General Comments:

The paper by Egli et al. presented research on deriving traceable total column ozone with the QASUME instrument. Traceability, or more specifically, a detailed uncertainty budget, is in great need for ground-based ozone observations. The methodology and analysis applied are comprehensive and solid. The instrument shows a good potential to be an independent/third-party reference to Brewer and Dobson ozone observations. However, some of the results still need more investigation or clarification. For example, the -1% offset between QASUME and Brewer or Dobson should be investigated (or, at least better described with more details), as the author claimed a 0.8% standard uncertainty. Overall, the paper is a well-written one. I recommend publishing it on AMT after addressing the following comments.

Specific Comments:

L63-64: please consider changing "the four wavelengths of … " to "the standard four wavelengths of … ". Please note that Kerr (Kerr, 2002) developed a scanning TCO retrieval method for the Brewer, and published the work back in 2002.

L159: If the term τ is a function of T and p, please make it consistent. In this line, the term was referenced as $\tau(T)$ and $\tau(T,p)$. Please clarify.

Eq. 3: Is this should be τ^{AOD}? (i.e., not "AOD = … ")

L212-213: Some justification of this assumption is needed.

L227: What is the major difference between these IUP_A and IUP cross-sections? Some description is needed.

L233: This could be a typo, i.e., "180 233 K". Please double check.

L248: The meaning of "(k = 1)" is not described (same request to the "k=2" in the later part of the paper).

L269-274: It might not be very straightforward for the reader to understand why IUP was selected (by WMO), if DBM shows less temperature dependency. Could you please provide some comments? For example, as I know, Pandora is using DBM.

L277-279: I am a bit confused here. Is the effective T been retrieved or not? If it is retrieved, the results should be presented, and some description is needed.

L 326-328 and Figure 5: The right panel has y label as "Standard Deviation/Uncertainty". What is this "uncertainty"? Is this the nominal one (i.e., "arithmetic results")? But the lines in the paper described it simply as "uncertainty", not "Standard Deviation/Uncertainty". Please clarify.

L358-359: I think this "averaged TCO" should be "averaged relative difference (of TCO)".

L362-364: The results here are very interesting and important. I do not want to be hypercritical. However, this information/claim might be very misleading too. First, lower variability for the Dobson-QASUME (D/Q) pair only can prove they (these two instruments) have a similar response (to the set of ozone cross-sections studied here). Second, the results depend on which cross-section is used. If we only select IGQ and IUP, the differences are very small (I doubt if the difference is statistically significant). Is this sensitivity due to instrument or due to cross-sections, or both? Anyway, the point is there is only one true ozone value. For any instrument, being sensitive to a "non-ideal" ozone cross-section might not be a bad thing. I would suggest rephrasing this part as "Brewer is more sensitive to some of the ozone absorption cross-sections (e.g., …) than … ".

Figure 7. Based on previous publications, Brewer has a remarkably low temperature-dependency (e.g., Kerr 2002), or at least "theoretically" better than Dobson (even with Bass&Paur, or IGQ here). However, the top panel in Figure 7 show that QASUME agrees better with Dobson; it shows a stronger seasonal structure when compared with Brewer. To me, this indicates QASUME data has a similar level of temperature dependency to Dobson. Any comments? As previously stated, I am not sure if the current QASUME algorithm retrieved effective T or used interpolated values (from radiosondes). The Pandora team seems also work on direct retrieve effective T, but had many challenges (at least, no published results or dataset yet). I would point out that Kerr 2002 also retrieved effective T, which shows pretty nice agreement with ERA reanalysis results (well, my unpublished research). It is very sad to see some knowledge in this community is not properly adapted and get lost.

Figure 7. I did not find any description of the red shading area in the paragraph. The caption here says "red bar" indicates D/Q comparison is consistent for all cross-sections. But even only eyeballing the areas, I could see some differences. Please clarify the meaning of the red shading area and provide necessary discussions. It seems this period (red shading areas) shows an opposite seasonality (decreasing with time) when compared with B/Q comparisons (increasing with time). The very strange/visible dimple in the red shading areas looks very likely due to the inaccurate effective temperature being used in the algorithm (or retrieved via the algorithm?). Please note, this feature is not shown in the B/Q comparison. Some investigation is needed. Also, please make the x-axis in Figure 7 has minor ticks. Otherwise, it is very difficult to tell the time of the observations/comparison.

Figure 4 says that when using IGQ cross-section, QASUME has a similar temperature-dependency as when using IUP or IUPA (i.e., about 0.1%/K). But, this may not be reflected in Figure 7. For example, Figure 7 shows that IUPA and ACS might have the best (lowest) relative seasonality, when compared with Brewer. Also, although when using IUPA, there is a 1% relative offset between B/Q and D/Q pairs (Fig. 7, 4th row), the seasonality difference is lower than the results for IUP (i.e., Fig. 7, 5th row, left panel). The worries are the good agreement between Brewer and Dobson (with use IUP cross-section,) might be due to wrong reasons. To me, IUP_A might be a good choice too. Relative seasonality between two instruments is always a clear indication/signal that one of them is wrong (at some level). Please provide some comments and reasons (if possible).

L370-371: I think that starting from here, all analyses were done with the IUP cross-section (QAUME, Brewer, and Dobson). If so, please include this information in the caption of Figure 8. Another question

is more challenging. I.e., do you see different slant ozone dependency when using different cross-sections? If yes, some results could be shared in this work (e.g., with a table, or bar plots). Some simple quantification could be made, e.g., using the parameters of the fitted lines in Figure 8.

L375-377: Unfortunately, I could not agree with this. Again, there is only one true ozone, although we cannot know the truth. But, less or even more sensitivity to many different (and selected) ozone cross-sections cannot prove it is an "advantage". Well, the finding itself here is important, but I would suggest phrasing the message carefully.

L377-382: This is an important finding, i.e., constant bias when compared with Dobson (no matter which cross-section was selected). If the IUP cross-section is the future WMO standard for both Dobson and Brewer data, do we expect to see QASUME instruments will always have this 1% offset (for all sites?)? Another interesting thing is, with current results, this 1% offset is not related to the selection of cross-sections when compared with Dobson (i.e., only small changes in bias from -0.72% to -1.01% with different cross-sections). Given the standard uncertainty of only 0.8%, what are the potential sources for this large relative bias? Some further comments and discussions on this offset are welcome.

Technical correction:

L27: define WMO here; move the definition from L48-49 to here.

L29: change "1980's" to "1980s"

L57: please provide the temperature here.

L160: define SO_2 where it was mentioned the first time.

L280: the link is broken.

L575: change "standard eviation" to "standard deviation".

L354-355: please rewrite this sentence, it is a bit ambiguous.

L361: % sign is missing for the number -0.72.

Reference

Kerr, J. B.: New methodology for deriving total ozone and other atmospheric variables from Brewer spectrophotometer direct sun spectra, J. Geophys. Res., 107(D23), 4731, https://doi.org/10.1029/2001JD001227, 2002.