This article collects the results of several in situ H2O/HDO campaigns and compares them to observations from the ACE-FTS solar occultation satellite mission. This provides an opportunity to place the in situ measurements into a broader context using the global coverage of ACE-FTS measurements. It also implicitly serves as a cross-validation of the different in situ measurements, if they can be shown to have similar systematic biases compared to the ACE-FTS results.

The paper is well written and seems reasonably complete. I only have a few comments, the most significant being an underemphasis of evidence of a systematic bias between the data sets. I should point out that this underemphasis appears to be a consequence of feedback that I, myself, provided the lead author in the past.

HDO enhancement in the North American monsoon is an interesting puzzle.

Major comments

> In Table 1, it is not clear how the ACE occultations are selected in order to get the reported numbers of occultations. I assume the selection took all occultations in the latitude and longitude ranges for all years from 2004 to 2022, but only for the same month(s) as the corresponding campaign? It should probably be mentioned in the text.

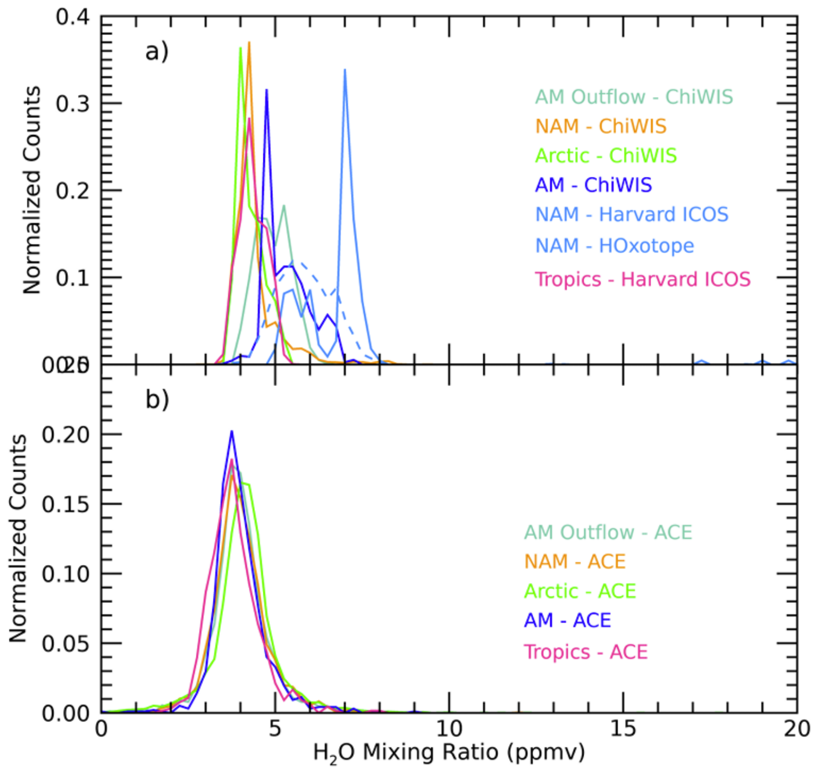
> Figure 4: perhaps include a dotted line in each panel showing an estimate of the tropopause height (from MERRA-2 or some other source). It is a significant consideration for interpreting the results. For example, the results near 16 km in the tropical measurements are in the troposphere, while in the Arctic, the results near 16 km are in the stratosphere. It might also be instructive to show the average altitude corresponding to the 400 K potential temperature level.

> Line 290: We note that the ACE-FTS retrievals are far more uniform than those from in situ campaigns, which likely reflects fine scale structure which is averaged out in the ACE-FTS retrievals.

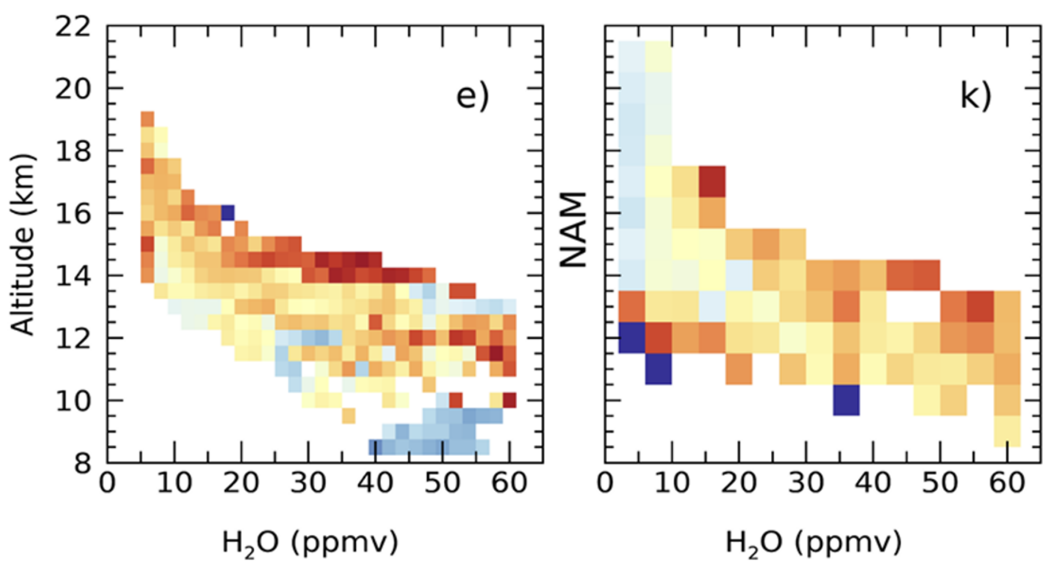
This actually seems to be more of a sampling issue. The two sampling quantities (seconds for the in situ measurements versus individual measurements for ACE) are not entirely compatible. Shown below is an excerpt of Figure 5. Consider the curves for “NAM – Harvard ICOS” and “NAM – HOxotope.” On a side note, I cannot tell which curve corresponds to the dashed line and which one corresponds to the solid line.

I assume these NAM curves are derived from a limited number of flights that targeted relatively moist events. If the plane is essentially “wallowing” in moist air, the distribution will of course skew to higher mixing ratio, but taking each second of measurement as a separate count is effectively counting the same “moist events” multiple times (many, many times). Also shown below is an excerpt from Figure 3, with the ACE results for North American monsoon contained in panel k. Note that at 16-17 km, ACE saw H2O mixing ratios up to 20 ppm, but the fraction of moist events in that altitude region seen was small enough that it doesn’t visually register in the probability distribution in Figure 5. ACE results agree that higher H2O concentrations occur, but not as frequently as the in situ plots might imply.

In summary, the fraction of seconds you measure high H2O mixing ratios when sitting in moist air is not equivalent to the fraction of times you measure high H2O mixing ratios when randomly sampling a particular geographic region with ACE.




Above: excerpt from Figure 5




Above: excerpt from Figure 3

> Line 326: In general, ACE-FTS retrievals of H2O are slightly drier than those of in situ campaigns.

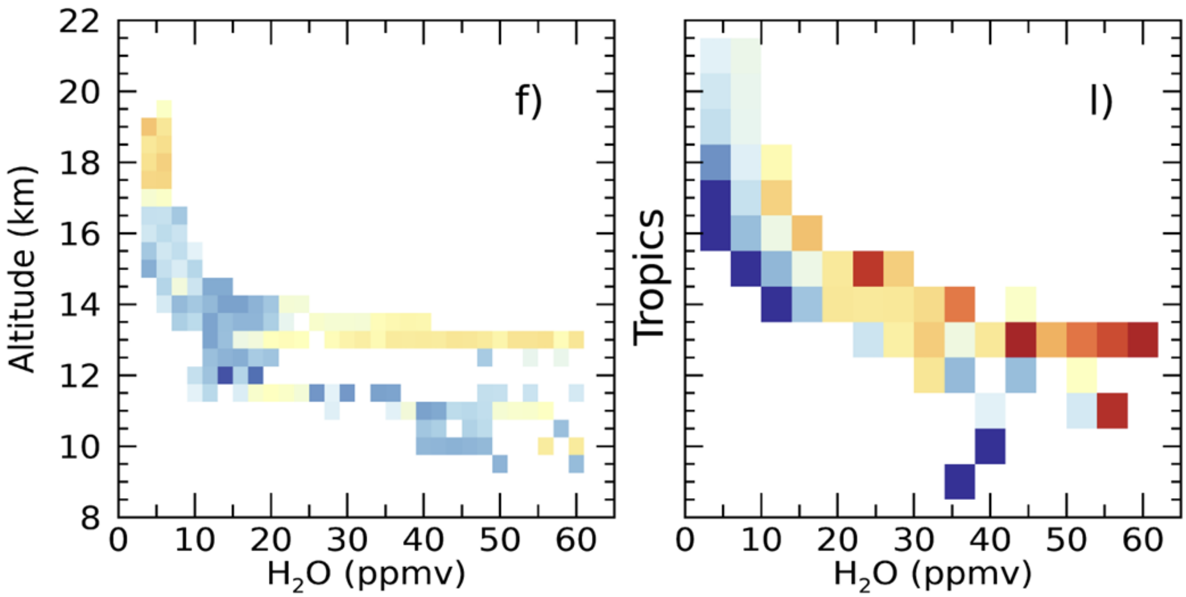
Figure 5 (and the discussion in the text) gives the worrisome impression that the in situ and ACE results are inconsistent, but I believe that is somewhat misleading. Consider the H2O distribution for the Arctic case in Figure 5. This is the only set for which 400 to 500 K potential temperature would be well into the stratosphere for all the measurements. The in situ measurements would presumably be random samples of the stratosphere along the route (rather than targeting a local moist air event), more similar in nature to the ACE measurements. For the Arctic, the probability distributions for ChiWIS and ACE appear to overlap fairly well, suggesting reasonable agreement between the two data sets.

For the other in situ cases, the 400 to 500 K potential temperature range includes a portion of the upper troposphere, and the in situ study presumably targets a most air event, leaving the impression that ACE is usually biased dry, but again I expect that is an artifact of the sampling. If one were performing validation or calibration, the pure stratospheric measurements in the Arctic would be the preferred data set for comparison. There are fewer complicating factors.

HDO, on the other hand, does not overlap particularly well for ACE and the in situ measurements, not even for the Arctic case. This is strong evidence of an inconsistency in the HDO results, but it is difficult to say which side has the problem (or perhaps it is both sides). From the ACE perspective, the likely source of a bias would be errors in the spectroscopy (i.e., line intensities). Because different lines are employed in different altitude regions, this bias could have an altitude dependence.

> Line 333: Even if both the in situ and ACE-FTS measurements are unbiased, it is still possible the instruments might return significantly different isotopic compositions from the same general region due to their very different sampling methodologies.

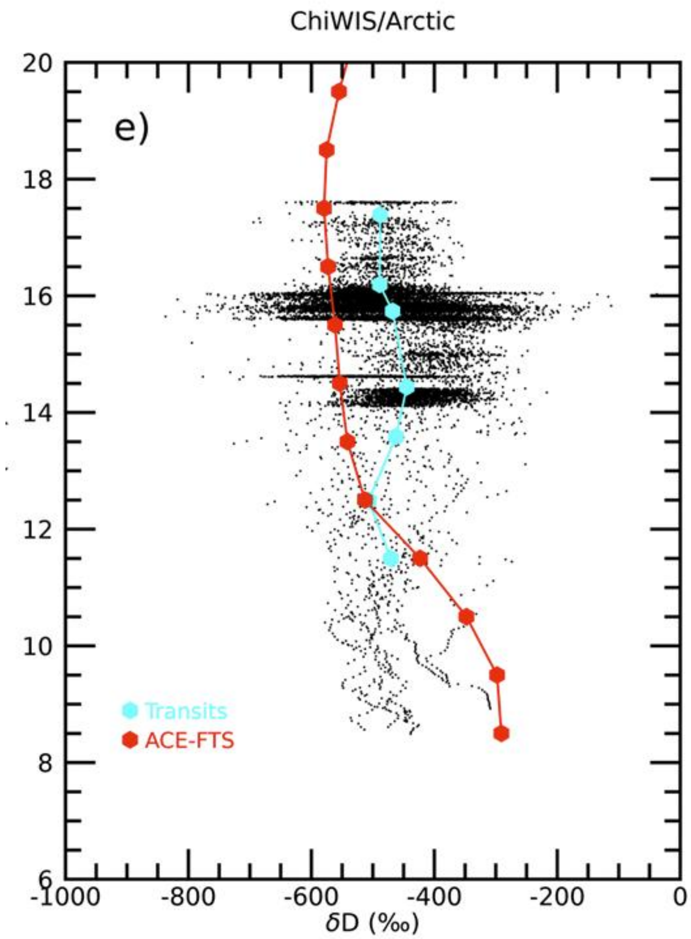
This is the subject I previously discussed with the lead author. There are instances in the comparison set that seem indicative of clouds impacting the results (through enhanced condensation of atmospheric HDO relative to main isotopologue H2O). For example, in the excerpt from Figure 3 below, showing the results from measurements in the tropics, the magnitude of the apparent fractionation in the 15 to 18 km altitude range is very large in the ACE results. In the tropics, this altitude range is in the troposphere, and ACE observes cirrus clouds relatively frequently at these altitudes in the tropics, frequently enough for lower levels of HDO to be seen in the average results. These results might be improved by using the ACE 1 micron imager data to identify occultations containing cirrus clouds and excluding them from the analysis, although I am not suggesting such filtering is necessarily required in revisions to this paper. It is simply worth keeping in mind that the larger discrepancies in the tropics have a logical explanation.




However, there is no similar argument available to explain differences in comparisons in the Arctic set, where 16 to 18 km is in the lower stratosphere (not the troposphere). During the summer (the time frame of the in situ Arctic measurements), there should be no clouds in the stratosphere for latitudes of 40 to 70 °N, so no possibility of them impacting the HDO fractionation. Over the years, there have been periodic volcanic eruptions that provided enhancements in stratospheric sulfate aerosols (liquid droplets of H2O and H2SO4 mixtures), which could provide a means of preferentially reducing atmospheric HDO through condensation, but volcanic eruptions that dramatically enhance stratospheric sulfates are not so frequent that they should pull the average significantly.

Looking at the excerpt of Figure 4 shown below, containing the Arctic results, the ~100 ‰ offset in the stratosphere cannot be explained away by clouds impacting the ACE results. Again, the most reasonable explanation would be that there is a systematic bias between the two instruments, some aspect in the analysis of one or both data sets that induces a systematic error in HDO fractionation, likely in the determination of HDO mixing ratio itself (rather than main isotopologue H2O). In my opinion, the strongest argument for the bias can be made by looking at pure stratospheric measurements in the Arctic set, but we also see similar offsets near 18 km for the other data sets in Figure 4, a persistent offset that strongly suggests issues in one or both data sets. Problems in the HDO line strengths wouldn't shock me, but there is no way to know for certain which data set is the biggest culprit in creating the bias.

Again, it would be appropriate to put stronger emphasis in the text on this systematic difference. Apologies if my previous feedback made the authors hesitant to highlight the discrepancy.




Minor comments

> Line 66: The instrument has been in operation from 2005 through the present day

The ACE-FTS has been in operation since 2004.

> Line 96: HOxotope

This instrument name is sometimes written as “Hoxotope” and other times as “HOxotope.”

> Line 153: high-inclination (75◦) circular orbit

The inclination of the orbit is 74 degrees

> Line 157: The H2O molecule is ideally sampled from 5–150 km altitude

ACE-FTS H2O retrievals extend up to 95 km.

> Line 174: (RINSLAND et al., 1998)

All capital letters

> Index to Table 2 (and Tables 3 and 4): Counts represent seconds of sampling time for in situ instruments and number of occultations for ACE-FTS

There are no counts in these tables.  There were counts in Table 1, so this is clearly a copy-and-paste issue.