Response to Reviewer #2

We acknowledge the detailed and helpful comments of Referee #2 on our manuscript. Below you can find our specific answers to the comments in blue. We have also revised the manuscript according to the suggestions by both Reviewers. The changes are marked with red in the revised word document. For extended changes we have indicated the line-number of the revised manuscript. We have acknowledged the two anonymous reviewers in the acknowledgements.

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.

Many thanks for this appreciation of our study. We agree that the 1% bias of the traceable TCO measurements compared to Brewer and Dobson is crucial point of the publication. However, one should note that this 1% difference is well within the combined uncertainties of the TCO retrieval of Brewer, QASUME, and Dobson. In fact, a strict uncertainty analysis has so far only been performed for QASUME while it is still pending for Brewer and Dobson TOC retrievals. If we assume a very low 1% expanded uncertainty for Brewer and Dobson, covering 95% of the coverage interval, then the combined uncertainty when comparing QASUME (combined uncertainty of 1.6%) with any of those instruments would be sqrt(1^2+1.6^2)=1.9%, which is larger than the observed bias.

TCO from QASUME is retrieved with a well described procedure. We have investigated the relevant parameters that could have an impact on the TCO retrieval of QASUME, as seen in the manuscript. For example, potential sources (e.g. Angstrom parametrization instead of linear parametrization, or different standard atmospheres) could not explain the observed bias of 1%. One can also speculate that a systematic bias could originate from imperfect Langley-plot calibrations form the Brewer and Dobson due to unknown residuals of ozone in the Langley plot. However, such speculations exceed the aim of the publication and requires more substantial scientific efforts.

We have revised the manuscript accordingly (e.g. lines 391 – 399, 405-423).

General comment from our side:

Due to the comment of Reviewer 1 who suggested a better guidance for the reader, we have re-worded some sentence of the manuscript.

During the revision of the manuscript, we have added two corrigenda:

1. *For the final standard retrieval, we have used the SO_2 cross section HITRANS from Hermans (2019) (Line 203)*
2. *The uncertainty from solar spectra is corrected to be 0.196 % since the factor 0.5 was not included in the calculation (line 285, Table 1).*

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.
Done – the proposed literature was included in line 304.

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

Eq. 3: Is this should be τ_{AOD}? (i.e., not “AOD = … “)
Done

L212-213: Some justification of this assumption is needed.
The impact of potential spectral correlations are included and discussed in the uncertainty assessment from two different solar reference spectra. Also suggested by reviewer #1 we have justified or assumption in the revised manuscript (lines 218-219, 290 – 292)

L227: What is the major difference between these IUP_A and IUP cross-sections? Some description is needed.
We have added more description in the revised manuscript (lines 247 - 249).

L233: This could be a typo, i.e., “180 233 K”. Please double check.
Typo – Done

L248: The meaning of “(k = 1)” is not described (same request to the “k=2” in the later part of the paper).
We have clarified the meaning of K=1 and K=2 in the first subsection (3.1) of section 3.

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.
The ozone absorption cross-sections of Serdyuchenko et al. 2014 were chosen by the scientific advisory group for ozone and UV of the WMO because the consistency between Brewer and Dobson TOC retrievals was considerably improved with this particular cross sections, compared to the others investigated (Redondas et al., 2014, Gröbner et al., 2021).

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.
The effective ozone temperature is not retrieved. Indeed, during the development of the algorithm, we thoroughly tried to retrieve this parameter by the least square fit algorithm, but this attempt failed. We have clarified in the revised manuscript (Lines 290 - 307).

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.
We have clarified (Lines 360 - 370) and changed the figure.

L358-359: I think this “averaged TCO” should be “averaged relative difference (of TCO)”.
We agree and have clarified in the revised manuscript accordingly.

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 … ".

Indeed this is an important comment. We fully agree with the reviewer that the discussion on the results need to be formulated carefully. It basically cannot be prooven what causes the sensitivity on the cross-section for Brewer, Dobson and QASUME. We have rephrased the section more conservatively (Lines 395 - 399)

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?
We have stated the similar temperature dependency of Dobson and QASUME more clearly in the revised manuscript (line 292)

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.

As clarified earlier, the effective temperature is not retrieved by the QASUME algorithm. We have used the linearly interpolated values from Gröbner et al. 2021. In the revised manuscript, we have acknowledged the findings of Kerr 2002, who successfully retrieved effective ozone temperature from the Brewer. This retrieval from Brewer could not be done with other ground based TCO instruments so far and should be highlighted here. We also see the importance to retrieve effective temperature, however we made substantial effort to do this in the presented study, but we were not successful (see lines 398 – 308 and 175- 179)

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.
The red shaded are should only highlight the offset numbers and not the corresponding data points in the shaded area. We apologize for this ambiguity. The red shaded is removed. The values of the offset are discussed in the manuscript.

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.

We have used the same effective temperature from the balloon soundings for the Brewer, the Dobson and the QASUME retrieval. We have clarified this in the revised manuscript (line 383)

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.

We have adopted the figure accordingly.

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).

As shown in Gröbner et al., 2021, the Brewer is more sensitive to changing ozone absorption cross-sections than a Dobson or QASUME (See Figure 5 in Gröbner et al., and figure 2 in this manuscript). Thus, small errors in the cross-sections will have a larger impact on the Brewer than either QASUME or Dobson. This is also consistent with the data shown in Figure 7, where changes in ozone absorption cross-sections show larger variability when comparing QASUME to the Brewer than to the Dobson. The conclusion that can be drawn from Figure 7 is that none of the ozone absorption cross-sections investigated in this study produce perfect consistent results between the three instruments, so small errors (spectral as well as absolute) remain in all these cross-sections. The selection of the most consistent cross-section based on the lowest bias and smallest seasonal variability is the IUP cross-section, even though future work is clearly needed to resolve the still observed discrepancies between these three instruments.

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.

Indeed the slant path analysis is done with IUP cross section only since we have chosen this cross section for the standard algorithm. We have added this information in the caption.

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.

Many thanks for this suggestion. We have additionally investigated the slant path dependency (differences of the linear fit at 300 and 1200 DU slant path) of the individual cross sections as follows:

	Brewer 156	Dobson 101
IGQ	0.29%	0.418 %
DBM	0.07%	0.03%
ACS	0.12%	0.40%
IUPA	0.09 %	0.43%
IUP	0.02%	0.43%

The results show that the slant path dependency is mostly insensitive to the selected cross section for the Brewer as well as for the Dobson. Except for IGQ applied on the Brewer shows significant higher slant path dependency, which is attributed to the seasonal variation when using IGQ. Since the solar zenith angle is correlated with season and corresponding effective temperature, the seasonal variation in temperature is also reflected in the slant path dependency. This detail is not further explained in the revised manuscript.
However, we have stated the finding of mostly insensitivity of slant path dependency on cross sections in the revised manuscript (Line 405 - 407), which is of interest for the reader.

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.
We agree that the statement of the "advantage" is inappropriate based on the sensitivity analysis regarding cross section. We have reformulated the statement more conservative (304-309).

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.
See answer of the general comment. Preliminary comparison results at two other locations e.g., at the Observatory in Izãna, Teneriffe, Spain in 2016 and in El Arenosillo 2019 in Southern Spain showed also about 1% bias compared to Brewer. The final approved results will be published in a WMO report in 2022.

We have revised the manuscript accordingly (Lines 413 -419).

L27: define WMO here; move the definition from L48-49 to here.
Done
L29: change "1980's" to "1980s"
Done
L57: please provide the temperature here.
Done
L160: define SO_2 where it was mentioned the first time.
Done
L280: the link is broken.
Sone - The link is check and it is working.
L575: change "standard eviation" to "standard deviation".
Done
L354-355: please rewrite this sentence, it is a bit ambiguous.
Done
L361: % sign is missing for the number -0.72.
Done