Response to Referee #3:

Thank you to referee #3 for the helpful comments. Our responses are given below in black with the comments in blue. The new text in the modified manuscript is given in red (italicized).

Referee #3:

Major issues: (1) To assign the label of cloudy, clear or intermediate, the variation of O4 along the day is taken into account. I think this criterion can be stricter for GBS instrument than for SAOZ instrument due to their differences in FOV. As can be seen in figure 4 (although please, see technical comments about this figure), it seems that there are more "clear" data in the case of SAOZ than in the case of GBS. I was wondering if this fact could be due to the O4 criterion. In figure 3, it is quite surprising that for year 2011 clear, cloud and intermediate cases are quite close for both instruments but this situation changes considerably for 2013 and it is clearly different for 2017. But both instruments are located in the same observatory, how is possible that the number of clear/cloudy days in 2017 can be that different? Maybe the O4 criterion is too permissive for SAOZ and too strict for GBS? This could also have an effect in the difference on the bias for both instruments when compared to Brewer. If the algorithm is not properly working for SAOZ, some clear days can, in fact, be affected by clouds and that would explain the better agreement between SAOZCS and Brewer than GBSCS and Brewer.

For 2011, the GBS performed measurements from March to August, and SAOZ performed measurements from March to August. So the percentages of clear/cloudy measurements from two instruments were very similar. For 2013, SAOZ performed measurements from March to April; while, GBS performed measurements from March to October. So the difference in the percentage of clear/cloudy measurements in 2013 was due to the different measurement periods. Please note the y-axis on Figure 4 is not number of days, but the percentage of data (spectra) that has been identified as clear or cloudy. For 2017, UT-GBS has measurements from May to September, while SAOZ has measurements from March to October. The 2013 UT-GBS colour index calibration factor change was due to the old 1 metre fibre being replaced by a 10 metre slit-to-spot fibre. The 2017 UT-GBS colour index calibration factor changes are mainly due to the use of an extra diffuser to decrease the signal (to enable MAX-DOAS measurements). These technical details have been added in the paper (Section 3.1). We also agree with the referee that the optimized O_4 criteria could be different for these two instruments, but to make it a consistent comparison, we used the same criteria for both instruments. A more detailed study could be performed in the future to fine tune this criterion.

The shifting of the calibration factor in 2013 is due to the fact that a 10 m slit-to-spot fibre bundle replaced the old 1 m single fibre. The shift in 2017 is due to a 200-grit UV diffuser that was used to attenuate the light signal (to enable MAX-DOAS measurements). Details about all instrument upgrades are provided in Zhao (2017).

(2) To be sure that the effect you observe in the bias when CS is applied to GBS TCO is only due to the presence of clouds, have you take into account that most of cloudy days happen out of the summer?

We have taken this potential seasonal effect into account. We divided the data into summer and spring/fall by using the largest available SZAs, and compared the clear-cloudy differences from these two periods. The summer period is defined as having the largest SZA of the day less than 85° (May to August). In general, when only summer data are included, the impact of the cloud-screening algorithm can be clearly seen. Figures R3.1 and R3.2 are similar to Figure 5, but present data divided into spring/autumn and summer using the largest SZA in the Langley plot.

In general, from these tests, we confirmed that:

1) The clear-cloudy difference in summer is statistically significant, regardless of whether Brewer or MERRA-2 is used as a reference.
2) If we use MERRA-2 as a reference, the clear-cloudy difference in spring and autumn data is clear. But if we use Brewer as a reference, the clear-cloudy difference in spring and autumn is not significant (due to limited coincident measurements). For example, for Brewer vs. GBS in spring and autumn, we only have 33 coincident measurements in cloudy conditions.
3) The proposed cloud-screening algorithm uses three sky-condition labels (CI value label, CI smoothness label, and O_4 smoothness label). For spring-time (when SZA >85°), the CI value label is not available. Thus, the efficiency of the cloud-screening algorithm is higher in summer than in spring and autumn.

Some of this information has been added to the paper (Section 4.1.2):

Since cloudy days mostly appear in the summertime, sensitivity tests were performed with the dataset divided into summer and spring/autumn periods to assess whether there was any seasonal bias. In general, we found that the clear-cloudy difference is still statistically significant in summer, no matter which reference is selected (Brewer or MERRA-2). For spring/autumn, the clear-cloudy difference is statistically significant only when MERRA-2 is used as the reference, but not if Brewer is used as the reference due to the limited number of Brewer measurements given the large SZAs in spring and autumn).

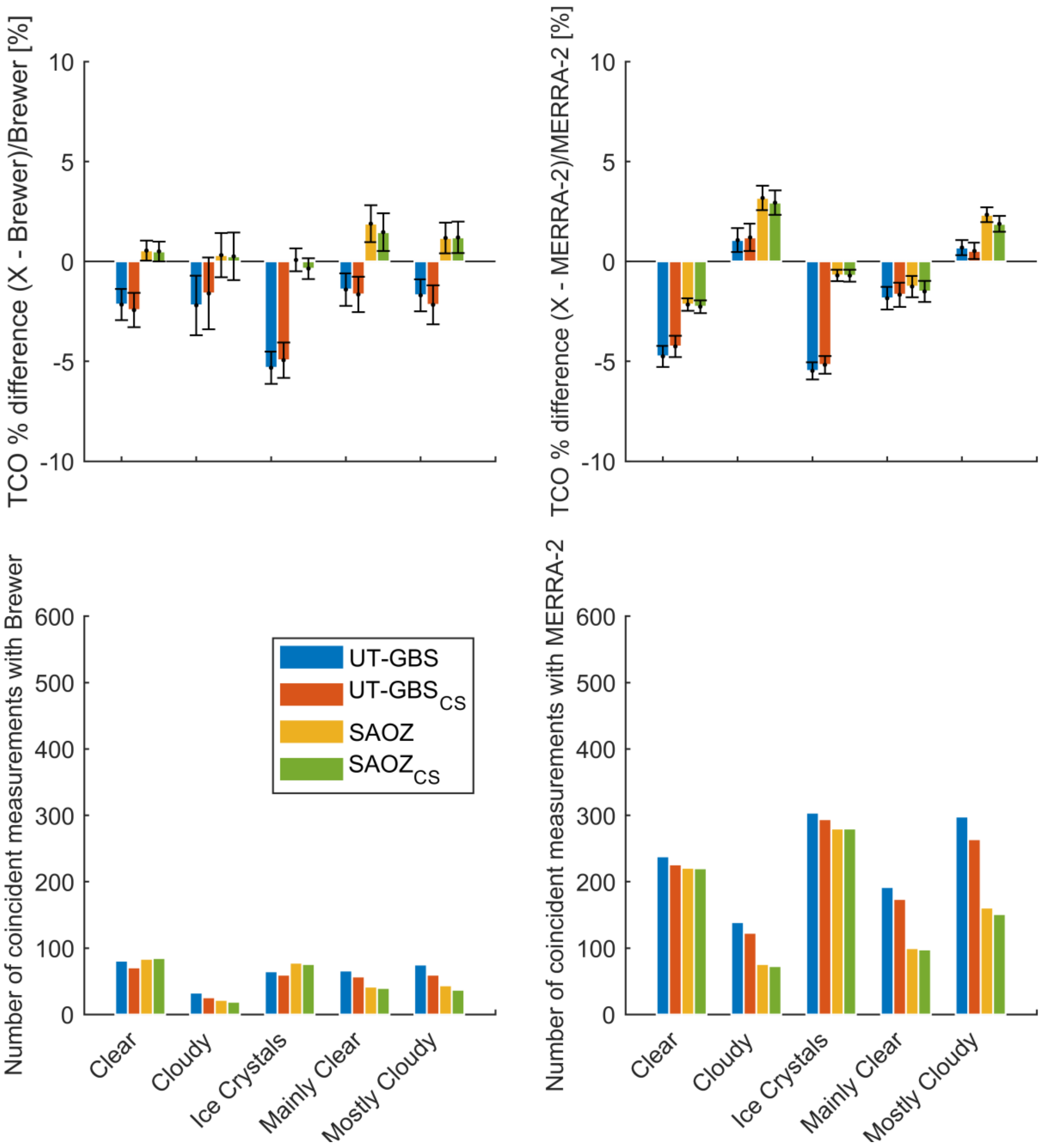

Figure R3.1. Same as Figure 5, but only including spring and autumn data (when daily maximum SZA > 85°).

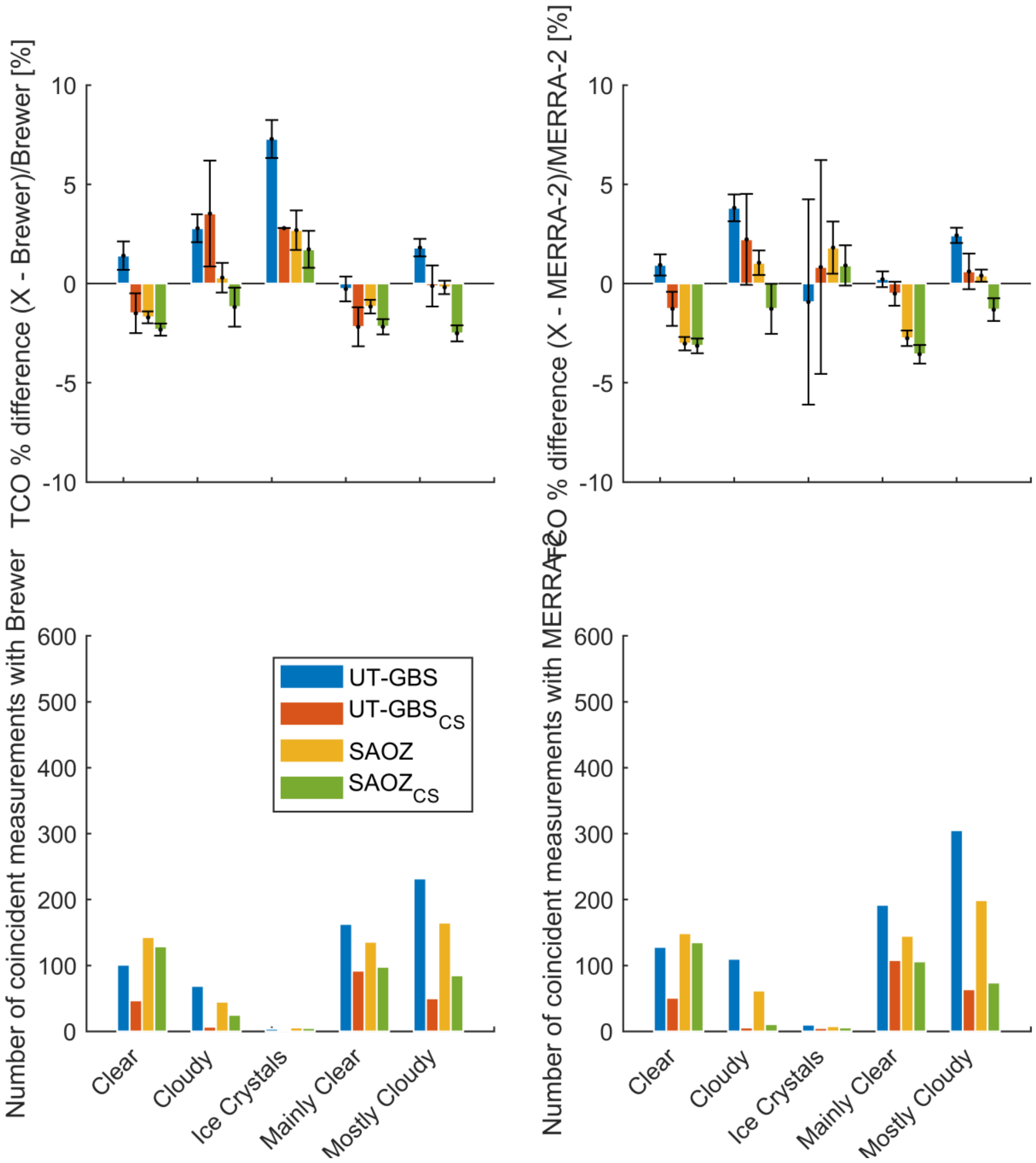

Figure R3.2. Same as Figure 5, but only including summer data (when daily maximum SZA < 85°).

What SZA do you use to calculate TCO at summer?

For summertime, when the NDACC-recommended SZA range was not available, we used the nearest available 5° SZA range. This information was previously provided in the manuscript. For example, on May 1, the SZA is in the range of 65° to 85°. Thus, we will use measurements made from 80° to 85° in the Langley plot.

Could the observed bias to Brewer have some to do with the major weight of summer days when you eliminate the cloudy days?

We agree with referee that the observed bias to Brewer may be to its greater weighting towards summer days. However, the bias due to Brewer measurements is inevitable for several reasons. First, the Brewer had limited springtime measurements (it only provides measurements when SZA < 82° , as stated in the manuscript). Second, the Brewer cannot perform measurements when heavy clouds block the solar beam. Thus, Brewer measurements are biased to summer and clear-sky conditions. This is the reason we included MERRA-2 in this work. For any study that only uses Brewer data to compare with NDACC-type UV-vis measurements, it is hard to assess the cloud impacts.

Minor issues: (1) Due to the high latitude of the observatory it is not possible to have DOAS measurements along the entire year. Please, in the description of the instrument include what is the annual period of measurements. From figure 4 and from data along the text it seems that the period is late winter to late autumn? It would be nice to know the months when DOAS and Brewer can measure.

The Brewer typically can provide measurements from April to August, while GBS and SAOZ can provide measurements from March to September. This information has been added to in Section 4.

The Brewer instrument at Eureka typically makes measurements from April to August, while UT-GBS and SAOZ can provide measurements from March to September.

(2) Section 4.1. Why the current agreement to Brewer and GBS is better in this work than in the previous work by Adams et al.?

The result (-1.4%) in Adams et al. (2012) was based on measurements from 2004 to 2011. For the current study, the result (-0.23%) is based on measurements from 2010 to 2017. There are several possible reasons for the improvement, such as year-round variability, improvement due to new NDACC ozone LUT, and more summertime measurements in the current datasets. During the 2004 to 2006 period, only springtime measurements were available. For the 2007 to 2009, the instrument was using a different grating for the summer measurements. In general, we could not apply the new cloud-screening algorithm to the data before 2010, thus we did not include 2004 to 2009 data in the current work. The 2004-2017 GBS data were reprocessed and used in a satellite validation paper (Bognar et al., 2018, submitted to JQSRT). In that work, we find that for the 2004-2017 period, the mean relative bias between GBS and Brewer is -0.9%, which is closer to the number reported by Adams et al. (2012). Also, Adams et al. (2012) defined the mean relative differences (Δ_{rel}) as:

$$\Delta_{rel} = 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{(M_{1i} - M_{2i})}{(M_{1i} + M_{2i})/2},$$

where N is the number of measurements, M_1 and M_2 are sets of coincident measurements. In Figure 5 (AMTD version), the mean relative difference was defined as:

$$\Delta_{rel} = 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{(M_{1i} - M_{2i})}{M_{2i}},$$

where M_1 was UT-GBS (SAOZ), and M_2 was Brewer (MERRA-2), indicated by the y-axis label (see the AMTD version).

To make this study directly comparable with Adams et al. (2012), we have revised Figure 5 and the relevant numbers (using the same Δ_{rel} definition as Adams et al. (2012)). These changes do not affect the conclusions.

Following Adams et al. (2012), the agreement between sets of coincident measurements (M_1 and M_2) was evaluated using the mean relative difference, defined as

$$\Delta_{rel} = 100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{(M_{1i} - M_{2i})}{(M_{1i} + M_{2i})/2}, \qquad (4)$$

where N is the number of measurements.

(3) Taking into account the current results, it seems that in the case of Hendrick et al., not all the observed discrepancies between DOAS and Brewer were due to the temperature dependence of XS used in Brewer analysis or in this work the Brewer analysis takes into account this dependence?

The Brewer data used in this work were processed by the standard Brewer algorithm. The temperature dependence due to the ozone cross section does exist in this Brewer dataset. This temperature dependence is different from instrument to instrument. Currently, we do not have an estimated temperature dependence factor for the Brewer instrument used in this study, so no temperature correction was applied.

The temperature dependence of Brewer data also depends on the location of the site. For example, if we assume the temperature dependence of a Brewer is 0.1%/K (as reported in previous studies, e.g.

Kerr, 2002), for a year-round 15 K stratospheric effective ozone temperature variation, the temperature dependence introduced by seasonal changes in TCO will be 1.5%. However, for Eureka, the Brewer only performs measurements from April to August, and so the temperature effect at Eureka is expected to be smaller (compared to year-round mid-latitude measurements). We calculated the effective ozone temperature (based on the method shown in Zhao et al., 2016) for 55°N and 75°N using ozone and temperature profiles from the Max Planck Institute for Chemistry (MPIC, Brühl and Crutzen, 1993) climatology to illustrate this. As shown in Figure R3.3, the estimated temperature-induced bias in Brewer TCO at 75°N is only 0.9% (while for 55°N, this is increased to 1.4%). Thus, to further separate the temperature dependence, cloud effect, and other potential seasonal effects, we will need more accurate temperature and pressure profile measurements or modelled values for Eureka.

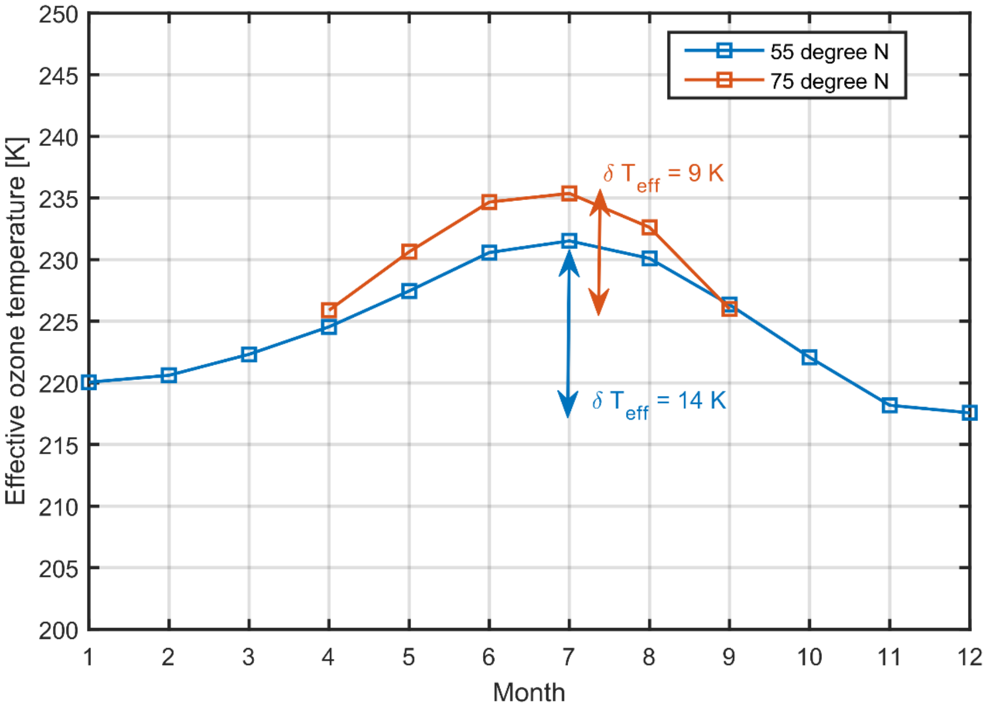

Figure R3.3. Simulations of year-round effective ozone temperatures (T_{eff}) at two latitudes based on climatological ozone and temperature profiles.

Do you observe also the same seasonal difference (taking into account that you cannot observe the entire spring and fall at 80°N) that Hendrick et al. in the bias against the Brewer?

The seasonal difference between UV-vis TCO and Brewer TCO at Eureka is weaker than reported values measured at mid-latitude sites (e.g., Hendrick et al., 2011). Figure R3.4 shows the ratio of SAOZ and Brewer TCO over the period 2010 to 2017.

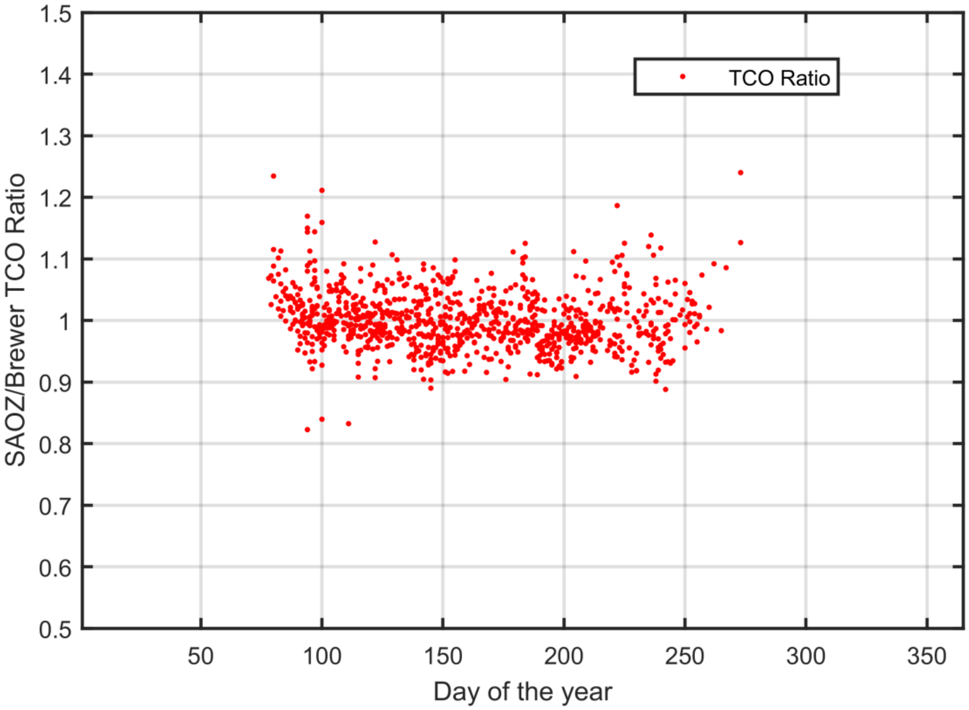

Figure R3.4. SAOZ/Brewer total column ozone (TCO) ratio as a function of day of the year for the period 2010 to 2017.

(4) Section 4.1.1, please indicate at any part of the text that the weather classification used here and in figure 5 is made by using meteorological data. If not, it is a little confusing.

The following text has been added in Section 4.1.1:

The weather classification used here and in Figure 5 is based on hourly observations of sky conditions made by a meteorological technician at Eureka.

Technical issues:

(1) Figure 4. Please, unify ticks in the horizontal axis. The lower graph is different from the previous ones and this makes very difficult to see properly the measurement periods. Grid in the middle of each year would be also very helpful. Colours in the legend are not coincident with the ones in the graphs. As GBSCS or SAOZCS are over imposed to GBS and SAOZ respectively, it seems that there are more data

for the CS filtered data than without any filter. This is a little bit confusing at first, I am not sure that it can be addressed, maybe using hollow symbols for CS case? If possible it would be nice a greater graph.

Figure 4 has been revised as suggested.

(2) Sometimes the DOAS instrument GBS is called UT-GBS, please unify nomenclature along the text.

UT-GBS has been adopted throughout.

References

Adams, C., Strong, K., Batchelor, R. L., Bernath, P. F., Brohede, S., Boone, C., Degenstein, D., Daffer, W. H., Drummond, J. R., Fogal, P. F., Farahani, E., Fayt, C., Fraser, A., Goutail, F., Hendrick, F., Kolonjari, F., Lindenmaier, R., Manney, G., McElroy, C. T., McLinden, C. A., Mendonca, J., Park, J. H., Pavlovic, B., Pazmino, A., Roth, C., Savastiouk, V., Walker, K. A., Weaver, D. and Zhao, X.: Validation of ACE and OSIRIS ozone and NO_2 measurements using ground-based instruments at 80°N, Atmos. Meas. Tech., 5, 927–953, doi:10.5194/amt-5-927-2012, 2012.

Bognar, K., Zhao, X., Strong, K., Boone, C. D., Bourassa, A. E., Degenstein, D. A., Drummond, J. R., Duff, A., Goutail, F., Jeffery, P., Lutsch, E., Manney, G. L., McElroy, C. T., McLinden, C. A., Millán, L., Pazmino, A., Sioris, C. E., Walker, K. A. and Zou, J.: Validation of ACE and OSIRIS ozone and NO_2 measurements using ground-based instruments at Eureka, submitted to J. Quant. Spectrosc. Radiat. Transfer, 30 pages, 2018.

Brühl, C. and Crutzen, P. J.: MPIC two-dimensional mode, in The atmospheric effects of stratospheric aircraft, vol. 1292 of NASA Ref. Publ., pp. 103–104., 1993.

Hendrick, F., Pommereau, J. P., Goutail, F., Evans, R. D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M. and Van Roozendael, M.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations, Atmos. Chem. Phys., 11, 5975–5995, doi:10.5194/acp-11-5975-2011, 2011.

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

Zhao, X., Fioletov, V., Cede, A., Davies, J. and Strong, K.: Accuracy, precision, and temperature dependence of Pandora total ozone measurements estimated from a comparison with the Brewer triad in Toronto, Atmos. Meas. Tech., 9, 5747–5761, doi:10.5194/amt-9-5747-2016, 2016.