

Review of Sheese et al. (submitted to AMT)

This is an important validation paper and the authors have done a nice job of putting the results into context using standards set by GCOS, for example.

P1L19: Is “dec” accepted as an abbreviation of decade? The editor can confirm.

P1L23: “being corrected for field-of-view modelling errors” -> “accounting for the finite height of the field-of-view” (my objection is that “errors” is plural)

P3L3: Delete “Since February 2004,” (repetitive)

P3L5: “orbital geometry” -> “orbital geometry and tangent height” (a suggestion since 6 km vertical sampling is very rare in the lower stratosphere for ACE-FTS)

P3L18: “accounted for” -> “taken into account”

P3L25: “210-1025” -> “285-1015” (see 1<sup>st</sup> page of McElroy et al., 2007)

P4L13: use “observations” or “measurements” in this sentence, not both.

P4L30: I think the upper tangent height for OSIRIS is ~70 km in the default “strat” mode, although there are many scans that go to ~100 km in the less used “strat-meso” mode. Has that changed in the later years?

P5L4: “and are retrieved simultaneously” is in disagreement with Degenstein et al. (2009), which states that aerosol and NO<sub>2</sub> are “pre-retrieved”.

P5L13: tuneable -> tunable

P5L29 (and elsewhere): Insert a space between the numeric value and the unit (µm)

P6L4: Don’t capitalize “Personal”

P6L12: No need for a hyphen

P6L29: “of the ACE-FTS comparison results” -> “for the ACE-FTS comparison”

P7L5: “Using 30-day mean...” Why is there a need to deseasonalize the data when daily mean values are used? Spatially, is the drift calculated using relative differences over the globe, or relative differences within a latitude band (i.e., zonally). It becomes clear in section 4.1 and 4.2, but at this point, the reader might be wondering already. If data from both hemispheres are used (i.e. global data), then seasonality largely cancels, does it not?

P7L7: Why is 99% confidence chosen? Does this follow Hubert et al. or some other previous work? If so, please state “following Hubert et al. (2016)”, for example. I think I would be more interested in 95% confidence intervals.

P7L26: examined -> examined,

P8L5: This sentence is a bit odd (revision is suggested but not necessary). If the line shape is improved, then it comes as a surprise that the bias is larger.

P8L9: “which improves the vertical sampling” is not correct. The vertical sampling relates to the measurement, not the modelling.

P8L13: Why not indent at the start of a paragraph? I believe this is conventional for AMT.

P9L12: The authors have provided important information on a drift with MAESTRO PM measurements.

P9L30:  $(2.4 \pm 1.5)\%$  is not different from 1%.

P10L18: I think it is worth mention that ozone has a significant diurnal variation above  $\sim 42$  km and that some differences with other instruments may be due to the temporal coincidence criterion of 6 hours. Also, ACE-FTS may have a retrieval issue due to horizontal inhomogeneity across the day-night terminator at these altitudes (and higher).

P22 (Fig. 4): The confidence bounds are the same colour for v3.6 and v4.1. At some altitudes for some correlative sensors, the confidence bounds are overlapping for the two ACE-FTS versions. I recommend using an outline on the bounds for v4.1 so that they can be distinguished from the v3.6 confidence interval profile.

P26 (Figs. 9-10): Please explain the third shading colour in the caption.... maybe something like “overlapping confidence intervals”

P27: Although no one would consider  $50-60^\circ\text{N}$  to be Arctic and  $50-60^\circ\text{S}$  to be Antarctic, I think it is acceptable here.