We thank all reviewers for their input, which we feel has clarified aspects of the manuscript and improved it. As we edited the text, we noted a few other areas where clarity could be increased. For instance, there were some discrepancies in significant digits that has been addressed.

A note to anonymous referees, any line numbers mentioned in the responses correspond to the track changes documents.

Review #2

Brockway and coauthors present a large set of AMAX-DOAS profiles of BrO leveraging a "porpoising" flight pattern to extensively sample the surface to 1 km altitude with high resolution. BrO profiles are divided into four clusters highlighting shallow near-surface BrO enhancements, more mixed boundary layers, and lofted layers of BrO associated with particles. There is a thorough and high-quality discussion of meteorological and chemical factors and the implications for satellite retrievals and modeling of Arctic BrO. The main text and supplement together indicate that the underlying BrO are likely sound, however, a more quantitative approach to uncertainty is needed to contextualize the underlying data.

I have classified this as a major revision because it is necessary to ground the underlying data, and it is not yet clear that all key points are supported.

Major revision:

- *More quantitative information is needed to demonstrate the validity of some retrieval choices taken by the authors. These are potential sources of uncertainty and should at least be bounded.*

1. *As the authors acknowledge in the supplement that the DOAS fitting of HCHO and BrO can have a tight anti-correlation arising from the similarity of their cross-sections (e.g. (Pinardi et al., 2013)). The authors note that HCHO is not significantly detected, and thereby infer that it can be omitted from the fits. However, this inference is not necessarily valid, without further information it is possible that the high BrO signal is limiting the ability to retrieve large HCHO columns. The critical criterion is the impact of including HCHO on the BrO dSCDs. The authors should report this and compare it to the fit uncertainty and/or the BrO columns themselves to bound the impact of this choice.*

HCHO concentrations near Utqiagvik in springtime are generally low, due in part to the lack of emissions (biogenic and anthropogenic) and the abundance of gas-phase halogens. Barret et al., 2011 observed a maximum of 200 ppt HCHO at the surface in Utqiagvik on a single day that corresponded with an ozone depletion event. To determine the impact of this much HCHO on our BrO observations, we added synthetic HCHO absorption to two different measurement spectra, one at low altitude and one at high altitude. HCHO SCDs were determined assuming well-mixed HCHO in the lowest 2 km with the modelled BAMFs and scaling the absorption cross-section to the calculated SCDs. Spectra were then fit with the original fit retrieval. In both scenarios, the maximum HCHO SCD added was roughly 1×10^{16} molecules/cm^2 which corresponded to 1.37×10^{13} and 1.63×10^{13} molecules/cm^2 for the low altitude and high altitude case, both less than the 1σ fit uncertainty. This is also much lower than the typical near-limb BrO dSCDs near the surface, which are on the order of 2.5×10^{14} molecules/cm^2. This has been summarized in lines 14-15 of the supplement.

Adding HCHO to the fit algorithm had the impact of increasing BrO dSCDs. For one flight (April 1, 2022) where we observed high BrO and flew through a NOx plume, we ran the fit routine with a HCHO reference. This resulted in an increased BrO fit uncertainty from 2.68×10^{13} to 2.88×10^{13} molecules/cm^2 and mean increase in the BrO dSCD of 1.7×10^{13} molecules/cm^2 due to mostly negative fit coefficients applied to the HCHO reference (on the order of -1×10^{16} molecules/cm^2). The change was therefore well below the detection limit of BrO. This point has been summarized on lines 15-17 of the supplement.

The negative HCHO fit coefficients are also observed in areas of high NOx where we might have expected collocated HCHO emitted from fossil fuel processing. Further, in these NOx plumes where we might expect to observe HCHO, BrO dSCDs are drastically reduced even without including HCHO in the fit routine, likely due to a direct gas phase reaction of BrO and NO_2.

Overall, the impact appears to be small and much less than other uncertainties in this work. An abridged version of this information has been added to lines 14-18 of the supplement.

2. *The authors choose to use a single constant stratospheric BrO profile for each flight. The authors note that flights are near local noon when stratospheric BrO is roughly constant. However, the springtime Arctic is expected to have high SZA (I estimate ~70° for most data from the information given but possibly more near takeoff) and therefore there might be a strong leverage on this. The authors should bound the possible impact of this choice or else provide more detail in the supplement about whether high SZA data were filtered, or what the range of SZA sampled to support their assertion.*

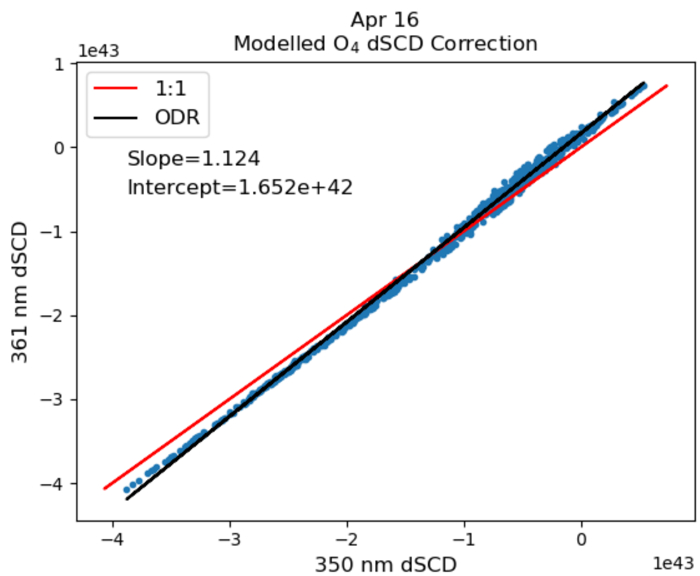
The median SZA of the observations during this field campaign was 70°, and the mean change in SZA during a flight was 6°. Due to the path length enhancement in the lower atmosphere of our near-limb observations, we chose not to account for diurnal variation in the stratospheric BrO profile, and since we are solving for the concentration profile near the surface, no diurnal variation needed to be applied in the lower atmosphere. Early in the campaign, some observations were made at SZAs greater than 80° (4.3% of all observations, with a maximum observation SZA of 82.2°) and these observations are not filtered out. To determine the impact of diurnal variation, we used the Theys et al. (2009) climatology with the minimum and maximum observed SZA for each flight and calculated a roughly constant 10% (1.35 ppt at the peak) uncertainty throughout the stratosphere. In this study, we propagated a 33% uncertainty in the BrO profile due in large part to the satellite stratospheric NO_2 observations. As a result, the uncertainty caused by diurnal variation is well within the assumed uncertainty of these profiles. This has been added to the supplement section "BrO Error Propagation and Sensitivity Studies" (lines 170-176).

3. *The authors have employed a random sampling of the flight to retrieve a single aerosol profile for each flight. Statistically, this approach minimizes the bias of the average of*

the random sample. Some information to assess the success and/or validity of this approach for individual profiles is needed. Could the authors provide statistics or even better an example graph comparing measured and modeled O_4 dSCDs? As above, can the authors provide an estimate of the uncertainty arising from imperfect aerosol retrieval, or is this already propagated in the BrO optimal estimation? If it is the latter that is not clear. The recently published O_4 cross section (Finkenzeller and Volkamer, 2022) has revised the bands included for the fits here compared to the prior cross-section used (Thalman and Volkamer, 2013) would this impact the results?

As stated in Finkenzeller and Volkamer, 2022, the novelty of their work is the extension of the O_4 cross-section to lower wavelengths, and the peak absorption centered around 361 nm is largely unchanged. However, the absorption band centered near 341 nm differs more significantly between the two works, and this is in the fit range used for the BrO and O_4 retrieval. Applying the 263 K O_4 absorption cross-section from Finkenzeller and Volkamer, 2022 to an entire flight (from April 1) resulted in no appreciable difference in O_4 or BrO dSCDs. The root mean square difference between the O_4 dSCDs with the different cross-sections is 1.7×10^{41} molecules^2/cm^5, with an R^2 near unity and a fit slope of 1.006. BrO dSCDs are similarly unaffected with a root mean square difference of 3.9×10^{12} molecules/cm^2 and a fit slope of 0.978. For both dSCDs, the fit error is slightly reduced (<2%) with the newer O_4 cross-section. We are therefore confident that a full reanalysis is not needed, and the differences seen in both O_4 and BrO dSCDs are well within the dSCD uncertainties attributed to the cross-sections (5% for O_4 and 9% for BrO). The supplement now states that Finkenzeller and Volkamer (2022) should be used in the future (lines 21-23).

To discuss the validity of the O_4 inversion, Figure S3 has been added to the supplement showing the observed vs modelled O_4 for the 200 data point (training) subset of each flight and for all porpoises throughout the campaign. To calculate the modelled dSCD for all porpoises, the BAMFs were used with the a priori O_2 profile (based on temperature and pressure), whereas the Lambert-Beer Law is used for the training subsets. This linearization could lead to some uncertainty between the two calculations. However, a larger impact arises from the fact that the training subset is modelled at 361 nm at a major O_4 absorption peak while the BAMFs are modelled at 350 nm for BrO. To account for this difference, the dSCDs modelled using 350 nm BAMFs were fit against the training 361 nm modelled dSCDs for each flight individually. The relationship between the two modelled dSCDs was largely linear, with the 350 nm dSCDs generally higher for low O_4 absorption values and lower for high O_4 absorption. An orthogonal distance regression was used to convert the 350 nm dSCDs for all porpoise observations to 361 nm. An example of this dSCD conversion is shown in the figure below. The average slope for this fit from all flights was 1.108, very similar to the Rayleigh extinction factor between these two wavelengths, i.e. $(361/350)^4 = 1.132$.



The RMSE for the entire porpoise dataset between observed and modelled (361 nm) dSCDs is 4.6×10^{42} molecules^2/cm^5. The following figure has been added to the supplement (line 133) to show how well the modelled O_4 represents the observations, and these statistics have been added to the main text (lines 244-245). The R^2 listed in this figure is also lower than that in the text, as the original manuscript listed the mean R^2 for all flights while this figure lists the R^2 of all training and porpoise observations together. Statistics have been updated in the main text, along with a reference to this figure (line 241).

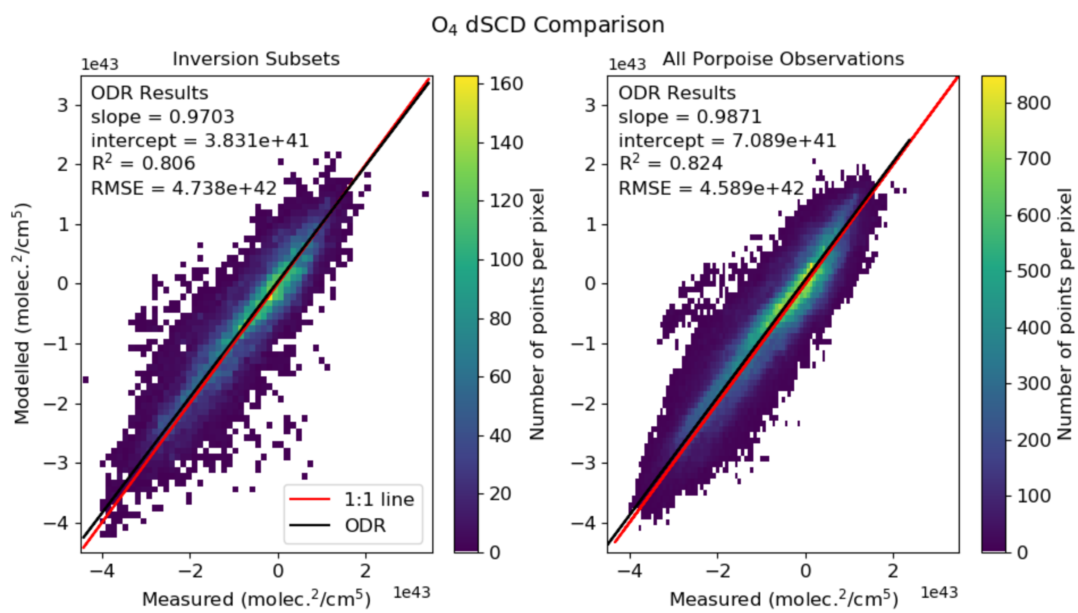


Figure S3

The particle extinction inversion results in an average relative uncertainty of roughly 20%. To account for the extrapolation of the retrieved extinction profile to points outside of the trained subset, this error is doubled in any figures. Further, a 50% uncertainty was used to determine the impact of particle extinction uncertainty on BAMFs. Sensitivity studies were performed to determine the impact of several uncertainties on calculated BAMFs. The resulting BAMF uncertainty was then used to vary the BAMFs and calculate the impact on the BrO inversion to include as a forward model error. The resulting variability in the BrO profile from sensitivity tests was then added as an additional uncertainty term to all BrO profiles. A new section has been added to the supplement (BrO Error Propagation and Sensitivity Studies) to discuss these sensitivity tests and a new section has been added to the results section (3.2 BrO Profile Uncertainty) of the main text to discuss the uncertainty of the individual BrO profiles.

- *Uncertainty needs to be addressed when examining differences and variability. How does the uncertainty of individual profiles compare to the variability in the clusters, e.g. in Figs. 5 and 7? In the current manuscript, there is not a clear demonstration that the increase of BrO with altitude in cluster-4 data is significant – the best case being a comparison of 50-100 m to 200-250 m in Fig. 7. If there is not a significant increase, then what is the meaningful difference between cluster 3 and cluster 4? The variability of the retrieved profiles can obscure significant differences, but some assessment of the significance of the underlying profiles is needed. How much of the increase in BrO in lofted layers arises from a decrease in light path from aerosol, and how much is driven by constant or increasing BrO dSCDs? The significance of this result especially needs more context and support.*

The K-means clustering algorithm that groups the BrO profiles is unsupervised, and simply looks for common profile shapes to minimize differences between the members of a group and the group mean. More information on the silhouette score test used to select four clusters as the best choice has been added to the text (lines 304-309). While it is true that one cluster may not significantly differ from another cluster at a given level, taking into account the entire profile shape reveals the differences between the clusters. For example, averaging the Lofted BrO cluster and the High Total BrO cluster between 200 and 400 m results in significantly different values, even accounting for the variability of the underlying profiles. The silhouette score test ensures that no two clusters are too similar, which is why we utilized the second highest score of four clusters.

While it is true that clustering the profiles can obscure some variability of the underlying profiles, the purpose of this work is to benchmark the general BrO profiles observed during this field campaign. A section (3.2) discussing the uncertainty of the underlying BrO profiles has been added to the results. In general, the uncertainty of the BrO profiles is comparable to the cluster variability above 300 m, and is lower than the cluster variability below 300 m. For this reason, the cluster variability has been shown as the error bars on the graphs.

As for the lofted profiles on 3-19, the retrieved profiles are influenced both by high dSCDs well above the surface as well as increased particle extinction. As discussed, this day was associated with higher particle extinction. Therefore, the same BrO dSCD as observed on a more normal day would result in a higher BrO mixing ratio. A figure discussing the dSCD profile has been added to the supplement so that readers can see the underlying data that supports the lofted BrO profiles (page 10 of the supplement).

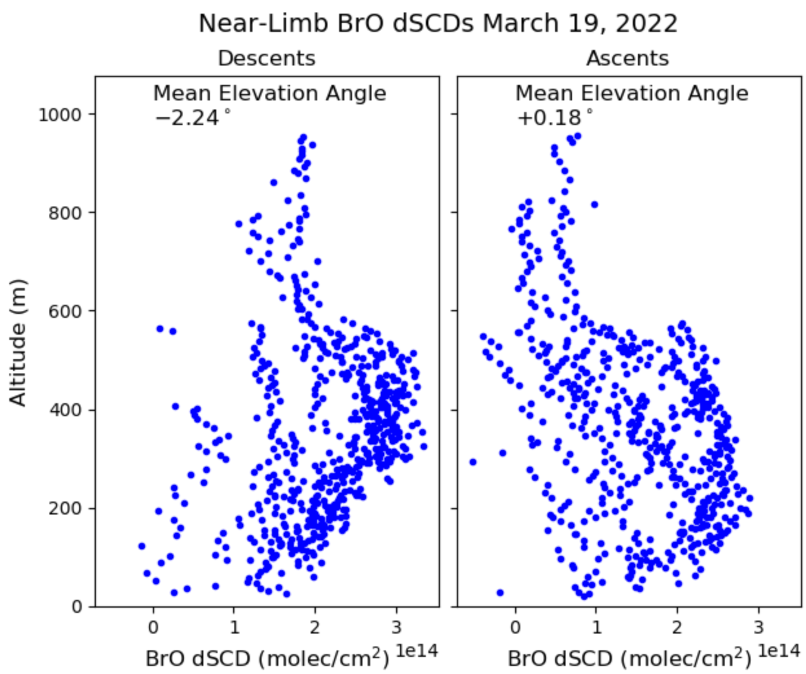
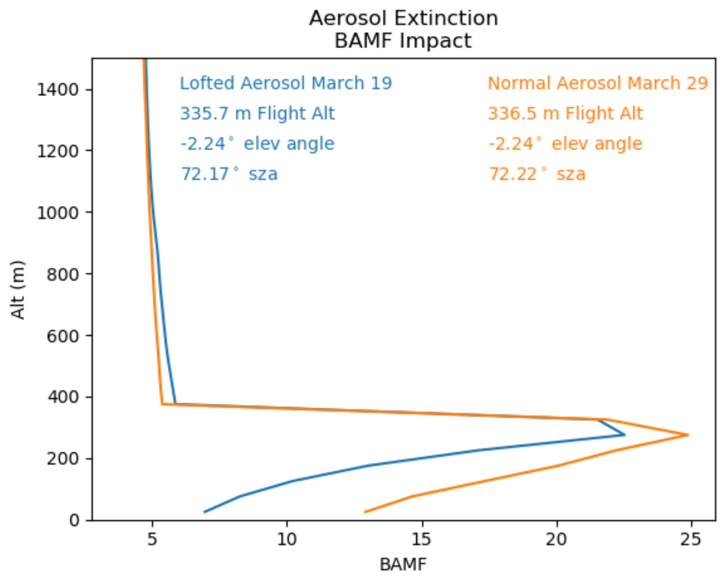


Figure S7

An example BAMF calculation is shown below with comparable geometry on a high aerosol day (3-19) and mean aerosol day (4-1). The BAMF is decreased on 3-19 due to an optically thicker atmosphere. But the difference does not appear large enough to create a lofted BrO layer from low dSCD values. It is difficult to determine the overall effect of these BAMF differences on BrO profiles. But as the increased particle extinction does not greatly impact the shape of the BAMF profiles but more the values, we expect the increased particle extinction to impact the magnitude of the retrieved BrO profile more than the shape.



Minor revisions:

Line 246-248. Grid resolution and DoF do not interact linearly in this manner. The resolution can be partly inferred from the off-diagonal terms in averaging kernels which from the example in Fig. 3 show the resolution is roughly equally valid for the full altitude range. From the examples provided in kernels which peak at lower values are also broader and that the slight loss of resolution is in the middle of the profile. This is roughly as expected for porpoising maneuvers with a rigid telescope which will vary pitch away from horizontal in the middle of the profile.

If this is the basis of capping Fig. 4 and other figures at 850 m, they should be extended if there are sufficient statistics at higher altitudes.

This sentence has been removed from the text.

Figure 4 is capped at 850 m because not all porpoises rose to this level. Extending the figure further just shows more influence of a priori assumptions.

Line 254 -257: It appears inconsistent that inhibited vertical mixing can explain high concentration below 50 m, but consistent mixing is invoked to explain near constant mixing ratios in the next 150 m. Can other hypotheses be offered, or else some discussion of variability in vertical mixing? From the cluster analysis it appears that this is a consistent feature of all clusters so I would suggest formulating a different hypothesis. Is it an effect from the 23% of data in clusters 3 and 4, or is it a result of an active source? This is partly addressed in Sects. 4.2 and 4.5, but the description should be consistent in different parts of the text. As discussed in Sect. 4.5 discussion is somewhat limited by filtering data from the lowest altitudes. Some more discussion of cluster 3 might be useful to addressing this.

The text now states that this may be due to a residual layer above the common surface inversions (lines 277-278).

Line 258: Above 250 m, the near-zero BrO omits some potentially important detail. Most of the retrieved DoF are above this altitude. How do the medians and ranges compare to the a priori?

The text now states: "BrO decreases to the a priori value of 1 pmol mol^{-1} with mean values comparable to the standard deviation" (lines 279-280).

Line 414: A recent publication (Wales et al., 2023) examines satellite and model surface BrO in the Arctic. Do the authors believe the latest findings address the limitations they outline?

This manuscript emphasizes the limitations listed in Wales et al. (2023), which discusses a case where their method was unable to reproduce large, near-surface BrO mixing ratios observed with ground-based instrumentation in Utqiagvik. Similarly, Table 4 of that study shows the lowest correlation between their modelled value and 0-200 m BrO columns observed via O-BUOY data. Our manuscript also suggests that satellite columns are likely less sensitive to surface-based differences in BrO, such as this difference between clusters 1 and 2. The reference has been changed to Wales et al. (2023) as this is a much more recent publication to calculate tropospheric BrO amounts (line 458).

Sect. 4.8: Some discussion of deeper AMAX-DOAS profiles such as those in (Volkamer et al., 2015) and (Koenig et al., 2017) is warranted. While it focuses on CHOCHO rather than BrO, (Volkamer et al., 2015) also includes a case study comparing AMAX-DOAS, surface MAX-DOAS (shipborne), and in situ (CE-DOAS) detection which is relevant to the discussion here.

More literature of MAX-DOAS DOFs has been added to the Section 4.8 as well as the impact of raising the MAX-DOAS platform above the surface (lines 475-477, 486-488).

Technical comments

Line 151: "dimer" should not be used to describe O_4 for the pressures and temperatures in Earth's atmosphere. The other language used such as "associative collision" is more accurate, but it is best described as "collision induced absorption".

This change has been made in the text at line 166.

Line 154: The mean elevation angles are not evenly spaced, as such it seems unlikely that the field of view is the same for all angles. Can the language be clarified?

The view is truncated at both ends, so the first and fourth elevation angle have decreased input on the extreme ends. As a result, we used the mean incoming elevation angle instead of the geometric center of each view to better represent the viewing angle of the telescopes. The reviewer is correct that they do not necessarily have the same overall field of view. The four views all have the same FWHM as calculated with field data, as now stated for clarity on lines 169-170 of the revised manuscript.

Line 183-184: Does the horizontal distance estimate include the mean light path averaged over by the BrO measurement, the flight distance, or both?

The manuscript has been updated to explain that this is only the flight distance (line 199).

Line 219: Can the authors provide more detail on the random sampling? What fraction of total measurements are the randomly selected 200? Are the 200 measurements selected independently or are later selections modified to ensure coverage?

The 200 points were selected randomly for the porpoises, which generally ensures coverage. Full coverage of the range of flight altitudes was also confirmed after model runs. The average flight consisted of 4,141 observations, so the training dataset was roughly 5% of the entire dataset. This has been included in the revised text (line 236).

Fig. 2: The Rayleigh extinction profiles should be shown for comparison.

This has been added to the figure. The new figure is shown below.

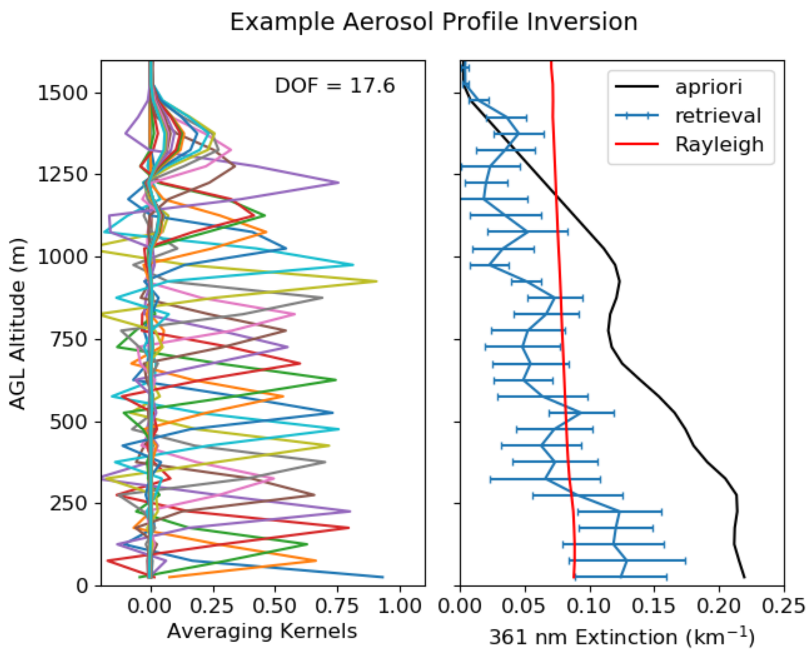


Figure 2

Line 368: I believe e.g. should be i.e.

This has been updated in the text (line 412).

Fig. 11: I recommend showing ozone to zero in the middle panel.

This change has been made to the figure.

Line 425: "prior profiles" here should be "a priori profiles"

This change has been made (line 470).

Supplement Fig. S1: The authors invoke a cutoff effect from intercepting the surface to explain the positive O_4 optical density. How relevant is this effect over a high-albedo surface? Since BrO (and perhaps NO_2) also are typically maximum near the surface why is this effect relevant for O_4 but not BrO or NO_2?

Even over a high albedo surface, light-path truncation occurs close to the surface, as the downwelling radiation is more a function of solar geometry. The reference spectra used throughout the field campaign are near-limb observations from roughly 1000 m altitude. The O_4 dSCDs often oscillate between positive and negative values based on the observation altitude and upward and downward aircraft pitch angles. In this case, the light path is truncated by the presence of the surface, so the observation "light path" is shorter than that of the reference spectrum. The same phenomenon occurs for the BrO and NO_2 observations, but the trace gas abundance at the surface is considerably larger than that aloft, leading to positive absorption relative to the reference. Clarifying information has been added to this figure caption.

References:

Barret, M., Domine, F., Houdier, S., Gallet, J.-C., Weibring, P., Walega, J., Fried, A., and Richter, D, Formaldehyde in the Alaskan Arctic snowpack: Partitioning and physical processes involved in air-snow exchanges, J. Geophys. Res., 116, D00R03, doi:10.1029/2011JD016038, 2011.