This paper describes an advanced version of the Chicago Water Isotope Spectrometer (ChiWIS), an instrument specifically designed for measuring vapor-phase water isotopologues (different molecular forms of water) in the dry upper troposphere and lower stratosphere (UTLS). This upgraded version employs a tunable diode laser (TDL) and off-axis integrated cavity output spectroscopy (OA-ICOS) for precise measurements.

ChiWIS was used in several airborne research campaigns, including the 2017 StratoClim campaign during the Asian Summer Monsoon and the 2021-2022 ACCLIP campaigns aboard different research aircraft. The instrument measures the HDO/H2O ratio, scanning absorption lines at a wavelength near 2.647 µm. It achieves high accuracy with a path length of 7.5 km under optimal conditions.

Key design features include a novel optical component that boosts the signal-to-noise ratio by threefold and ultra-polished cavity mirrors that minimize scattering losses and optical fringing. In lab tests, the instrument demonstrated high precision, achieving a 5-second measurement accuracy of 3.6 ppbv for H2O and 82 pptv for HDO.

The paper highlights the instrument's advancements in airborne isotope measurement technology, emphasizing its successful deployment and precision in capturing isotopic data in challenging atmospheric conditions.

The paper is precise and exhaustive in its description of the new version of the Chicago Water Isotope Spectrometer (ChiWIS). The authors provide a thorough explanation of the instrument's design, including its advanced optical components and the successful application in multiple airborne research campaigns. The level of detail in the methodology, as well as the demonstration of the instrument's precision in laboratory settings, showcases the robustness of their work.

The article is highly informative and well-written, making it a valuable reference for the future; therefore, it merits publication.

However, before publication, I recommend a minor revision to enhance clarity in a few sections where technical details could be further simplified or clarified for a broader audience.

In the following, (row).

(60-61) You claim that a sensitivity of 50 per mil is required to resolve a convective streamer. Could you justify further this assumption, maybe quoting previous research?

(62-63) Again, could you add some justification to this?

(Table 2) Please give a description of the parameter in the header of the table, both in the caption and in the text

(87) Spend a few word to note the power of the baseline increases, also in view of introducing the pedestal correction section later on

(96) maybe here you can add "under the hypothesis of negligible one-pass intracavity absorption and add some reference.

(Figure 1) Start the Y axis from 0, explain (in the text, not in the caption) why the power of the baseline increases, describe what is meant by n samples or transform the X axis into cm-1 (preferable). In the caption you use percentages for the line depths, please explain percent with respect to what. Change “dry air” to “WV depleted air” or something like that. Keep in mind that among your readers there might be those unaccustomed to how the mode coupling between lasers and cavities works. So, when writing about “mode hop” add that it is an instrumental artifact due to large shifts that happen when the laser switches from one longitudinal mode to another (if that is so).

(106) You should expand this and try to be more didactical toward a broad audience. I would suggest: “In OA-ICOS, the cavity supports certain resonant modes, and noise can arise from fluctuations in the coupling between the laser and these modes (mode noise). High laser tuning rates can reduce the impact of mode noise by quickly sweeping through the cavity modes, preventing prolonged interaction with any single mode. This rapid tuning reduces the likelihood that fluctuations in the laser's frequency or intensity will coincide with the cavity resonances, thus minimizing mode noise. On the other hand when the laser tuning rate is high, the laser frequency changes quickly relative to the cavity's ring-down time, which in our case is ... and the cavity may not fully respond to the rapid changes in laser frequency, leading to a situation where the cavity doesn't have enough time to build up the intensity that would normally correspond to a sharp absorption feature. This causes the absorption signals to be smeared out over a broader frequency range, making them appear broader and shallower than they would be at lower tuning rates.” Play with it.

(110) Also here you should explain why larger mirrors are preferable: Suppress optical noise by enhancing the stability of the optical cavity, increasing the effective path length, reducing sensitivity to misalignment and mode noise, improving light trapping.

(116) No need to be so specific, I do not think there would be many of your readers who know where Bay IX is “... to fit in a bay of the aircraft where...”

(119) Those who do not know the way the instrumentation is housed in the Geophysica aircraft is unable to appreciate this information, and maybe do not even know what MIPAS is. Add references and make this phrase more general.

(147) Please be more didactic for the unaccustomed reader. As instance, explain what is a ringdown and why it is useful to measure, so to introduce here what you will detail later on. Something like "The ringdown trigger is sent just before the end of the ramp to initiate a ringdown scan. This scan focuses on the decay of light intensity inside the cavity, which is known as the ringdown event, and measures the cavity's ringdown time, which provides information on the losses within the cavity, including absorption by the sample." Is the laser turned off or detuned during such events?

(178-180) If I understand well, D does not measure the instant power of the laser, but the interference between the instantaneous and a delayed emission (this in turn modulated by the etalon). This latter, in principle, could have a different intensity if fluctuations are fast. It would be interesting to mention the relative contributions of these two members to the interference figure, given that the smaller the contribution of delayed emission, the lower the interference modulation on D, and therefore the measured data can at best quantify both the instantaneous power (as a dominant contribution to the revealed light) and the processing of the laser tuning with the etalon modulation. Moreover, I guess the delayed emission has a slightly different frequency from the instantaneous one, as the tuning proceeds, If this is the case, therefore it should cause beats of the interference figure on D. Which of the two effects is dominant in the interference figure, the etalon modulation or the frequency shift? Please clarify.

(251) “TRB”, here and everywhere expand the achronyms at their first appearance.

(254) 40 hPa. Why this particular value was chosen?

(257) Those who are not accustomed to the different instruments' housings of the Geophysica, cannot appreciate this information. Please reformulate in a more descriptive way.

The rest of the instrument’s description is very clear and well written.

(421) FLASH and DLH, add brief description of the instruments and references

(465-477) This paragraph is not very clear. The authors should explain more clearly how on the reference detector is present both a fraction of the laser output to monitor the laser changing power, and the alternating maxima and minima that arise from the interference within the etalon (the “fringing”). As the fringing pattern is modeled with a squared cosine, I am assuming the etalon quality factor is not very high, but it would be worthwhile to mention it. Moreover, it would be beneficial to add some reference, or in-depth justification of the choice of the f(s) function in the cosine argument. I guess that, since both wavelength shift and laser power change vs current are temperature dependent, the fitting to f(s) is made on every laser ramp. This is worthwhile mentioning explicitly.

(534) The optical fringe is introduced well, but a quick clarification of how this affects the measurement process (e.g., "introduces a sinusoidal error pattern") might help readers unfamiliar with fringe effects.

(546) The description of how vibrations reduce the fringe effect is insightful but could benefit from a bit more detail on how this trade-off affects overall data quality (i.e., is vibration a bigger or smaller problem than the fringe?).

(594) The discussion on tuning curve uncertainty is solid, though perhaps expanding on what "residuals from the tuning curve fits" are and how they impact measurements could help those unfamiliar with these technical specifics.

(636-651) The comparison between ChiWIS and DLH using the Allan deviation plot is well-conceived. It effectively highlights the difference in instrument performance and measurement sensitivity. Specifically, ChiWIS shows a characteristic minimum deviation at 0.5 seconds, while DLH never reaches a minimum deviation, suggesting that DLH is capable of capturing atmospheric variability down to extremely short timescales.

The fact that the DLH instrument measures natural variability more effectively than ChiWIS on short timescales is a key point. The DLH's ability to measure at integration times as short as 0.05 seconds provides a useful benchmark, illustrating how different instruments can be optimized for different applications.

It would be helpful to include a qualitative explanation of why DLH outperforms ChiWIS on short timescales. For instance, ChiWIS's reliance on a cavity-based measurement inherently limits its ability to track rapid fluctuations due to the time needed for air replacement within the cavity. In contrast, DLH’s direct measurement approach (free air sampling) naturally allows for faster response times.

The inclusion of Figure 10 in the text is a good reference, but a short description of the key features of the figure would aid in interpretation. For instance, explaining the downward trend followed by the rise in deviation after 0.5 seconds for ChiWIS, compared to the continuous downward trend for DLH, would provide additional clarity.

Please declare somewhere in the text that “T” stands for “Integration Time”

(753) Are there any limitations or trade-offs associated with using the filter in terms of data quality or instrument sensitivity? Additionally, elaborating on how this validation process compares ChiWIS data with other instruments or campaigns (such as ACCLIP) could emphasize the reliability of the correction.

(777-801) This section presents a comprehensive overview of the field data collected during the StratoClim and ACCLIP campaigns, focusing on the performance of the ChiWIS instrument in capturing isotopic field data. The narrative is well-organized, with clear references to figures and tables that support the findings. However, to improve clarity, consider providing a brief explanation of the scientific goals or significance of these campaigns, especially for readers who may not be familiar with them. This context can help underscore the importance and significance of the data being presented.

(781) The statement that ChiWIS returned science-quality data from 6 of 8 flights in StratoClim and 3 of 4 test flights in ACCLIP is informative. However, it would be beneficial to include a brief discussion on the criteria for determining data quality. Or the instrument simlpy did not work?

(Figure 16) This effectively illustrates natural isotopic variations during Flight 2 of the StratoClim campaign. The explanation of how distinct isotopic ratios correspond to different origins is informative but lacks context about what caused such differences in isotopic ratios and why these differences are significant for atmospheric science. What do these differences tell us about the processes occurring in the atmosphere? When discussing the isotopic measurements as indicators of airmass origins and ages, it might be useful to elaborate on the implications of these findings. The reference to Bucci et al. (2020) provides important context and may be expanded.