

We would like to thank the Referee for the thoughtful and constructive comments that helped us to improve our manuscript. We addressed each comment individually and made revisions in response to their suggestions, as detailed below. Our replies to the Referee comments are highlighted in blue. Modifications to the manuscript text are indicated in *italic*.

Referee #2 (Markus Miltner)

General comments

This paper is of interest for the AMT audience for several reasons: 1) It demonstrates that laser absorption spectrometers are a promising alternative to cryogenic frost-point hygrometers under the particularly difficult measurement conditions encountered in the upper atmosphere (variable pressure, low and variable water concentration). 2) It gives a blueprint of a laboratory validation for such measurement devices. 3) It shows the importance of choosing a more advanced (compared to the Voigt profile) line shape model (here the qSDVP) and demonstrates how to obtain the necessary parameters, which are not contained in the HITRAN database.

The experiment is well-designed and the underlying measurements and data analysis are presented in detail, allowing fellow scientists to reproduce the measurements if they wish.

The paper is well written: It is easy to follow the authors through the chapters, thanks to the clear structure, the precise language and the supporting figures.

Specific comments

1 Introduction, p2, 125; State that only pressure broadening parameters are assessed and the line strength is not (since only pressure was varied, while temperature was held constant). This was not immediately clear to me when first reading the paper.

We agree that the description was not entirely complete and added a clarifying sentence to the Introduction as suggested:

"[...] such as the line strength, its temperature dependence, and the pressure broadening parameters. Especially, a detailed knowledge of the latter is a prerequisite for an accurate spectral retrieval. Our focus was on the broadening effects, while for the line strength and its temperature dependency we take the values from the HITRAN 2020 database."

2.3 Gas handling system, p4, 113; Flow rates not consistent with what is stated in Figure 1 (0.05 to 3 slpm vs 0.05 to 4.5 slpm)

Text revised: 0.05-4.5 slpm is the correct range.

Figure 1; Should there not be a vent somewhere between the mixing solenoid valve and MFC 3? If not, where does the excessive flow go? Please clarify this.

Indeed, an overflow valve was present between the solenoid valve and MFC 3, which was neglected in the previous schematics. It is now included in the revised version of Figure 1.

p4, 119; how was the temperature controlled?

The experiments were performed in an air conditioned laboratory at METAS with set temperature of 20 °C. The spectrometer was operated inside a plexiglass chamber (at ambient pressure) to further suppress any sudden temperature fluctuations from outside. The temperature inside the PMMA chamber (next to the multipass cell) was continuously monitored by the Vaisala HMP110 sensor and found to be stable at 23.5°C with a standard deviation of 0.03 °C. This information was now included in Section 2.3 of the revised manuscript, and the PMMA chamber is now also indicated on the schematics in Figure 1.

"All measurements were performed in an air conditioned laboratory and the spectrometer was operated inside a custom-made PMMA plastic chamber (thickness 15 mm) to further suppress any sudden temperature or humidity fluctuations. The temperature was monitored in the vicinity of the multipass cell (inside the PMMA chamber)."

P4, 125; it would be interesting for fellow scientists trying to reproduce the experiments to know how the gas cylinder used for the secondary reference gas mixture was “conditioned” and what kind of synthetic air (upper boundary of water content?) was used to prefill and pressurize the bottle.

The only purpose of this experiment was to demonstrate that the QCLAS maintains its linearity over an extended range of water content, beyond the levels obtained from the permeation method. We used a regular steel cylinder (previously containing synthetic air) and after its evacuation, we applied the procedure as described in the manuscript. The conditioning was achieved by repeated purging (with synthetic air) and evacuation steps. The synthetic air used to fill the gas cylinder contained < 5 ppm H₂O (according to the specification sheets, Messer, Switzerland). The water spiking was immediately followed by the filling process. After filling, the gas bottle was lied down and rolled multiple times at few hour intervals to facilitate mixing, and finally, it let sit for a few days for equilibration. No attempt was made to investigate the long-term drifts in this gas cylinder, e.g. due to wall adsorption and other surface effects, and therefore, we do not consider our approach as a reference method to be followed by others (and most notably, it does not fulfill SI-traceability). We slightly modified our description of this procedure in Section 2.3 of the revised manuscript to address the Referee's concerns.

"For this purpose, a regular steel cylinder (previously containing synthetic air) was evacuated, filled with synthetic air up to 1 bar and then a syringe with distilled water was used to introduce a given amount of H₂O, followed by pressurizing the gas cylinder with synthetic air (with <5 ppm H₂O content, Messer, Switzerland) up to 100 bar, resulting in a humidity content of about 180 ppm H₂O. After a few days of equilibration, the gas was expanded into a 34 L SilcoNert®2000-coated stainless-steel cylinder (Essex Industries Inc, USA) to further minimize any potential surface effects during the measurements. It should be noted that this custom-made secondary reference gas does not fulfill SI-traceability, and it is subject to well-known long-term stability issues. Its sole purpose is to assess whether ALBATROSS is capable to measure significantly higher water vapor amounts than can be generated by the permeation method."

P4, 128; it would also be good to specify the type of the SilcoNert®2000-coated stainless-steel cylinder into which the reference gas mixture was expanded. Despite the coating, surface effects might be different for different bottle sizes.

We used one 34 L Air Sampling Cylinder produced by Essex Industries Inc (St. Louis, USA). This information is now included in the text.

3.2 Pre-processing, P9, 12; I imagine that the spectral range was about 0.845 cm⁻¹, so you obtain a spectral-point resolution of 1.69×10^{-4} for your stated 5000 datapoints? Please mention the spectral range to make clear how you got to the spectral-point resolution.

The details of the spectral range are already indicated in Figure S1. We have now specified the values in Section 3.2 of the revised manuscript as well.

"The spectral range covered by the QCL is 0.88 cm⁻¹. The number of data points per spectrum is reduced by a factor of 4, i.e. from 2.1×10^4 to 5×10^3 , by using the moving average approach, resulting in a uniform spectral-point resolution of 1.67×10^{-4} cm⁻¹."

3.4. Quadratic speed-dependent Voigt profile (qSDVP), p13, 114; I do not agree with the interpretation of the residuals obtained with the qSDVP as being free of any structure exceeding the normal noise level, although admittedly the features visible in Figure 6c and d are very small. Have you tried to see if an additional fit parameter ($\Delta_2 \neq 0$) or a higher order line shape model would be able to suppress this feature? If so, it would be nice to mention this. In any case, in my opinion it would be preferable to state that there is still some small structure observable, but that it is largely reduced compared to the VP. Stressing the argument that the QF obtained with the improved fit equals the SNR, one could nonetheless justify the choice of the qSDVP as line profile?

We appreciate this remark and realize that there is some contextual issue with this paragraph. The following text was added at the end of Section 3.4 in the manuscript to better describe the arguments for our decision regarding the selection criteria for the line profile parameters :

"It can be tempting to include additional line profile parameter, e.g. Δ_2 to further reduce the remaining structures. However, our primary aim is to find an optimum compromise between establishing a reliable and accurate spectral retrieval, while maintaining the high temporal (spatial) resolution of the spectrometer during balloon flights. The latter requires that we evaluate spectra from the flights at 1 s acquisition time rather than averaging them over, e.g., 50 s. However, as the noise scales by \sqrt{t} (assuming random fluctuations), its amplitude is about 7 times larger in the 1 s data compared to the situation shown in the Figure 6, and thus, random noise-induced statistical effects dominate the spectra. In our opinion, including another degree of freedom for the spectral fit under such circumstances is largely questionable. Another aspect is the consideration of various artifacts and their impact on the measured line profiles. As indicated by the Figure S1, the FSR determination uncertainty in our case is about 1.3×10^{-5} cm⁻¹. Furthermore, the frequency stability of our free running QCL at longer time scales was found to be between 1.2 and 5.5×10^{-5} cm⁻¹ (mainly determined by the laser heat-sink temperature stability). While the former term affects the frequency scale accuracy, the latter has a random bias on the line profile when averaging over multiple acquisitions. These influences can easily induce slight asymmetries or subtle line-shape distortions that can be than erroneously assigned to the Δ_2 parameter. Moreover, our trial of using Δ_2 as a free fitting parameter in the MSF routine resulted in a weakly constrained value with large uncertainties, indicating difficulties of a proper assignment. Similarly, we also found that considering other parameters, e.g. collisional narrowing, does not improve the QF index. Therefore, we conclude, in full agreement with previous works (e.g. Lisak et al., 2015), that the observed line

shapes can be well reproduced by assigning the non-Voigt effects to speed-dependent effects rather than collisional narrowing. These facts justify the choice of a reduced model against a generalized higher-order (HTP) parameterization."

4.3. Linearity (extended-range validation), p18, 114; What about the repeatability of this measurement? Have you determined the ~180ppm concentration several times?

Due to the limited amount of the secondary reference gas (~1000 L), the measurement at ~180 ppm H₂O (i.e., the undiluted secondary reference cylinder) could only be performed once. In particular, this measurement was performed at the end of the extended-range validation routine (i.e. after 64 hours of measurements). Therefore, it cannot be excluded that a slow decrease with time of the H₂O amount fraction in the secondary reference gas (due to loss of H₂O molecules to the walls of the cylinder) would indeed result in an underestimation of the initial H₂O amount fraction. However, the magnitude of this drift can be estimated from the repeated (3×) measurements performed with diluted samples, e.g. at 140 ppm. To do this, we adopt a conservative approach and consider the statistical error on the undiluted reference gas measurement to be the sum of the standard deviation of the three individual measurements (repeatability) performed at 140 ppm (310 ppb H₂O), and the precision of the single measurement at 180 ppm (60 ppb H₂O). This results in a total statistical uncertainty of ± 370 ppb H₂O (i.e., roughly ± 0.2 %) on the H₂O content of the undiluted cylinder. This is now shown by the error bars on the ~180 ppm data point in Figure 9b. Furthermore, the following text was added to Section 4.3 of the revised manuscript to discuss this aspect

"As the H₂O content of the undiluted cylinder was only determined once, we estimate its uncertainty based on the repeatability of the measurement performed at 140 ppm (i.e., 310 ppb H₂O) and the precision of the single measurement at 180 ppm (60 ppb H₂O). This results in a total uncertainty of ± 370 ppb H₂O, i.e., roughly ± 0.2 % (as shown by the error bars in Figure 9b)."

From Figure 9b it looks like all measurements (except for the 180ppm one) are slightly too high, as if the undiluted gas actually had a slightly higher water concentration...

Following this comment, we re-evaluated the dataset used for the linearity assessment and tried to find the reason for this deviation. It turned out that the expected H₂O amount fractions were calculated using a wrong estimate of the H₂O amount fraction in the undiluted cylinder (181.47 ppm), i.e., a preliminary result of a spectroscopic retrieval performed with exploratory conditions (namely, FSR = 0.024 cm⁻¹, $\Gamma_0 = 0.0985$ cm⁻¹ atm⁻¹, $\Gamma_2 = 0.0155$ cm⁻¹ atm⁻¹). The actual H₂O amount fraction retrieved with the appropriate settings as given in the manuscript (FSR = 0.02429 cm⁻¹, $\Gamma_0 = 0.0992$ cm⁻¹ atm⁻¹, $\Gamma_2 = 0.0135$ cm⁻¹ atm⁻¹) equals 183.54 ppm. After recalculating the expected H₂O amount fractions based on this value, we found that the systematic overestimation is actually removed, and all measurements lie now 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.

Conclusion, p21, 110; “without systematic residuals”, I do not agree, as discussed above.

Rephrased ("residuals" replaced with "bias").

Technical corrections

P9, 115, replace “while secondary reference gas mixtures (panel d) at 3 pressure levels (60–200 mbar)” by “while secondary reference gas mixtures (panel d) were measured at 3 pressure levels (60–200 mbar)”

Done.

P 14, 14, delete double “the” in “while the the qSDVP fitting uses”

Done.

The color-code used in the figures for different pressures (Figure 3, 4, 6, 7) should be unified. A different color scheme should be used, since the different blue lines are hardly distinguishable.

Done: the 250 mbar lines were changed from blue to black, and the same color-coding was applied consistently to Figures 3, 4, 6 and 7.

P 25, 115-26; references are not in correct order (K after L)

Order of references corrected.