Review of “ The Airborne Chicago Water Isotope Spectrometer: An Integrated Cavity Output Spectrometer for Measurements of the HDO/H 2 O Isotopic Ratio in the Asian Summer Monsoon” by Clouser et al., https://amt.copernicus.org/preprints/amt-2024-98/amt-2024-98.pdf

The manuscript describes an update of the well-known Chicago Water Isotope Spectrometer ChiWIS for measuring the stable isotope ratio HDO/H2O ratio in water vapor. Due to the extreme technical challenges to precisely measure the H2O isotopologues in the dry mid-/upper troposphere and lower stratosphere, only a very small number of relevant instruments have been developed. The paper thus gives an important contribution and as being well structured and focused, it’s a pleasure to read.

Thank you for these detailed comments. They are highly appreciated and have made the manuscript better. A note on my formatting here. Items accepted and incorporated into the manuscript without comment have been highlighted in green. Other items have been highlighted in yellow, and are accompanied by some explanatory text.

Still I have a couple of issues that need to be clarified (see minor concerns) and one general comment.

General comment:

1) I like the paper very much. It’s very detailed and nicely describes the instrument and its performance during three campaigns. However, one crucial issue is missing, namely the measurement of an isotope standard, e.g. SLAP diluted to low mixing ratios. This is usually provided and some/many PIs just use the measurement of a standard (in regular intervals of minutes to ~1 hour) to verify and demonstrate the quality (uncertainty) of the data. I see that this would be a very difficult or maybe impossible exercise during flights on Geophysica / WB57 (with considerable additional uncertainties), but in the lab such a calibration should be feasible, although not simple with the voluminous cavity. I do not request such an additional lab test, but you should add a section where you discuss this point and provide all the information you gained (e.g. at the AIDA chamber) to enhance the trust in the data and your uncertainty estimates in section 6.2., also as the isotope data you show in Figure 14 are on the very high (isotopically rich) side of the data collected so far.

I have added a subsection in the future work section briefly detailing our plans for an isotopic comparison between ChiWIS-lab, ChiWIS-airborne, and a commercial isotopic instrument. I also note the success of the ChiWIS-lab instrument in extending measurements of the HDO/H2O fractionation factor to 190 K.

Regarding the isotopic data shown in Figure 14 and associated discussion – this section was originally going to involve a more in-depth analysis of the differences between the satellite and our in situ measurements, but quickly became a paper in its own right. A preprint of that paper was recently published on EGUsphere, and I have included here a brief summary and citation to direct the interested reader to that topic. The top-line conclusion of that preprint is that ALL extant UTLS measurements of delta-D (ChiWIS, Hoxotope, Harvard Isotope) are similarly enhanced above both ACE-FTS and ATMOS retrievals of the same.

Minor concerns

- L. 26/27. This increase with altitude in the TTL is not a general feature, but (likely) indicates the recent influence by convectively lofted air masses. Please correct.

- L. 29. Add citation.
- L. 41. Add after table title "(references further below)".
- Move the sentence "Limited in situ isotopic measurements of water vapor isotopologues … and briefly described below." above the table.

I will make a note to the publishers during the typesetting stage to ensure that the table is properly positioned relative to this text.

- After L. 53, add something like "commercially available in situ laser isotope instruments (e.g. by Picarro or Los Gatos) are not adequate for H2O mixing ratios of below ~1000 ppm".
- L. 64. … and the inadequate/poor vertical resolution of ACE-FTS (although for a satellite instrument it's quite good)
- L. 91. You state that "OA-ICOS is necessary to satisfy this requirement …" so that only OA-ICOS can do the job? The OF-CEAS technique invented by Danielle Romanini could do it, too, but was (to my knowledge) never been applied in the mid-infrared. And there may be other potential techniques.

You are correct that OA-ICOS is too specific. It is more proper to say that Cavity Enhanced Spectroscopy (CES) of some flavor is necessary to observe HDO features in the context of an aircraft-based instrument. I have incorporated text to this effect in the manuscript.

- L. 118. As MIPAS will not be known by all, use quotation marks … "MIPAS dome".
- L. 127ff. The title list of the subsections A – F does partially not agree with Figure 4 and the subsections further below. Please make them (more) equal.
- L. 132 or at L. 155. Specify the lasers (manufacturer, type).

I have clarified that the JPL MicroDevices Laboratory both manufactured and provided the DFB tunable diode lasers.

- Figure 4 (or somewhere else). Give dimensions and weight of ChiWIS … and it would even better to have an instrument spec sheet (table) with most relevant parameters, i.e., also measurement range, response time, measurement speed, …

We have included a table in the text which summarizes the most critical characteristics of ChiWIS

- L. 175. Section ??
- Figure 5. Give dimensions.
- L. 188. To avoid confusion with Figure 5, I would use M-in and M-out (or so) for the two mirrors, instead of M1 and M2.
- L. 190 (or around L. 200). "tiny fraction". Do you know how much of (1 - 0.9998)*100% ? Or what is the transmission T ?
- L. 217. Here you write that the surface roughness by Optimax is ~0.3 nm, but on L. 210 you write that such a roughness already causes a decrease in the path length from 7.4 to 2.8 km. And already in the abstract you wrote that the path length if 7.5 km. I don't get it here.

This is a critical point about the instrument, and I have tried to rewrite this section for clarity. This passage is meant to provide an illustrative calculation highlighting the importance of low mirror roughness. In this calculation I show what the mirror losses *would* be if the mirror substrates had a roughness of 3 nm instead of 0.3 nm. I hope this passage serves as warning to future developers of ICOS instruments in the NIR or MIR to avoid substrates with large grain sizes.

- Table 3. Explain the parameters / abbreviations.
- L. 235ff. If I interpret this para correctly, the "set of eight identical wedges" is identical with the nonax component in Figure 6. First, use an identical name / specification and secondly, shortly explain how this component increases the signal-to-noise ratio by a factor of 3 (by an improved focusing onto the detector?) and by how much you have

increased the number of photons on the detector?

- L. 241ff. The sentence "Even though better noise characteristics could be had from a smaller, colder detector, calculations indicated …" I don't fully understand. And the consequence? Next time you would use a larger detector?

I have rewritten this section for clarity to indicate that ray-tracing calculations indicated that it was more important to the overall signal-to-noise ratio to maximize signal by using a detector with a large area than it was to minimize noise with a small detector, which could also be cooled to lower temperatures.

- L. 333. "… can be changed prior flight through …" or it's during flight via access from the ground?
- L. 394. "… that valve angle is incrementally stepped up to 90 degrees over the course of several minutes" … in order to allow the top valve to independently control the cell pressure? Please, add a subclause.
- L. 396. Exchange "very wet" with "humid" or so. Above ~400 hPa all airmasses are quite dry.
- L. 416. What means: "Spectra … are rebinned in software"? It's apparently not a simple "binning" (including how many individual data points?)!? Shortly specify or refer to the relevant section.
- L. 437. Shortly comment, whether there was a power reduction of the pump due to overheat.
- L. 450. Exchange "less than two seconds even at the lowest concentrations" with "of 0.? s (at x ppm) and 2 s (at 3 ppm)" or so.
- L. 460. This means that the contamination occurs between the campaigns and in the lab? Please specify and explain.
- L. 472/473. Where the procedure and equation come from? Add citation or whatever.
- L. 487. Here you write 15 ppm, in the figure caption 18 ppm and in L. 494 18.5 ppm.

I understand the source of confusion in this passage, but each of these numbers refers to a different situation and therefore can't really be harmonized or matched up.

- L. 490-495. Give the ranges where you average over 2s, 5 s, and 10s. Cold/depleted is quite subjective.

There are not strict mixing ratio ranges over which I adhere to these averages – it is really more of a reflection of how much time one has to stop and restart the fitting routine to deal with problematic scans. Nevertheless, I have included approximate ranges in this section.

- L. 496. "raw data (gray)" … I see black.
- Section 6.1 (Precision). In this section, some things I don't understand.
  - L. 512. "The measurement precision σ χ is defined as the standard deviation of the mixing ratio as a function of time". What means "as a function of time"? It's the standard deviation e.g. in a 60 s interval or so? … and includes "atmospheric" variability (not "natural"). Please specify.
  - L. 516. You write "for all field data sets the H2O measurement precision is significantly larger than the detection limit", but in L. 520 you write "measurement precision of 3.6 ppbv. This is roughly half the detection limit" which likely shows that during flight atmospheric variability dominates, right?
  - Assess how the instrumental variability in the lab compares with the one at flight and then you could/can differentiate between instrumental and atmospheric variability.
  - You should better explain all this based on the numbers shown in Table 4, e.g. "see line/column x/y in Table 4".
- Table 4. Add in the caption: "χ is the H2O or HDO mixing ratio". The unit in column 4 (χ) is

wrong.

- L. 543. Exchange “natural” to “atmospheric” everywhere is the text and again, you do not well distinguish between instrumental and atmospheric noise. Fig. 8b indicates that ChiWIS basically works during flight as in the lab, because the atmospheric variability of HDO is much smaller than the instrumental noise. You can thus argue that the much higher Allan variation in H2O (blue trace) compared to the lab is atmospheric variability … and the “removal” of the fringe at longer integration times you have described … I would stress the great performance in the field, that is, the “external Piezo” (aircraft, vibration,…) that makes your instrument even better in the field than in the lab.

I have rewritten this section to clarify these important points.

- L. 558. I guess, the difference is mainly due to the general higher noise level or lower signal, respectively, which let us see the fringe at higher averaging times and lower Allan deviation, resp.
- Sect. 6.1.2. Usually one compares the measured noise (e.g. when the laser is off) with the theoretical noise sources, such as shot noise and Nyquist noise which can be calculated easily (in contrast to the 1/f-noise). Can you add this, please?
- … and then you have nicely separated the different noise levels: electrical noise (is apparently constant or only depends on temperature), optical noise (alignment dependent) and atmospheric variability. It would be cool to see this noise partitioning somewhere at the end in the conclusion or presented in a table or even with additional columns in Table 5 (it’s up to you).
- Section 6.2. All/most the uncertainties discussed and shown in Table 6 cannot be assessed by the reader. Thus, write at the beginning, how the uncertainties have been retrieved. For instance, some can simply be measured such as the cavity length uncertainty, others are given in the literature (e.g. Hitran), others come from retrieval algorithms and some may be estimated.
- L. 619. Again “very wet” in the atmosphere means everything above 25000 ppm or so.

I have clarified in the text that this refers to intervals with greater than 1000 ppmv H2O.

- Section 6.3. Here you don’t discuss the cavity flush time (it can simply be calculated based on the pump volume flow … add this), but the cavity or better instrument response time, which is also dependent on memory effects of the sampling lines and the cavity. Please correct!

The available data on the Triscroll 300s volumetric flow rate at a given sample pressure are available, but are all taken in ambient air. That is, the scroll pump exhausts gas into air with a pressure of 1 atmosphere. I do not believe the manufacturer’s data can do more than provide a lower bound on the volumetric flow rate, since presumably the impedance on the flow through the pump is lower when the ambient pressure is lower. This was part of the motivation for performing the modified Allan deviation study above.

I have changed the section to “Instrument Response Time” and defined that this is inclusive of cavity flush time and memory effects.

- Fig. 9. Give the temporal resolution of the shown data. The ChiWiS data appears smoothed and with an averaging time of >2s and I actually wonder (having the short cavity flush time and high precision in mind) why you don’t provide the H2O data with ~2 Hz (the cavity flush time / instrument response time). It appears at least, that the “data averaging” smears out fine atmospheric structures.

The work of Singer et al used 1 Hz H2O data from the ChiWIS instrument. I have included this in the text.

- Fig. 11. Write what low, medium, high N2O m.r. mean in ppb/ppm/%
- L. 688. Write where the bandpass filter has been placed. Refer to L. 175 and to Fig. 5.

- Fig. 12. Give the fit results (red line) in the graphs.
- Fig. 12 caption. Bottom panel. Add “data without correction” or so.
- Fig. 13 caption. After “… because contamination”, add “(likely from memory water at the cavity wall)” or so
- L. 737. Exchange “real” with “atmospheric”
- Fig. 15 caption. Delete “natural”.
- Fig. 14/15. Understand it correctly, that in Fig. 14 only the stratospheric data (i.e. above 400 k) are shown, but in Fig. 15 that are very close at the troposphere and at theta-levels below 400 K. Please detail this a bit. I also ask as at 17.5 km you should have seen pot. Temperatures close to 400 K.
- Table 2. Add theta (K), see my question before.
- Fig. 14. I know it’s an AMT paper, but clear is that the dD-values a high compared to most other data taken at these potential temperatures (altitudes). Thus, at least a short discussion is required. Do you e.g. believe that your data / profiles are already representative and your instrument is so accurate that you can claim that the picture has to be revisited a bit, that is that water may enter the stratosphere more heavy than previously thought … as we have the most accurate device that has been existed on this planet …

I had originally intended for this section to include a comparison of our in situ measurements with retrievals from the ACE-FTS satellite, but that analysis quickly grew into a paper in its own right. That manuscript is nearly complete and ready for submission to AMT. In it I have also included delta-D measurements from Hoxotope and the Harvard water isotope instrument, and found that in the stratosphere they cluster around the same range of delta-D values as ChiWIS, and that the in situ instruments are 100-150 per mil enriched above ACE retrievals in every campaign region. Obviously either or both of the in situ or satellite measurements could be biased for a variety of reasons, which are described in the forthcoming manuscript.

- L. 765. Do you think of silcosteel, which can be used for aluminum? Please write.
- L. 769. Instead of “clumped”, is “multiply substituted” or “doubly substituted” the better word?

I have made sure that both clumped and doubly-substituted are included in the manuscript, although I prefer clumped. The reason is that the singly-substituded water isotopologues HDO and $H_2{}^{18}O$ are also used in the so-called ‘doubly-labelled water’ method used by doctors and biologists to study animal metabolism.

- Fig. 16. If you can measure even at 3500 ppm, the cavity is apparently not optically thick or lines saturated and the instrument could potentially be used in almost the entire free atmosphere, right? … see my hint to the spec table at the beginning at, to L. 785 where you write that it can be operated only above 12 km and to the entire conclusion.

While it is true that the cavity is not optically thick, even at high mixing ratios, the spectra become much more difficult to interpret above about 500 ppm. This is primarily due to the nonlinear response of the spectral features as they become saturated at line center.

The limitation to operation above 12 km only is not due optical saturation, but primarily due to concerns about condensation due to rapid expansion of moist sample gas as it traverses the top valve where there is a large pressure differential. Lab tests suggest that condensation at this point can reach the mirrors where it rapidly makes them dirty and poorly reflective. Ground based operation related to Fig. 16 involves a different inlet setup and lower flow rates, and is not as susceptible to this effect.

I have made note of this limitation in the System Design section.

- Conclusion. I haven’t checked all other relevant publications, but your instrument is likely the most precise and maybe most accurate one worldwide, no? If so, write or at least add a sentence that outlines how ChiWIS performs compared to others.