Comment on amt-2021-300
Anonymous Referee #2


Overview
--------

The paper presents a comprehensive analysis of available satellite datasets covering water vapour measurements in the upper troposphere, using balloon frost point hygrometers as the main reference source, but also satellite-to-satellite intercomparisons. The analysis is certainly comprehensive and the authors have taken care to explain all the assumptions and caveats associated with their results.

I have only a few significant suggestions (along with a number of minor points) which I think may improve the paper, but I leave it to the authors to decide whether or not to incorporate these.

1) While the problem that in-situ measurements sample just a single location within the volume of the satellite measurement is discussed, it is slightly waved off as an inevitable cause of discrepancy without any attempt to assess the magnitude of H2O variability, apart from a reference to 20-30%
variability from MOZAIC. I think it would have been interesting to see, for the BFH sites chosen, just how consistent the day-to-day values are when averaged over, say, to the vertical resolution presented in the plots. This would give a useful upper limit to the contribution of local atmospheric variability to the discrepancies (assuming over 24 hours that the wind blows air a much larger distance than the satellite resolution) and could, perhaps, replace the rather artificial example used for Fig 2.

2) (specifically the gridded map comparions) Wouldn't you expect MLS to be more 'moist' than AIRS? Presumably MLS is sampling over the whole volume, cloudy or not, while AIRS - even though nadir-viewing - is still restricted to cloud-free, and therefore presumably, drier volumes? On the other hand I would not expect AIRS to be retrieving H2O in the drier regions above clouds either (assuming the presence of a cloud in nadir-viewing flags the whole spectrum as unusable), while MLS would, so perhaps above cloud tops MLS would become drier than AIRS. Perhaps this contributes to the MLS 'exaggeration' of AIRS?

3) The conclusions section is far too long (over 4 sides of text). I would suggest moving much of the material to summary sections at the end of each type of comparison and limit the conclusions to a discussion of Fig 23 and recommendations.

Typos, minor comments & general pedantry
-----------------------------------------

P1 L1: SCIAMACHY is neither occultation nor passive thermal.

P1 L3: Another major problem with this region (less so for the microwave instruments) is the presence of clouds

P1 L4: 'per million BY volume' ?

P2 L16: 'And' should be capitalised, according to sparc-climate.org

P2 L18: 'there HAS been'

P2 L19: Can it be a 'reassessment' if these are 'new' resources which,
presumably, have not yet been assessed?

P2 L25: Table 1 actually has 25 entries under 'Data Set' rather than 21, since you have split all four MIPAS data sets into 2.

P2 L28-30 and L34-38: I'd prefer these lists of data sites in table form, which would be much easier to read and refer back to. Maybe even a map?

P3 Table 1: might be worth distinguishing between nadir and limb-viewing, which are two very different techniques within the 'IR Thermal Emission' category

P4 L41: Perhaps mention here what units are used for H2O concentration (ppmv?) If the original data were in different units, presumably they have all been converted to the same at this stage?

P4 L45: Maybe a colon after 'methods' rather than a comma, and a new paragraph for 'coincident comparisons' since the other two methods merit new paragraphs.

P4 L61: 'sun-synchronous' is hyphenated here, but not in L51.

P4 L63: Section headers seems to have gone a little awry. S4 is titled 'Coincident Comparisons' but includes subsections on the two other comparison methods.

P5 Fig1: Could probably have been incorporated into Table 1.

P6 L71: '20-30%' needs an en-dash in here, ie '--' in LaTeX.

P6 Fig2: not sure this figure is necessary to explain a Gaussian distribution

P6 L81-90: I take the point but I feel there ought to be a mathematically robust method of showing this, otherwise how do we know it's not
just a characteristic of this particular sample?

P8 Fig 3: Although averaging kernels have not been applied to the BFH data it may be instructive to superimpose the corresponding points if, say, triangular averaging kernels of different widths had been applied since this may account for some of the inability of the satellites to track the extreme values (ie accounting for vertical but not horizontal atmospheric variability).

P9 Fig 4: Use 'BFH' rather than 'FP' in plot titles (also for Fig 5). It would be useful to give some indication of the FOV size of the different instruments eg in the figure caption, and how this maps (approximately) on to the vertical scale used here. This would give the reader some idea of the degree of smoothing expected from the averaging kernels.

P9 L120-127: In order to make the Figures 'standalone', most of this should be in the caption to Fig 5 rather than here.

P14 L180: 'also HAVE a seasonal cycle'

P14 L189: An obvious question arising here is: how well do the Vaisala and BFH instruments agree with each other? Presumably that's covered in the Dirksen et al but no reason not to produce your own analysis or reproduce theirs.

Fig 12: A more elegant way of displaying amplitude and phase would be polar plots for each pressure level, amplitude indicated by length of radius. In this way phase differences between low amplitude fits are not exaggerated as they are when plotted separately from amplitude.

P21 L200 (and subsequently in this paragraph) hyphen introduced within RS92, inconsistent with elsewhere (and Vaisala web-site, which has no hyphen)

P21 L203: Unnecessary (and inconsistent) capitalisation of Satellite.

P21 L219: 'based' is superfluous here
As with Fig 5, I think it would be better to have this information in the Figure caption.

Fig14/15 captions should make it clear that scatter plot is only for MLS.

Fig 15 caption: Sodankyla mis-spelt, also missing the umlaut over the final 'a'.

Fig 16 caption: some people might argue that it should be 'data ... are shown', but I'm not one of those people.

Given that ACE-FTS only measures at local sunrise/sunset, while MLS-Aura at local 13:45 (and presumably 01:45), how are there more MLS-ACE coincidences at the equator than MLS-MIPAS?

You seem to have shifted to a different set of pressure surfaces for the gridded maps (ie 50hPa increments). Any reason for the change?

Space within '10ppmv' for consistency.

Again, a new pressure level, 175 hPa.

Perhaps semicolon instead of comma after 'maritime'.

South Atlantic Anomaly should be capitalised (at least Atlantic should be).

Figure caption - also mention the time period used for averaging.

Use LaTeX $-$ rather than hyphens for minus signs.
P34 L379: Comma after 'closing'.

P34 L381 & L393: 'Thermal lapse'? Do you mean negative temperature gradient?

P34 L393: Use en-dash: 10--20 ppmv

P36 L406: Comma after 'troposphere' (to match that after 'instruments')

P36 L431: I’m not sure this is due to forward model differences. In my experience it is most likely due to the inverse models, ie regularisation, a priori assumptions, convergence criteria etc