

## **Anonymous Referee #1**

Brunamonti et al. report a synthesis of results from the AquaVIT-4 project, providing a comparison of state-of-the-art atmospheric hygrometers for use in conditions found in the upper troposphere and lower stratosphere. Accuracy of hygrometers in these dry conditions is fundamental to assessment of ice microphysics, which controls the formation of cirrus and limits the transport of water vapor into the stratosphere. Substantial differences in the measurements reported in this region by various hygrometers has historically motivated a continued set of campaigns at AIDA and in the atmosphere, to assess the skill of research hygrometers.

This article is excellent. The experiment and analysis are well done, the figures are high quality and text is very well written. I essentially only have minor editorial comments for the authors to address, in addition to a couple of potential small changes that they could consider for the content and analysis.

**Authors:** We would like to thank the Referee for the constructive feedback that helped us to improve our manuscript. Below are the individual comments from the Referee (in black) and the replies from the Authors (in blue). Please note that page and line numbers given below refer to the revised manuscript without tracked-changes.

**Page7 Line 5:** Suggest to mention what kind of data acquisition mode is used for ALBATROSS, e.g. scanning with DC or 2f modulation, etc?

**Authors:** An additional paragraph was added to describe the method in more detail: *"ALBATROSS uses rapid spectral sweeping of the QCL by periodic modulation of the laser driving current. A highly energy-efficient strategy, referred to as "intermittent continuous-wave" (iCW) modulation (Fischer et al., 2014), is implemented, in which the driving current is applied in pulses, typically 200  $\mu$ s long, followed by a short period of complete shutdown of the laser. The transmission data, consisting of  $25 \times 10^3$  data points, are digitized by a 14-bit analogue-digital converter (ADC) at 125 MSs-1 and real-time processed by an FPGA (STEMlab 125-14, Red Pitaya). The signal-to-noise-ratio is further improved by averaging up to 3000 individual spectra in real time, leading to an effective measurement rate of 1 Hz."* (page 7, line 9).

**Page 8 Line 6:** Suggest instead of saying "25+ years" state the first year that it was operated.

**Authors:** Changed as requested ("*since 1994*").

**Page 9 lines 18 – 27:** Is this paragraph really needed? As I understand it, this is summarizing results that are already reported in Ghysels et al 2024. The large differences stated of "+/- 23.6%" and "about 30%" I assume are differences in the air that was measured by the instruments on those experiments and not due to accuracy problems with the instruments, but this isn't fully explained here. I would either provide more detail or remove that part.

**Authors:** Indeed, this paragraph summarizes the results of Ghysels et al. (2024). We agree with the Referee and hence removed the paragraph.

**Page 13 Line 15:** need to delete word "of"

**Authors:** Done.

**Figure 2:** Recommend using a discrete colorbar with ~5 degree intervals.

**Authors:** Implemented as suggested.

**Page 14 Line 7:** "instrument" -> "instruments"

**Authors:** Done.

**Page 26 Line 8:** "Evaporates" should be changed to "sublimates". However, I'm having trouble understanding that this is really the mechanism that results in a loss of ice. Is it rather that the ice crystals sediment to the bottom of the chamber? It doesn't make sense that the ice cloud would sublime while it is supersaturated.

**Authors:** The cloud loses ice volume due to warming from the chamber walls, leading to sublimation. Sedimentation can contribute as well depending on the ice particle size. "Evaporates" was changed to "sublimates" and a short explanation was added: *"the total H<sub>2</sub>O decreases with time as the cloud evaporates sublimates due to warming from the chamber walls (sedimentation may also contribute, depending on ice particle size)"* (page 28, line 19).

**Page 26:** Does the result in Figure 9 imply a positive bias in APicT and is it worth making a comment about how that might impact the comparisons shown previously where that hygrometer was the reference? Or is it rather the case that the uncertainty range shown for 100% RH<sub>i</sub> in Figure 9 is considered uniform such that the correct value could be anywhere in that +/-5% range with equal probability?

**Authors:** The deviation of +5 % between APicT and the ice saturated mixing ratio (H<sub>2</sub>O<sub>IceSat</sub>) is indeed at the upper edge of the APicT uncertainty range ( $\pm 5$  %), and might potentially indicate a positive bias of this instrument with respect to the true value. However, it should be considered that also H<sub>2</sub>O<sub>IceSat</sub> has a comparable uncertainty ( $\pm 4.6$  %, green shading in former Fig. 9/now Fig. 10), due to the uncertainties on AIDA temperature and pressure. These uncertainties are derived from error budget calculations and account for both systematic and random effects. The measurement precision (i.e., random error component) is in both cases much smaller than the total uncertainty: ~0.1% for both APicT (see new Fig. 9, Section 5.2) and H<sub>2</sub>O<sub>IceSat</sub> (estimated from the timeseries shown in former Fig. 9/now Fig. 10). Therefore, in both cases, the systematic error component dominates the uncertainty. This implies that, in the absence of an independent reference with a better uncertainty, the positive bias of APicT cannot be excluded, but also cannot be estimated based on this measurement.

The text was revised to point out that the observed discrepancy between APicT and H<sub>2</sub>O<sub>IceSat</sub> is at the upper edge of the APicT uncertainty range (page 28, line 25).

One other comment I have is on the lack of substantial discussion or analysis about instrument precision. Precision is discussed a bit towards the end of the manuscript, but is not quantitatively summarized e.g. in Table 1 or assessed elsewhere. This paper would be a useful venue for comparison of the precision of the hygrometers as well as accuracy so I suggest that the authors consider addressing this somewhere. One suggestion if the authors do this is that some markers indicating the observed precision could be added in e.g. Figure 8, or a separate figure could be generated comparing the precision of the different instruments in this relevant P, H<sub>2</sub>O space.

**Authors:** We thank the Referee for this comment. We agree that a detailed assessment of the instrumental precision was lacking in the previous version, therefore we added a new dedicated Section ("5.2 Precision") and figure (Fig. 9) to the manuscript to address this aspect.

The corresponding text and figure can be found in the revised version of the manuscript (page 25, line 6 to page 27, line 7). In summary, the precision of the instruments was evaluated by a detailed analysis of the timeseries of a setpoint interval representing average tropopause conditions (~6 ppm H<sub>2</sub>O and ~100 hPa pressure). First, a linear fit was applied to all timeseries to define the time evolution of the H<sub>2</sub>O mixing ratio, allowing to distinguish instrumental variability from small changes in H<sub>2</sub>O occurring inside the AIDA chamber and/or sampling lines during the measurement interval. For the extractive instruments (MBW373LX and ALBATROSS), an additional sinusoidal fit was also applied to account for the small-scale fluctuations in H<sub>2</sub>O associated with the heating controller of the sampling line (discussed in Section 5.1). Then, the precision of each instrument was quantified by calculating the frequency of occurrence distributions and standard deviation at 2 s resolution ( $\sigma_{2s}$ ) of the detrended timeseries over the entire (10 min) setpoint interval.

We believe this addition provides a very valuable insight to the paper, therefore we are grateful to the Referee for this suggestion. A short paragraph summarizing the results of the precision analysis was also added to the Conclusions section (page 30, lines 22-26).

Note that following this change, Section 5.1 was renamed (from "Statistical intercomparison" to "Accuracy"), and a short sentence was added to introduce the structure of Section 5.