**General Comments**

The primary subject of the submitted manuscript is the long-term trend in vertically resolved stratospheric ozone based on the measurements of Dobson and Brewer spectrophotometers operated at Arosa and Davos using the Umkehr technique.

Overall, this is an important topic and a very valuable dataset. Further, the work is essentially sound and certainly well within the scope of ACP.

I recommend publication, but only after a significant revision to improve the clarity.

Unfortunately, in its present form, I have to say I found the manuscript very difficult to follow. In numerous places the discussion went straight into details without sufficient introduction, terms were used without explanation (including acronyms) or the reader was assumed to be aware of facts that hadn't been presented yet.

Even the overall scope of the work is unclear.
As a prime example, the place of the alternative NOAA homogenization remains mysterious to me even after several readings. In lines 260-262 the text states "The NOAA homogenization has been developed ... Our approach is different ...". This can only mean the NOAA work is not part of the current work, but is being presented as an alternative for comparison. The MCH homogenization is described in much greater detail. However, I note two co-authors are included for performing the NOAA homogenization, according to the stated author contributions. Even more confusingly, section 2.3 gives details about measurements at Boulder and OHP which seem of almost no relevance. (Incidentally "OHP" is never explained or located on the globe).

It seems to me the core of the work is the post-2000 trend using DLM based on Dobson 051 using the MeteoSwiss homogenization, but the manuscript also at times covers the pre-2000 trend as well as post-2000, the MLR trend as well as DLM, the Brewer and AURA-MLS trends as well as the Dobson, and the NOAA homogenization as well as the MCH. If such a broad scope is going to be covered then the authors need to provide a lot of assistance to the reader to make sure the right message is being communicated.

As an example, from my understanding of your description, the NOAA homogenization is designed to minimise the offset with AURA-MLS (line 258). The discussion of Figure 5 implies AURA-MLS is being used as a reference to compare the quality of the four datasets shown, which is inconsistent with the first approach. Then Figure 8 compares the trend determined from Dobson 051, Brewer 040 and AURA-MLS, as if these are independent datasets. These three discussions will seem to be mutually contradictory to the reader without better explanation.

To be frank, I would recommend a complete rewriting of the text. The authors should first decide the logical pathway they would like the reader to follow and then provide clear signposts and guidance to enable this.

I note there is no discussion of the physical basis of the deduced trends at different heights or the implications for ozone recovery, but I think that is quite reasonable given the scope is already very large.

Only the statistical uncertainty of the trends are provided. Clearly the overall uncertainties are much greater than these given the different results produced by the different instruments, the different homogenizations and the different trend analyses.

Specific comments
The title is too general, as it implies all Umkehr datasets from Dobsons and Brewers around the world are going to be considered rather than those from just one location. The acronym "DLM" should be expanded.

Line 6 – You talk about there being six instruments available but then go straight to D051, without explaining why this one of the six is the focus.

Lines 7-11 This sounds like the main goal of the work was to homogenize Dobson 051 using the co-located Brewers as a reference. However in other places (eg the title) the implication is you are calculating trends from all available instruments. I would prefer the abstract provide much clearer guidance to the reader about what the main idea of your work is.

Line 8, 9 – I am not sure "technical intervention" is really the best term, I think something like "instrumental changes" would be more accurate.

Line 11 Throughout the manuscript, it is not clear to me what exactly you mean by "OEM" – do you mean an optimal estimation method (ie Rodgers) or the particular optimal estimation method as implemented in the standard Brewer algorithm (ie Petropavlovskikh)?

Lines 12-14 You should explain better to the reader how the alternative NOOA homogenization fits in – is it part of the scope of this work or is it just going to be used for comparison?

Line 15 MLR trends were also calculated.

Line 20 As a comment, the fact the trends depend on the dataset raises the interesting question of their true significance.

Lines 22-63 You should also include the findings of Godin-Beekmann et al. 2022 here. This is a very important reference for this discussion.

Lines 29-63 The tenses are inconsistent, you have Chipperfield 'pointed to', but Wargan 'confirms', and also 'have been reported' (Bognar)
Line 40 You say MLR has been used for "ozone trend estimation" but the discussion that follows seems to be only about ozone profiles and not total ozone. You should either give references for the use of MLR in total ozone trend estimation (where of course it has also been widely used) or say "vertically resolved ozone trend estimation" or words to that effect.

Line 65 Ozonesondes, not radiosondes (this occurs later too at line 166)

64-66 Not just that, I would have thought there would be many Umkehr records at stations for which there weren't ozonesondes at the time, or in fact, where ozonesondes have never been flown.

Lines 76-78 Do you mean that Umkehr records only exist in the Northern Hemisphere mid-latitudes? (And not the Arctic, Antarctic, tropics or southern-hemisphere?) What about the Umkehr results shown in Godin-Beekmann et al. 2022?

Lines 80-81 I don't think that can be right – I have listed at least one example in the references (Fitzka et al.) Perhaps I have misunderstood.

Lines 81-83 This sentence seems to contradict the previous one – re-word to make the distinction clear.

Line 91 This makes it sound as if the NOAA homogenization is not part of this work but is being used as a comparison – in other places quite a different impression is given

Line 96 I would suggest starting with a very brief (one sentence) introduction of what an Umkehr observation is before getting into all the details.

Line 98 "relocation" rather than "relocalisation"

Lines 100-105, 109-11 Would it be possible to re-write this section to make it easier for the reader to follow? Perhaps a table would help?

Line 102 Is this the last we hear of D015?
Considering the different stray light characteristics of single and double monochromator Brewers, how well do their Umkehr results agree with each other?

I think your description of the Umkehr method could be made clearer. Stone et al. 2015 did a good job. If the shorter wavelength is scattered above the ozone layer, wouldn't its intensity decrease due to passing through more ozone?

You could reference Götz as the originator

The reader can't assess from this description whether the empirical correction is a good idea or not. Do you think it has any significant effect on the trends? Presumably it hasn't been used the whole time since 1956?

"is commonly retrieved" – from this wording it is not clear if you are talking about what other people do, or what is being done here.

Does it matter that the a priori profiles are now 20 years out of date, and ozone has increased since then in the upper levels, as your work shows?

Information below level 4 "is not independent" – independent of what? Each other?

You say a generic stray light correction "can be applied" – but have you applied it? Is it different for Dobsons and Brewers? Does it affect the results?

I am very confused about why this section is here – it seems completely irrelevant.

The way this is written it seems to contradict itself – you say "in this study we use ozone profiles ... given on 55 pressure levels " but then you "only consider" 10-75 km - please reword for better clarity.

I assume this is the latitude and longitude of Arosa (and not just "Switzerland") but it doesn't seem to be actually given anywhere in the manuscript.
Line 191 It don't think it's reasonable to cite Ziemke (2017) for this calculation – it appears in Godson (1962) in his discussion of Umkehrs.

Line 194 – The implication is Petropavlovskikh et al. 2022 defined the definitions of the Umkehr layers.

Line 196 I don't think 'the 2008 homogenization' has been mentioned before – what are you talking about?

Line 201 If you use the wording "a technical issue in the metadata", it sounds like there is a problem with the metadata. I think it would be better to say something like "an instrumental change recorded in the metadata".

Line 205 "retrieval iterations higher than 3" has not been explained.

Lines 205-208 Given there are small but identifiable differences between the Dobson and the Brewer, does this limit the effectiveness of your approach?

Line 210 It sounds like good metadata is only available from 2000 onwards – does this affect the confidence of the trends for earlier periods (eg Figure 4)?

Lines 214-215 This effect also seems evident in 2009?

Lines 216-217 It only looks lower in the lower altitude levels?

Line 218 But not at the level around 30 km?

Lines 219-221 It seems a shame that the observing routines changed in this way – was there a reason for the change?

Lines 222 I found the table somewhat hard to follow and perhaps therefore slightly unconvincing. I suggest at least considering whether it would be better to relate the changes in the instrument to the observed discrepancies using text. At present it is hard for the reader to assess what you've done and its validity.
Lines 228-229 Given the Brewer uses different wavelengths, with a different assumed stratospheric temperature, and has different stray light characteristics at high SZA, is it reasonable to apply a constant offset to the N values?

Line 244 You need to include some introduction to this section rather than going straight into the details, explaining in broad terms how this part fits into the overall picture of what you're doing, and secondly, how this homogenization is different to yours in its basic approach.

Line 244 The implication is you do not use the 'correction for the stray light effect' – does that matter?

Lines 249-250 I don't follow this. What is the difference between using a reanalysis and using specified dynamics? What are the specified dynamics based on if not the reanalysis?

Line 251 "accounting" -> "accounting for"

Figure 3 – I think this diagram is very helpful – I like it.

Lines 257-258 Why would you try to reduce mean bias compared to AURA-MLS? The implication is MLS can be used as a reference to adjust the Umkehrs to. In that case why bother with the Umkehrs at all?

Line 258 Why stop at 2018? This makes it seem as if the NOAA homogenization is a separate piece of work.

Line 261 "Our approach is different" – this strongly makes it seem as if the NOAA homogenization is a separate piece of work.

Line 265 The term "MCH" has only appeared in a caption until now and should be properly explained

Lines 268-269 You don't say why this correction would cause high variability
Line 283 I am not convinced the use of the abbreviations 'MS' and 'UpS' is beneficial, all things considered, but you might disagree.

Lines 282-297 It seems to me that you are using AURA-MLS as a reference, in that you are judging the quality of the homogenizations by how well the data agrees with MLS. Is that what you really mean?

Figure 5a – This plot is helpful but the black line is very hard to see. All four datasets seem to agree better prior to 2010.

Figure 5b Why is the black line only visible for a short period? Does it lie underneath the red line?

Figure 5c This is a very helpful plot.

Figure 5 Caption - You should state the heights or pressures corresponding to the layers DL5 and DL8 for the convenience of the reader.

Line 296 Why wouldn't the model take into account the 'local variability'? Is the resolution too coarse?

Lines 298-302 I found this hard to follow, sorry.


Lines 308-311 and 323-324 Did you find that this selection of proxy variables gave good results? Did you need all of them?

Line 309 The link is not for the QBO

Line 310 When I follow this link, the data only seems to go back to 2004, not 1970 or 1956.
Figure 6 Caption – You should explain in the caption what the blue and black lines are.

Lines 331-345 Would this be easier to represent in a table?

Lines 348 Here, I think you should use the word "percentage" rather than just the symbol.

Lines 360-372 Wouldn't it be more meaningful then just to calculate trends on the independent Umkehr levels?

Lines 386-390 This information is presented without any context. Is the point that Boulder and OHP are also northern hemisphere mid-latitude sites?

Lines 393-394 – "... has been homogenized on the observation data level" – what does that mean?

References


