and analyses, the paper does not come back to clear answers to guide how past LOTUS results concerning Lauder can be updated. Nor does the paper provide recommendations for TOAR II activities on how to use the findings in trends analyses. For example, should one try to merge the various datasets for tropospheric ozone analysis? Why or why not? If so, how would that be done?

In this paper we do not yet focus on the merging of the data sets. The main outcome is to see if there are spatial/temporal/instrumental mismatches between the instruments that can cause a different representation of the ozone field properties. The topic of merging measurements is very interesting though, so we have included a statement about how, thanks to an increased temporal sampling, the trend estimates from merged data sets are improved compared to the individual datasets. This is also work carried out by (a) co-author(s) for future publication.

Additionally, more explicit conclusions have been added on how the results help the research of both LOTUS and TOAR-II.

The paper needs to be re-outlined and clear conclusions on how, if and why each of the 5 datasets can be used in ongoing LOTUS and TOAR II analyses.

We have made significant changes to the abstract/introduction to provide a clear context of our goals of the paper concerning all the datasets in context of LOTUS and TOAR-II and we provide a systematic discussion on the results and the repercussions of those results on the usability of separate measurement techniques in the partial columns. This also means that we added more explicit discussions of the issues that cause the biases/drifts in some of the measurement techniques in section 4.2.6.

Second, there are more fundamental questions about the Lauder datasets relevant to LOTUS and TOAR II. Here are several:

1. In the TOAR II HEGIFTOM activity, presumably the FTIR, Umkehr, sonde records have been homogenized. The paper gives no information about the data version, archive, etc, for each of these data sets. Are these the HEGIFTOM files at the RMI ftp repository? The customary doi information is lacking

Information has been provided in the ‘Data availability’ section for each of the instruments. A doi is provided for the collection of all data sets used in this article (DOI is currently pending but will soon be available.)

2. With respect to the ozonesonde data in particular, papers by Stauffer et al (2020; 2022) and updates (through 2021, see Figures below) find total ozone column and stratospheric ozone in particular, suffered the “Ensci dropoff” artifact at Lauder. The upper figure is a satellite comparison – Aura MLS for stratosphere, OMI, OMPS and European TCO comparisons. The lower Figure is based on the Lauder Dobson as archived at WOUDC, Dobson presumably the source of the Umkehr data. Have ozonesonde dropoffs been corrected in the HEGIFTOM files? The wording about the version of sonde data (page 9 of the manuscript) is vague. Reprocessing via the Smit method, even the WMO/GAW, 2021, Report, as referenced (line 217 to 220) do not give a procedure for correcting for the dropoff. Was the process of Nakano & Fujimora, AMT, 2023) to correct the dropoff applied to the Lauder record? “Claim to be removed” is your wording – what does that mean? If the dropoff has been fixed, it would be good to have a supplementary figure showing that.

In the papers by Stauffer et al. (2020; 2022), Lauder has never been identified as a drop-off site (see Fig. 2 in Stauffer et al., 2020 & Fig. 1 in Stauffer et al., 2022). In Table 2 of Stauffer et al., 2022, the average Lauder EnSci ozonesonde TCO change relative to OMI pre-EnSci and post-EnSci S/N 25,250 is -2.6%, which is below the -3% threshold (defined in Stauffer et al., 2020) for an “Ensci dropoff” site. The overall average of this metric over the entire EnSci network is -1.8%.

Thank you for providing those figures for the Lauder time series. An update of this figure, together with the comparisons with the unhomogenized Lauder ozonesonde time series was presented at the WMO Technical Conference on Meteorological and Environmental

Instruments and Methods of Observation (TECO-2022) in Paris on 10-13 October 2022 and can be downloaded here:

<https://ozone.meteo.be/uploads/media/634ea9e9d6f09/vanmalderen-p102-wmoteco2022.pdf?v20231106-1421>. From this figure, one could also argue that there is an overall decline in the total ozone content of the ozonesondes w.r.t. the co-located or satellite overpass total ozone measurements, instead of a sudden dropoff. Moreover, Stauffer et al. (2020) also mentioned that “Some sites (e.g., Lauder in 2015) switched radiosondes again from RS-92 to the RS-41”, which might have an impact on the ozone profile calculation (through the pump temperature and pressure measurements) as well. Based on these arguments, we do not refer to Lauder as an Ensci dropoff site in the paper.

Surely, Nakano & Fujimora (2023), reported differences in the pump motor specifications of the ozonesondes delivered to the JMA before 2013 (serial numbers ≤ 24000) and after 2013 (serial numbers > 24000), which might have an impact on the total ozone of around 1% (their Fig. 17). Those “measured” JMA pump efficiency correction tables, for each serial number series, might be used instead of the Komhyr et al. (1995) empirical pump correction factors, but there are three problems with this:

- according to the WMO-GAW #268 guidelines, the official reference document for ozonesonde data processing, the Komhyr et al. (1995) tables should be used for En-Sci ozonesondes
- just using the Nakano & Fujimora (2023) pump efficiency measurements instead of the Komhyr et al. (1995) tables will, for SST0.5 and SST1.0 solutions, completely alter the ozone distribution in the upper parts of the profile (starting for pressures lower than 100 hPa) and, as a consequence, the total ozone content of the profile, because the Komhyr et al. (1995) “empirical correction” tables combine decreasing pump efficiency, increasing conversion efficiency, and typical memory effects in the background current for the standard buffered solutions SST1.0 and SST0.5 (Tarasick et al., 2021).
- To solve the incompatibility of the WMO-GAW #268 guidelines with the use of the Nakano & Fujimora (2023) pump efficiency correction tables, the new methodology as described in Smit et al. (2024) and in Vömel et al. (2020) is indeed a possibility. But these procedures require some accurate and additional pre-launch information (IB0, IB1, sensor (fast) response time, time between IB1 measurement and launch) and might introduce noise in the data. Practical guidelines to implement those methods under less controlled and well-established conditions as in the JOSIE campaigns, on which data the methods have been developed, are still lacking. As such, these methods are still in research mode and should be implemented and assessed at a global scale before being used in an intercomparison paper like this. Finally, the calibration functions introduced in Smit et al. (2024) to refer the ozonesonde measurements to the photometer in the Jülich simulation chamber, have been determined using the new absorption cross sections for the photometer. To have a consistent comparison between those corrected ozonesonde time record and other co-located techniques, this new set of absorption cross sections should also be used for the other techniques.

To summarize: we will not change the processing of the Lauder ozonesonde time series. It has been processed through its entire time series according to the WMO-GAW #268 guidelines, the official document. The ozonesonde time series at the HEGIFTOM ftp-server have been processed according to those guidelines for all sites, creating a consistency among

those sites at the HEGIFTOM ftp-server. The WMO-GAW #268 does not correct for a possible EnSci dropoff, which has never been identified as such in the Lauder time series. At the moment, the EnSci dropoff has not been corrected for in any site, even in the so-called dropoff sites. New proposed methodologies (Smit et al., 2024 & Vömel et al. 2020) are still in experimental phase and should not be widely used before assessed globally (and after practical guidelines).

We included in the manuscript the sentence: “In a worldwide ozonesonde comparison with satellite and ground-based total column ozone and with satellite stratospheric O₃ profiles, Stauffer et al. (2022) mentioned a negative ozone bias in the homogenized Lauder ozonesonde time records in more recent years (see also the plot in https://acd-ext.gsfc.nasa.gov/anonftp/acd/shadoz/nletter/stations_vs_satellites_timeseries.zip). This feature might be related to the so-called “post-2013 dropoff in total ozone” identified in a number of ozonesonde stations (not Lauder) in Stauffer et al. (2020), but, as no clear cause has been determined yet, no correction strategy has been implemented in the here applied WMO/GAW 2021 homogenization procedures.”

However, in this response, we want to show the impact of the new ozonesonde processing strategy according to Smit et al. (2024) on the comparison with the FTIR. We have added this in the discussion 4.2.6. where we mention the following about our tests with the new processing:

“While it is decided in the present study to use the sondes data sets from HEGIFTOM that follow the WMO/GAW 2021 homogenization procedures (See Section 2.5), we have performed as a test the drift study on a sonde data set in which a “dropoff” correction as suggested in Stauffer et al (2022) has been applied. The bias and dispersion with FTIR are worsening with this newly processed sonde data set. In the middle stratosphere, where the effect of the dropoff is usually most significant, we see a bias of -9.3% and a scatter of 4.3%. However, it should be mentioned that the drift with FTIR is significantly positive (1.3 ± 1.1 %/decade). This effect is very small in the case of Lauder (1.2 %/decade), but does seem to go in the good direction towards the other ozone stratospheric trend measurements at Lauder where we see a similar positive drift of the measurements with respect to FTIR. To confirm the changing trend when applying the dropoff correction, we perform a similar drift analysis of the ozonesonde data sets to the lidar measurements. In the middle stratosphere we see that the drift (when using (lidar-sonde)/sonde) changes from 2.0 ± 1.3 %/decade for the original ozonesonde data set to 0.7 ± 1.6 %/decade for the newly processed data set. This seems to be consistent with the earlier results comparing sonde to FTIR, because the significant drift between lidar and sonde that is present with the original sonde data is not there for the newly processed sonde data, putting the trends of the ozonesondes more in line with that of lidar.

However, since this newly processed data set is only a temporary test to analyze differences with the original data set, we will remark here that while the EnSci dropoff seems to be better resolved, future attention is needed concerning the ozonesonde trends when a new official ozonesonde data set is available.”

3. If the dropoff has not been corrected, the authors need to implement the Nakano and Fujimora (2023) procedures; ideally the new reprocessing by Smit et al (AMT, 2023) would lead to an even more accurate, referenced result. For LOTUS applications the FTIR-

referenced comparisons make sense but for the TOAR II application in the troposphere, the optimized sonde data should also be used as the reference.

As explained in the response to your previous comment, we stick to the official WMO-GAW #268 processing of the ozonesonde time series. However we do mention in the paper the effect of the new processing on the bias and drift with respect to FTIR and lidar.

4. *In the case of TOAR II/HEGIFTOM, calculations for 2000-2022 trends being prepared for publication (VanMalderen et al) show the following. Note that trends for the HEGIFTOM ozonesonde data at Lauder (surface to 300hPa) and trends for Umkehr and FTIR at Lauder diverge somewhat as shown below. Graphs of this information were presented to the HEGIFTOM Teams meeting of 7 March. (Based on calculations from NOAA and GSFC)*

2000-2022 Trends	Surface to 300 hPa	Not rounded to sig fig	QR L1	
(ppbv/dec)	QRL3 (ppbv/dec)	MLR L3 (ppbv/dec)		
Lauder O3S -45	169.68	0.134324342	0.01106383	0.133214349
FTIR -45.04	169.68	1.544135587	1.638209739	1.673699546
Umkehr -45.04	169.68	0.358046	0.377753	0.579331805

It is assumed that the data used in the above Table are the same as Björklund et al are using but more details are required in Section 2.

The FTIR data used in our study uses a new retrieval method which is found to affect the resulting trend (as can be seen in Figure 7). The new trend of FTIR is lower than with the 'old' FTIR, which then helps to resolve the diverging trends in the troposphere at Lauder.

RELATED COMMENT IN RESPONSE TO OWEN COOPER COMMENT ON THIS PAPER. (see <https://doi.org/10.5194/egusphere-2023-2668-CC1>). *The table above shows that there is sufficient variation in the surface to 300 hPa trends for sonde, Umkehr and FTIR that "averaging the data" (as Cooper recommends) or averaging the trends is not justified. The current manuscript and the trends analyses show that, in a revised manuscript, more analyses need to be carried out, with careful uncertainty comparisons, on the FTIR, Umkehr and sondes before merging of data can be considered, as suggested by Cooper. It is particularly important that uncertainties for the 5 different instruments being considered are compared. Note that Figure 1 in the manuscript suggests that FTIR and sonde TCO had some declines, albeit not monotonic or identical, after 2014.*

A further comment on the Cooper et al Comment on this paper. Reference is made to the Pope et al RAL paper: Atmos. Chem. Phys., 23, 14933–14947, 2023

<https://doi.org/10.5194/acp-23-14933-2023>. That paper was accepted prior to the reprocessing of OMI (2014-2021) data that displayed a drift artifact in total ozone. The latter issue is discussed in with corrected data by co-author Ziemke in Gaudel et al: <https://egusphere.copernicus.org/preprints/2024/egusphere-2023-3095/>. The Pope et al., RAL product overestimates tropospheric ozone trends.

As mentioned in a reply above, we decided not yet to focus on merging the data sets in this study. The comments of Cooper et al are nonetheless interesting to explore in order to provide a potential way to improve the uncertainties on trend calculations. This is why we chose to include a short discussion on the topic of merging in the conclusions, however also remarking the care that is needed to merge datasets, especially if there is evidence of large dispersion or low correlation between the involved data sets.

In summary, the paper in its present form should not be published. In a revision the authors need to:

- *clarify the source of their data – the customary DOIs and references on the datasets are absent.*

The explanation on the sources in the ‘Data availability’ section have been expanded and a DOI for the collective dataset will be available for reference.

- *If the sonde data are not corrected for an artifact stratospheric ozone loss after 2014, that needs to be done before re-analyzing drifts. Intrinsicly, the sonde data are more accurate than FTIR in the troposphere and possibly in the lowest and mid-stratosphere. Drifts in FTIR for those segments relative to corrected sonde data should be carried out and discussed for the troposphere, lower and mid-stratosphere.*

Aside from our explanation on the issue of the dropoff in the reply above, we have included an analysis and discussion when comparing the corrected ozonesonde data to both FTIR and lidar in the drift discussion in the results.

- *Most important, please think through and describe clearly the significance of the new*

results for LOTUS and TOAR II/HEGIFTOM. The paper currently presents interesting technical details but does not relate a clear scientific story of interest to the TOAR II community.

Per your suggestion, we have significantly changed the introduction, discussion of results and conclusions to present clearer motivations related to TOAR-II and LOTUS and to provide more concrete and clearer scientific explanations to our results combined with more explicit suggestion of the use of the ground-based measurements in future studies.

Lesser comments:

- *Section 2.5. Note that the sonde instrument type and solution used at Lauder should be added.*

We changed this to “At Lauder, ozonesondes from the two different ECC ozonesonde manufacturers (SPC and EnSci, switch made in May 1994) have been launched, and different sensing solution types (SST1.0 and SST0.5, changed in August 1996) been used as well.” We also added a reference to the manuscript “Analysis of a newly homogenised ozonesonde dataset from Lauder, New Zealand” by Zeng et al., 2023 in this section: “More details of the Lauder ozonesonde time series and the homogenization procedure can be found in Fig.1, Table 1, and Appendix A of Zeng et al., (2023).”

On line 214, end of sentence, the following reference for the variations in types of instrument and solutions should be inserted.

1. G. J. Smit, A. M. Thompson and ASOPOS, Ozonesonde Measurement Principles and Best Operational Practices, ASOPOS (Assessment of Standard Operating Procedures for Ozonesondes) 2.0, 165 pp., WMO/GAW/IO3C/NDACC/GRUAN, WMO/GAW Report 268, Geneva. (Online at https://library.wmo.int/index.php?lvl=notice_display&id=21986#.YaFNSbpOlc8).

Alternatively this can be called WMO/GAW 2021 but the citation is missing from the Reference list at the end of the manuscript

Thank you for this suggestion. We implemented it.

Reply on RC2

Thank you for your comments. Calibration histories are indeed highly relevant in the discussion of an intercomparison of ground-based measurement techniques. Therefore, we have added information about the calibration of the instruments with a focus on FTIR and Dobson instruments, which is added in a fully new appendix (Appendix D).

Concerning the technical comments:

Page 1, Line 5: Abstract, between => among. Please check usage of "between" here in the abstract and in the manuscript body.

This change has been made in the manuscript.

Page 5, Line 109: Technically speaking, this does not show the sensitivity of the instrument itself, but rather of the chosen retrieval strategy.

This has been changed to refer to the retrieval and not the measurement itself

Page 23, Line 559: "... thanks new regularization ..." => "... thanks to new regularization ..."

This change has been made in the manuscript.

Page 24, Line 578: "... microwindows are chosen ..." => "... microwindows, which are are chosen ..."

This change has been made in the manuscript.

Page 25, Line 597: where is this -5.7% shown or derived from?

The value is derived exactly as in the main body of the paper for the updated FTIR data set. The value here is only mentioned in the text and not in any table.

Page 26, Line 625- 626: “and we even find that there is complete agreement of FTIR with ozonesonde over all partial columns.” This sentence is rather vague.

This has been changed to “... find that there is no significant drift between FTIR and ozonesonde for any of the partial columns”

I think the authors should compare and contrast the results with the results of this publication as well:

Steinbrecht et al., An update on ozone profile trends for the period 2000 to 2016, Atmos. Chem. Phys., 17, 10675–10690, 2017 <https://doi.org/10.5194/acp-17-10675-2017>.

A discussion of their results has been added to the introduction in relation to the similar results found in Godin-Beekmann et al. 2022.