Comment on amt-2022-133

Chris McLinden (Referee)

Referee comment on "An improved OSIRIS NO₂ profile retrieval in the upper troposphere–lower stratosphere and intercomparison with ACE-FTS and SAGE III/ISS" by Kimberlee Dubé et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-133-RC2, 2022

Review of "An Improved OSIRIS NO2 Profile Retrieval in the UTLS and Intercomparison with ACE-FTS and SAGE III/ISS" by Dubé et al.

This is a solid paper on a topic that fits in AMT. It describes a new OSIRIS NO2 data product (version 7.2) that improves upon previous versions, addressing deficiencies in the retrieval methodology such as how the reference radiance is calculated and changes in spectral resolution with instrument temperature. Comparisons are made with co-incident observations from occultation instruments. It is well written and the subject is worthy of publication in AMT.

My only criticism of substance is that considering "… version 7.2, was designed to fix a low bias and to improve performance in the UTLS …" it really does not do an adequate job evaluating the performance of the new product in this region. This paper should attempt to better answer questions such as "precisely how good is v7.2 in the UTLS" or "is v7.2 actually better than the old version", and if this cannot be answered, then why not. This is the region where diurnal effect errors can be quite substantial. Of the three datasets used, it was ignored in OSIRIS and ACE-FTS, and corrected for in ISS/SAGEIII, which muddies the waters. For example, the OSIRIS-ACE and OSIRIS-SAGE results seem to conflict. Does that mean there are biases between ACE and SAGEIII? Is it related to the diurnal effect? I made some suggestions below as to how to investigate the diurnal effect as, especially for OSIRIS, it is not that easy. That said, the lead author is well equipped to examine this further. It is also difficult to know what to make of the comparison that go all the way down an altitude of 10 km. Is there sufficient data for these to be meaningful? And does the variability swamp the signal? All-in-all this part of the paper needs additional analysis such that they would support some clear summary findings.

Once this, along with several smaller clarifications outlined below, is completed I would support publication.

Line 20: you spelled out the acronyms in the abstract, but do so when first used in the main test as well.

Line 27-28: use "… PRATMO stratospheric photochemical …" and " … McLinden et al. (2000) and later Adams et al. (2017)." (https://doi.org/10.5194/acp-17-8063-2017)

Figure 2: what time period is this for?

Line 95: "worse agreement between the OSIRIS measurements and the forward model" Worse how? Shouldn't this be one of the parameters one can optimized by simulating OSIRIS radiances accounting for precision and sampling/field of view considerations. The 2-3 km over most of the stratosphere argument would seem to apply for any of the values of alpha used. Would a smaller alpha help in the UTLS?

Line 110: Be more specific about where v6 and v7.2 begin to differ. E.g. do they both solve for the minimum in equation

Line 125: "The log-normal distributions are less physically realistic" … why is this? This could be true but maybe it is not that obvious. Please give some rationale for this. If this is in the upper troposphere one might expect something non-Gaussian for if there is occasional lightning-NOx.

Line 139 or thereabouts: what is the DOFS, or range in DOFS, for this data product?

Line 144: please motivate the use of 1.5 km as the threshold a little more.

Line 155: "… kernel filter, as used in version 6.2"?

Line 170: "change in the processor"… what does this mean? Same for "global environment settings" later on.

Line 181: would the lower altitude of the ISS, relative to SCISAT, mean there is ~1 more orbit per day… I think there are generally 16 orbits per day for the ISS

Line 192: does OSIRIS not need to worry about O4 absorption in their NO2 retrieval?

Line 203: So this is an off-line (unofficial) retrieval, correct? This should be clarified. Related: later on you use "SAGE III/ISS DV v5.2"… what does this mean? It implies, I think, that the product is v5.2 but with a diurnal correction. If you did your own retrieval then I would not use the 'v5.2' label which implies to me that it is an official product.

Line 221-224: this is good information to have in section 3.2.1

Line 225: "The effect of changes in the input parameters on the PRATMO NO2 was estimated by perturbing them in the model." What was this used for?

Line 228-229: "… it is not always enough to account for the difference between sunrise and sunset occultations from a single instrument." Is the purpose of this statement to point out that systematic errors may still remain suggesting that the scaling is not perfect? My guess is that there could be sampling issues that would also cause the SR and SS to not be the same anyway.

Figure 7: "… coincident NO2, when shifted from their individual local times to 12:00 pm, for several altitudes…" is this what you mean? If not, please clarify.

Line 245: Shouldn't v7 have more data lower down based on Figure 5?

Line 245: showing the standard deviations of the average differences, or the average ± s.d., would suggest if this is due to fewer coincidences or not.

Section 4.1: any discussion of the differences between ACE and OSIRIS below 20 km needs to factor in diurnal effect errors. Neither ACE nor OSIRIS corrects for them, but you should be able to deduce the impact of these based on your work on ISS/SAGE III and other publications on the subject. Previous comparisons, e.g., see Brohede et al. (doi:10.1029/2006JD007586) Figure 9, show the expected high bias in ACE-FTS. The effect in OSIRIS is harder to generalize due to the varying geometry, but was estimated in McLinden et al. (doi:10.1029/2005JD006628) Figures 4 and 5. You might consider limiting OSIRIS to

SZA<85â□° OR

[SZA<85 or (SZA<88 and 75<dAZ<105)]

Section 4.2: What is the impact of the diurnal correction on the SAGEIII profiles? It is a little hard to tease in the figure. What do you make of the OS-S3 comparisons vs the OS-ACE below 20 km?

Line 274: "…from a data set…" what does this mean?

Line 284-285: "At sunrise ACE-FTS samples a region of the atmosphere that has not been illuminated long enough for the NO2 to reach equilibrium, however this is not an issue at sunset." Are you agreeing this is true? I am not sure it is. Or are you merely restated the argument from Sheese et al.? Neither would be in a pseudo-steady-state at the terminator.

Line 291: was the trend removed in these timeseries?

Line 316: "ACE-FTS retrieval it would likely improve agreement with OSIRIS" … would it not make ACE-FTS values larger, thereby increasing the bias? I agree the diurnal effect will be similar for ACE and SAGEIII, and thus, either way you could be more quantitative here.