Response to anonymous reviewer #2

We thank the reviewer for their very helpful suggestions and their positive review.

1) The abstract states " In clear-sky UTLS conditions (H2O < 10 ppmv), mean differences between ChiWIS and FLASH were only -1.42% and those between FISH and FLASH only -1.47%. Agreement between ChiWIS and FLASH for in-cloud conditions is even tighter, at +0.74%. In general, ChiWIS and FLASH agreed to better than 10% for 92% (87%) of clear-sky (in-cloud) datapoints." I'm a bit confused between the order 1% values noted in the first sentence, and the 10% value noted in the second. Is the second sentence valid for cases where H2O is larger than 10ppmv? Perhaps instead, the uncertainty needs to be included in the first sentence (ie, what's the Standard deviation on the -1.42% for ChiWIS and FLASH?) Please made this clear. Also, include the range in the statement on line 253 on page 10.
→ We received a similar comment from another reviewer about the confusing nature of these two sentences in the abstract. The following comment has been sent to both reviewers:

The goal of the first sentence was to describe the mean differences between instruments across the whole campaign. The goal of the second sentence was to describe the variability of the differences in simultaneous measurements, essentially providing the same information as a standard deviation to the mean values would, but the authors thought this format of 'what fraction of simultaneous measurements agreed better than 10%?' was more intuitive. Furthermore, this way of presenting the statistics aligned nicely with the dotted and dashed lines shown in Fig 3 (previously Fig 2).

However, clearly this phrasing has created more confusion rather than helping. We have chosen to remove this second sentence and instead have added the standard deviations (in addition to mean) for instrument differences. This information is also now presented in Table 3, as suggested by the reviewers. Thank you!

We have also added the standard deviations shown in (now) Fig 5 to the text as suggested. Thank you.

2) Figures in supplement: It would be easier to look at on a laptop screen if you include the values on the horizontal axis for figure s2 and s4
→ X-axis labels for all subplots have been added to Figures S2 and S4. Thank you.

3) Re: MLS, the most recent retrieval is version 5, which at least partially corrected for an instrumental drift. It would be worthwhile redoing the comparisons with the new retrieval.
→ This was commented on by all reviewers and the following reply has been sent to all reviewers.

Thank you for alerting the authors to the newest version of MLS. At the time this analysis was started v5 had not yet been released. We have updated the analysis to use MLS v5. The results are similar, but in the UTLS (~68 hPa), v5 is about 15% drier than v4 across the whole globe (not

shown). In consultation with the MLS team at JPL, there is not a known cause of the larger than expected (from changing sideband fractions) discrepancy at these levels between the two versions, which is about twice as large as reported by Livesey et al. (2021) in section 4: https://doi.org/10.5194/acp-21-15409-2021/. For completeness we have included both v4 and v5 in Fig 9 (previously Fig. 8). Now MLS appears drier than in situ measurements (aircraft and balloon) in both the warm/wet and cold/dry halves of the campaign, but still captures the qualitative shift of drying through time observed by the aircraft instruments across these two periods. We have added a short discussion about the differences between MLS v4 and v5 to the text:

"Here we use 126 water vapor profiles spatially and temporally co-located with the StratoClim flights as a point of comparison (shown in Fig. 1a). We use version 5.0 (v5) profiles which were selected in the region between 20–30°N and 78–92°E during the campaign dates of 27 July – 10 August 2017, using screening criteria from Livesey et al. (2022). We also show MLS version 4.3 (v4) profiles (only 118) which were selected using screening criteria in Livesey et al. (2020). We interpolate the H2O profiles onto a potential temperature grid using the MLS temperature product provided at the same pressure levels. MLS v5 includes a correction on the H2O retrievals described in Livesey et al. (2021), which results in an approximately spatially uniform drying at 68 hPa of about 15%.
…
Overall MLS v4 shows a wet bias compared to v5 (of about 15% between 380-500 K), but both versions are able to discern trend across the campaign of a cooling/drying of the UTLS seen by the aircraft measurements.
…
During the warm/wet period, MLS v5 shows a significant dry bias compared to the aircraft instruments of (-19 ± 7)% and (-22 ± 6)% for ChiWIS and FLASH, respectively.
During the cold/dry period, MLS v5 shows an insignificant dry bias of (-12 ± 12)%, (-11 ± 12)%, and (-5 ± 15)% compared to ChiWIS, FLASH, and the balloon CFH, respectively. Because MLS v4 is 15% wetter compared to v5 in this altitude range, MLS v4 actually agrees more closely with the in situ measurements, reporting no statistically significant differences with any of the instruments during either period of the campaign."