We would like to thank the Referee for the constructive comments and helpful suggestions on the manuscript, which helped us to further improve the clarity of the paper. Below, we give detailed responses (in blue) where appropriate.

General comments:

(…) I believe that the performance of the spectrometer could have been improved by improving the fit routine, therefore I suggest to the authors to explore this suggestion. This hopefully would not require them to redo the analysis, if they saved data for playing with the fit parameters at posteriori. However, I would say that this is not compulsory since the authors seems to have achieved the target precision on the measurements (although it would further improve the precision, and it would perhaps allow to remove or at least minimize the corrections on the pressure and concentration dependencies). I therefore recommend this manuscript for publication on AMT, after considering the comments below.

We fully share these thoughts with the Referee. Using more sophisticated line profiles in the spectral evaluation has the potential to reduce the current pressure dependence, but most likely the impact will only be marginal. At the moment, however, we don't have a fitting routine going beyond the Voigt-profile, but we are working on implementing the HTP-profile fitting algorithm in the near future. One serious issue here is, however, the four missing line parameters for the molecular transitions. Nevertheless, the raw spectra have been recorded and they can be post-processed later, once the parametrization is available. This aspect is now included in the conclusions as further potential for future improvement.

Specific comments:

The whole story regarding the absorption line saturation need more discussion in the paper. Authors should mention the optical power in the multipass cell as well as the beam waist. With those two parameters, the optical path length and two parameters available from the HITRAN database (the Einstein coefficient and the degeneracy of the excited state) they should be able to estimate the saturation parameter (s), which is the ratio between the intensity (I) used and the saturation intensity (I_sat) that is calculated. Then, the difficulty to find the exact saturation effect is due to the fact that it depends on the regime "homogeneous" (Lorentzian broadening dominate) or "inhomogeneous" (Gaussian broadening dominate). But authors could calculate this assuming that at 5 mbar they are in the inhomogeneous regime. Authors should therefore calculate their saturation intensity (I_sat in W/cm2) and s = I/I_sat, and therefore the effect of saturation on the absorption lines (1/sqrt(1+s)). An s of 0.05, for instance would corresponds to a 2.5% effect on the absorption lines. Then they should check if this value can be in line with an estimation of the amplitude stability of their laser source. (A good reference for saturation spectroscopy is Giusfredi, et al.: Theory of saturated-absorption cavity ring-down: radiocarbon dioxide detection, a case study, J. Opt. Soc.Am. B, 32(10), 2223, doi:10.1364/josab.32.002223, 2015).

We agree that the saturation parameter could be roughly estimated, but this will not help us improving our instrument performance. Based on experimental evidence, we know that for sample pressures above 10 mbar the saturation is negligible (< 0.2 %). The same is true for further reduced laser intensity (< 1mW). However, the gas pressure is defined by the sample availability. Therefore, we expect the gas pressure to be ≤5 mbar. Unfortunately, a further reduction of the laser intensity has a negative impact on the SNR, i.e. on the analytical precision (see also our reply to Reviewer #1).
Nevertheless, the well-founded remark of Reviewer 2 motivated us to have a closer look at the saturation effect by (i) comparing absorption profiles, recorded at a fixed gas pressure, but with two different laser intensities, and (ii) by changing sample pressure at fixed laser intensity. From these empirical data, and assuming the inhomogeneous regime as proposed by the Referee, we estimated a saturation intensity and thereof the saturation factor s of 0.014. This effect can account for about 0.7 % reduction in the line absorption. Given the discrepancy between observed and estimated values, we checked the input parameters of the fitting routine and found a mistake in the pressure conversion (hPa to atm). Thereby, a systematic bias of 2.6 % was introduced to the retrieved concentrations. Since this mainly affects the calibration factors, we re-calculated these values and also added a short paragraph about the saturation. The revision in our pressure value and calibration has no further implications for the quality of our measurements.

We are very thankful to the Referee for drawing our attention to this issue.

L51-52. In the paper you mention about discrete measurements in ice-core and here you mention a continuous extraction system. Please be more consistent on this point. And if possible, please refer to a paper "in preparation" for the extraction technique.

Text revised to: "*In this publication, we will present in detail the laser absorption spectrometer and its performance, while the continuous sublimation extraction system providing a cm-scale vertical resolution will be described in a separate paper (Mächler65et al., 2020).*"

L84-85. In the selection of the absorption lines did authors take into account the temperature dependency of the absorption lines? This should be then discussed in the manuscript.

Yes. Added a short paragraph on this topic in the revised paper.

Figure2. This wiggled shape in the residual is very large, I think authors should consider using a different line-shape profile which would include Dicke narrowing and perhaps also speed dependence collisional broadening. The intensity of the peak absorption is only 150 time higher than the peak-to-peak on the residual, which is in with the 11 permil precision on a single spectrum. But this also show that on the d13CO2 precision authors are limited to those wiggles and that they could further improve their measurements. Furthermore, looking at HITRAN2016 simulation the intensities at 5 mbar and 20°C do not seems to have the same intensity as showed here. Is there an error in the HITRAN database? If so, authors could mention it in the paper and say which line has the wrong cross-section and by how much is wrong (based on their gas standard mixtures). Last, they should be mention that this is a single scan spectrum (acq. Time 75 us). What is the strategy then? How many spectra the authors average before perform the fit? This should be also mentioned.

See above our reply to the general comments. Our midterm goal is to implement the HTP-profile fitting algorithm for spectral evaluation. However, this will mainly remove systematic biases, i.e. improving accuracy rather than affecting precision. Furthermore, as we use the integral of the absorption profile instead of the peak amplitude, the situation is much more relaxed as the integral over the residual is approaching zero. The current performance allows us to reach the precision targets, but we have to perform careful calibration to achieve the required accuracy. The information about spectral averaging is given at line L190 and a link to Fig.2 is provided at L196. Added sentence on future potential development in spectral analysis in the Conclusions.

L116-119. Please provide the power injected in the MPC. Are authors not annoyed by optical fringes from the ND filter?

The laser power after the ND filter entering the MPC is 5.4 mW. The ND filter was custom-made using a wedged substrate. This reduces the unwanted etaloning effect and the remaining modulation is as broad as our spectral coverage. Nevertheless, mechanical and temperature effects still may affect the instrument long-term stability. Therefore, we use a water-cooled optical plate, stabilized at 5 mK, and a hermetically sealed housing around the optical module. We added this information to the manuscript.

L121: how the accuracy on the temperature was measured? With an independent temperature probes? is this the accuracy of the temperature stabilization or just the precision of the temperature sensor itself? Please specify. This could then be related to the temperature dependency of the absorption lines and to the achieved precision on the d13CO2.

The base-plate temperature was measured indirectly based on the T-sensor included in the thermochiller used to circulate the cooling liquid. The gas sample temperature in the MPC was measured by high-precision NTC sensors, which were not calibrated in our laboratory and may have some bias (~0.5 %), while their precision is much better (mK). The temperature effect on $\delta^{13}C(CO_2)$ can be easily estimated considering the lower state energies of the involved transitions, which results in a temperature sensitivity of ~16 ‰/K. Although it may be difficult to maintain long-term stability of the absorption cell to better

than 0.05 K, changes in the cell temperature can be measured with a precision on the order of mK using high-precision thermistors. Since the measured temperature is used continuously for interpreting the absorbance spectra, the temperature dependence of the line strength is not a major impediment to obtain isotope ratio precisions better than 0.02 ‰ for d13CO2. We added this information to the manuscript.

L130-131. What about for the d13CO2? It is weird that author fixed they required optical path length based on the CH4 and N2O required precision rather than on the d13CO2 measurement.

Fig.2 shows the absorption of the different species with their corresponding axes (left and right). Please note the scaling difference between these axes. The CH4 and N2O absorptions are less than 0.5%, while those for CO_2 are about 60 %. This clearly shows that the required optical path length was mainly driven by CH_4 and N_2O. We changed the wording to include this detail.

L139. Significantly longer. Could you provide a value?!

Up to several hours. The required duration was strongly dependent on the conditioning history of the cell, but even by baking, purging, and evacuating the cell over few days the pump-down time of the cell after filling with a sample gas took more than 20 minutes. This information has been added to the manuscript.

L155-156: How authors can simulate the frequency of the optical fringes produced by the multiple passages?! It is just a maximization of the distance between the neighboring spots as mentioned few lines below or is something else?

Yes, but not only. Besides the lowest possible overlap of the individual reflection spots, we also take into account the optical path difference of the neighboring spots and try to avoid those patterns where the interference fringe frequency generated by the optical path difference is comparable to the width of the absorption line. Especially, overlaps of the beam spots with small pass number (e.g. 4 and 6) differences are problematic. Searching for patterns that mostly fulfil these criteria is expected to result in a reduced optical fringe level or at least have less impact on the absorption line retrieval. These considerations have been added to the manuscript.

L203. Critical orifice. Can authors mention the size? Well, in flow measurement this would probably fractionate...please explain.

Added a short text to clarify this issue: "*The critical orifice size (20 µm) as well as the output pressure of the regulators were all carefully selected to avoid isotopic fractionation as the gases flow through the system. A series of experiments was conducted in flow-through and static regime to investigate potential fractionation effects due to flow-restriction elements such as critical orifice and the optional metering valve used in the flow-through mode. Their effect was found to be lower than our detection limit. With this setup it is possible to operate the instrument both in flow through or in batch mode for discrete samples.*"

Table 1. When second digit precision is reported (eg. 0.01) then concentration should be reported with the same number of digits (eg. 248.8x ppm). Did authors monitor their standard mixtures over time? Do they see any drift?

Added the second significant digit to the values.

Figure 7. It would be interesting to see the same AW statistical analysis for a static experiment. Can authors add this information? Since the ice core measurements will certainly done using discrete samples and in static mode, I guess.

We performed several experiments of this kind, also because we had the same concern as the Referee. These tests resulted in very similar precision values as those conducted in flow-through mode. Of course, in this case we had to limit the measurement period to shorter time (< 1 h), to avoid drift effects caused by surface effects. The performance in batch mode is also demonstrated by the following sections of the

manuscript, where we systematically focused our attention on discrete samples and demonstrated analytical performance obtained using such 1 mL samples. In our opinion, these data are more informative then showing a replicate of an Allan-Werle plot.

L281: This sequence is repeated several times: Can authors provide information about the frequency of those 50 measurements?

The time required to measure consecutive sample/reference pairs was 10 minutes, thus, for the 50 repetition cycle, 8.3 hours was necessary. We added this information to the revised manuscript.

L293-294. "Even for discrete samples...". What do they mean with this sentence? Because for a single discrete measurement the performances are maintained for a time given by the AW analysis, which is shown only a maximum integration time of 16 minutes. What authors mean with this 24h of continuous operation? Can they prove it with an AW plot?

Individual or discrete gas samples can be measured with a precision as determined by the AW analysis. However, when switching between sample gases collected in dip-tubes further additional artifacts/biases, mainly originating from gas handling, can affect the accuracy of the measurements. Of course, during the time that is required to replace the sample gas with the next sample the spectrometer can drift. However, we demonstrate that with our cyclic measurement approach, i.e. quickly alternating between dip-tube sample and reference gas measurement, these drifts can be accounted for and thus their effect can be minimized even for longer periods of time, e.g. 24 h. Figure 8 demonstrates exactly this situation.

L311-312: "This two-point calibration..." This is normally not a linear relationship, but rather an exponential with discrepancy becoming larger and larger at low concentrations...It is possible that authors would be able to approximate it to a linear relation in the range of concentrations they will have in the ice cores? Why authors did not just apply different dilution factors to on of their standard to study this effect on a larger concentration range? Just to mentioned, I see an artefact of 0.94 permil on the working range of 157.7 - 345.5 ppm, which is more the 20 times larger than the claimed accuracy.

In our experience, the linear relationship is a very good approximation (see also the fit residuals on Fig.9) for such a narrow concentration range. We did several measurements in the past using the dilution approach as also suggested by the Referee, and we always found a tight linear correlation even for various molecular species, e.g. Tuzson et al., Appl. Phys. B 92, 451–458 (2008) and Waechter et al, Opt. Exp. 16, 9239-9244 (2008). Our focus was only for the concentration range expected for ice core samples, because this is the main application for which we developed the instrument and which we fully cover with our custom-made standard gases. The dilution is feasible, but it is a time-consuming process that we want to avoid when analyzing ice core samples. That's why we opt for a robust two-point verification approach. Unfortunately, we are unable to follow the Referee regarding the artefact.

Figure 9. So, if I understand well, the difference between the measured and the reference corresponds to the difference between the measurement done by the commercial instruments used for calibrate the standard gas mixtures and your new instrument, is that correct? But, you should say what is your approach for your instrument. Do you stick with the intensities provided by HITRAN database or you calibrate the spectrometer with one of the standard bottles? Because if it is the second option, then you should have at least one point which would match well the reference measurement. And what about the d13CO2? Because for that you have to rely on a reference measurement. So, I do not get why there is this offset between reference and measurement for all isotopic ratios rather than a crossing point.

Yes, that is correct. The concentration of the trace gases in the reference cylinders were determined using the WCC instrumentation, while the $\delta^{13}C(CO_2)$ values were determined by IRMS, as described at L240-L243. The raw values reported by our spectrometer are purely based on spectroscopic (HITRAN) and physical parameters (OPL, T, and P) without any additional calibration. In the case of the $\delta^{13}C(CO_2)$ the same approach was used, but included the scale conversion from natural abundance used by HITRAN

and the *Rs* value defined by the VPDB-scale. Finally, a constant offset of 9 ‰ was added to bring the "spectroscopic"-scale closer to the VPDB values. This last step could also be included in the calibration function. The calibration curves in Figure 9 then illustrate the conversion of the instrument values to international scales. We revised the corresponding paragraph to include these details.

L343. "This indicates that..." You should explain your approach here. I believe that you leave the Lorentzian contribution on the absorption lines free to adjust with pressure. This should be mentioned in the text. As well as that the concentration in mixing ratio is then corrected for the total gas pressure in the cell. I believe that the artefact you see could be removed (or at least minimized) by changing the fit profile as mentioned above.

Exactly, the Lorentzian width is calculated based on the actual pressure readings and using the pressure broadening parameters γ_{air} while the concentration is determined using the ideal gas law. We added this information to the manuscript. We also think that the artifact shown in Fig. 10 could be reduced by using a more appropriate line profile model that includes at least the Dicke narrowing term. Further contributions can be due to the uncertainties in the pressure broadening coefficients, line strengths and frequency scaling as well as optical saturation.

L345. Can authors estimate the dependency for d13CO2 around 5 mbar? what precision in pressure is required the 0.04 permil precision and how it compares with your accuracy on the pressure measurement?!

Yes, the dependency of $\delta^{13}C(CO_2)$ on pressure around 5 mbar is 0.81‰/mbar, thus we need 0.05 mbar accuracy on the pressure measurement to achieve 0.04 permil precision on $\delta^{13}C(CO_2)$. Our pressure sensor has a stated accuracy of 0.2% of the measured value, i.e., an uncertainty of 0.01 mbar for the 5 mbar sample pressure. We included this information in the revised manuscript.

L356: "possibly at value >= 5 mbar." Here you should rather report 5 +/- xx mbar, with the error estimated in order to stay within the 0.04 permil (comment above).
Done.

L370-371: mention that those are 1 sigma precisions.
Done.

Technical corrections:
L9. I would rather call it multi-pass absorption rather than direct absorption (or directmulti-pass absorption if authors want to keep the word direct).
We do not agree. The technique used here is the direct absorption spectroscopy, while the multipass is just to enhance the signal.

L61-80. This part should go to the introduction rather than in the methods. The latter should be focus on the description of the methods.
We agree and implemented the suggested changes.

L108. Mention the photodiode model.
Done.

L173. I suggest By applying.
Done.

L293: replace "targets" with "target performances"
Done.