Referee #3 <u>(Daniele Romanini)</u>

Very nice work and paper, clearly exposed and overall rigorous in contents and discussions.

Your kind words are very much appreciated. Thank you.

First of all, I fully agree with all comments, questions and suggestions of the other two referees.

In particular, as already underlined, this works illustrates and confirms the interest of using laser absorption spectroscopy in particular in the mid infrared, to obtain high accuracy and selective measurements of atmospheric species. The presented instrument additionally possesses the uncommon feature of an open absorption cell which allows fast renewal of the atmospheric sample with minimal surface effects, especially valuable with sticky molecules such as water. This advantage is clear when looking at the long transient times (hours) observed when closing the cavity as is done for this study. About this point, it would be nice that the authors add a short discussion about the fact that the performance of the instrument can be considered the same when the cavity is open to ambient air. In particular, are there any turbulence effects which add noise to the measurements, or thermo-mechanical deformation of the optical setup which may increase measurement drift relative to what obtained here?

This is a key remark and a decisive aspect, and we are thankful to the Reviewer for raising this point. At the beginning, we asked ourselves exactly the same question and we spent quite some time in trying to establish a fact based verification of it. A detailed description of our finding was summarized in a recent publication (Tuzson *et al*., AMT, 13, 2020) demonstrating the performance of the QCLAS aboard drones. The environmental conditions (turbulence, temperature and pressure fluctuations, and mechanical vibrations) in such configuration are much more severe than flying aboard a weather balloon. We found that the highly reduced optical complexity of the QCLAS demonstrates an excellent robustness against environmental impacts and its performance was not deteriorated. The engineering solutions that allow to achieve such characteristics are detailed in the PhD thesis of Manuel Graf (see https://doi.org/10.3929/ethz-b-000429788). We added a short paragraph to Section 4.3 of the manuscript to summarize this aspect:

"Although, these assessments were performed under well controlled laboratory environment, we expect that the performance of the instrument will remain the same also during flight conditions. The demonstration of this behavior is beyond the scope of this study, however, some related aspects (the effects of turbulence, thermo-mechanical deformation, etc.) were described in our earlier studies, e.g., Tuzson et al. (2020), Graf (2020), Graf et al. (2021). We found that the highly reduced optical complexity of ALBATROSS exhibits an excellent robustness against environmental impacts and its performance was not deteriorated."

This work also illustrates that the Voigt line profile is largely inadequate for an accurate description of collisionally broadened molecular absorption lines. This is actually a well-known fact since the time that high-resolution spectra are being obtained by using narrow laser spectral sources - compared to spectra obtained at lower resolution with traditional Fourier or grating spectrometers. Nonetheless, it is very instructive to see the impact of the choice of spectral line shape on the retrieval of molecular mixing ratios, and to see that linear and accurate results may be obtained by using an advanced line shape model - with parameters determined by a multiline fit at several pressures, as shown in this work.

I have only one criticism concerning the long-term stability. The Allan-Werle stability analysis was performed for timescales only up to 500s (8 minutes). However, the importance of AW stability plots goes beyond the determination of the time for which a minimum AW deviation is attained, which defines the time for optimal averaging used in this work. The behavior of AW deviation at longer times provides essential information on instrumental drift. Do the measurements keep falling close to the optimal value or else do they drift away, and by how much and over how long? The AW deviation at long time scales provides this information on long-term stability of measurements and allows to assess the need for a re-calibration in case measurements are required to a level below the long-term drift. As it may be complicate to run an AW plot over more than one day, one can replace that by taking individual measurements averaged over the optimal time periodically and during a period which may be representative of the duration of a measurement campaign (one or 2 weeks).

We agree that longer measurement time would allow a more detailed insight about drift mechanisms acting at various time scales. However, in case of water vapor there is a particular technical difficulty related to the reference source. This has to be dynamically generated and the system involves multiple mass-flow-controllers, valves, and tubing. All these elements can, in principle, have long-term artifacts and thus affect the spectrometer performance. Nevertheless, we investigated a situation where we measured the dry synthetic air for up to ~1.5 hours. The results of the Allan-Werle variance analysis of these data is included now in the Supplementary Material as Figure S3. In this case, the spectrometer shows a precision of 0.5 ppb H_2O reached after about 2 min averaging and then it remains constant over at least on a one-hour time-scale. We added the following text to Section 4.1 of the revised manuscript to highlight this aspect:

"*Investigating the time-series of the zero air measurements over longer time scales indicated that the spectrometer maintains a stable operation over a few hours, at least. The Allan-Werle deviation minimum in this case stayed at a constant value of 0.5 ppb H_2O after about 2 min averaging (see Figure S3 in Supplementary material).*"

Drifts appearing at even longer time-scales, due memory effects in our sampling line and multipass cell (see discussion in the replies to Referee #1), were accounted for in the laboratory by re-measuring the empty-cell spectrum (i.e., dry synthetic air) every time before and after each experiment, which is then used to normalize the respective raw spectra. The stability is well demonstrated by the results of the accuracy validation (see Figure 8), which are based on repeated measurements of the same H_2O amount fractions (at different pressures) over the course of ~5 days (8× experiments of 15 hours each).

Besides, in the present case, long term drift may explain some observed effects. For instance, if the measurement at 180 ppm used for determination of the amount fraction of the secondary reference

was performed a sufficiently long time before the measurements for the other amount fractions shown in figure 9, this might explain their systematic positive offset from the expected values.

The systematic overestimation of the measured H_2O amount fractions in the linearity assessment was discussed in the context of the replies to Referee #2. This artifact was unfortunately due to a mistake in calculating the expected H_2O amount fraction levels. Applying the right value, the systematic overestimation is removed, and all measurements lie within ±0.6 % of their expected values. Figure 9, Table 4 and the text of Section 4.3 were updated according to the revised values and results. We thank the reviewer for this comment and apologize for the mistake.

Specific comments/questions:

What is the finesse of the Ge etalon used for spectral calibration of the laser scans? Is it just an un-coated flat Ge slab, or does it have reflective coatings on its faces?

The finesse of the Ge etalon is very low, i.e. about 2.44. It is just an un-coated flat Ge slab. This limitation and its impact on the line profile is now discussed in the context of our replies to Referee #2 and in Section 3.4 of the revised manuscript accordingly.

Concerning the spectral resolution, the authors discuss briefly the "point resolution" and how that is fixed by the number of data points over a laser current scan, and their group averaging, however they should address the more fundamental question of the spectral resolution afforded by the line-width of the laser. Is it narrower than the point resolution?

For the icw-driver, we estimate a current noise density of 0.74 nA/Hz$^{-1/2}$. The typically observed 1/f laser noise PSD is ~10^7 Hz^2/Hz at 1 kHz (see Liu et al., 2018). Hence, an approximate laser linewidth of about 650 kHz can be deduced. This about 8 times narrower than the point resolution that we used in our frequency scale. Considering its rather technical aspect, we decided to not include these details into this manuscript and keep the instrumental description compact.

Do they observe any excess noise in correspondence with the sides of an absorption line, before averaging several scans? At low pressure where the line sides are steepest, and at high H2O content for better S/N, excess noise relative to the spectrum baseline could reveal the effect and magnitude of the laser line-width.

We looked into the spectra recorded at 1 Hz using elevated H_2O content (180 ppm) at the lowest pressure (60 mbar) that we used in the experiments, but we could not find any excess noise in the indicated range.

Any physical or practical reason for fixing Delta$_2$=0 in the fit?

Thank you for pointing out this weakness in our description regarding this aspect. We discussed this issue in detail in our replies to Referee #2 and added a more comprehensive description of the arguments for our decision regarding the selection criteria for the line profile parameters in the manuscript.

In table 1, the caption states that the relative uncertainty levels vary between 1.4-1.47% for all conditions but that does not seem to hold for the first point in the table (0.04/2.51=1.59%)

This is due to rounding the reference values given in Table 1 to the second digit after comma. The actual H_2O amount fraction generated at this setpoint is 2.514 ppm with an expanded uncertainty of 0.037 ppm (i.e., 1.47 %). The full value with three significant figures is now given in Table 1. Thanks for noticing this issue.