Review of Björklund, Vigouroux, Effertz et al: Intercomparison of long-term ground-based measurements of tropospheric and stratospheric ozone at Lauder, New Zealand (45S) 12 March 2024

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? 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.

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
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.

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.
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. *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 montonic 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.

In summary, the paper in its present form should not be published.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

(1) clarify the source of their data – the customary DOIs and references on the datasets are absent.

(2) If the sonde data are not corrected for an artifact stratospheric ozone loss after 2014, that needs to be done before re-analyzing drifts. Intrinsically, 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.

(3) 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.

Lesser comments:

(1) Section 2.5. Note that the sonde instrument type and solution used at Lauder should be added. On line 214, end of sentence, the following reference for the variations in types of instrument and solutions should be inserted.
H. 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#.YaFNSbpOIc8). Alternatively this can be called WMO/GAW 2021 but the citation is missing from the Reference list at the end of the manuscript

(2) The authors have done a fine job in English but there remain many English errors. Please ask authors 3, 5 or 6, as appropriate to review and correct them.

(3) The Stauffer references for figures below:
Stauffer, R. M., A. M. Thompson, D. E. Kollonige, J. C. Witte, D. W. Tarasick, J. M. Davies, H. Vömel, G. A. Morris, R. Van Malderen, B. J. Johnson, R. R. Querel, H. B. Selkirk, R. Stübi, H. G. J. Smit, A post-2013 drop-off in total ozone at third of global ozonesonde stations: ECC Instrument artifacts?, *Geophys. Res. Lett.,* doi: 10.1029/2019/GL086791, 2020.
Stauffer, R. M., A. M. Thompson, D. E. Kollonige, D. W. Tarasick, R. Van Malderen, H. G. J. Smit, H. Vömel, G. A. Morris, B. J. Johnson, P. D. Cullis, R. Stübi, J. Davies, M. M. Yan, An examination of the recent stability of ozonesonde global network data, Earth Space. Sci., https://doi.org/10.1029/2022EA002459, 2022.

Figure showing ozonesonde ‘dropoff’ for TCO and stratospheric ozone in the Lauder record (Stauffer et al., 2020; Stauffer et al, 2022 & updates). Files were downloaded from RMI ftp site, 2021. The lower comparison is sonde TCO vs TCO from the co-located Dobson.