Author's reply to the referees comments to manuscript AMT-2017-120 - Anonymous referee 1

The original referee's comments are written in black and the author's reply and changes to the manuscript are colored in blue/green respectively. References to page and line numbers as well as figures refer to the original manuscript, but references to sections refer to the corrected manuscript. In cases where we insert figures, tables and equations into this document, they are referenced with R1, R2, R3 …

The paper "A new instrument for stable isotope measurements of ^{13}C and ^{18}O in CO_2 - Instrument performance and ecological application of the Delta Ray IRIS analyzer", by Braden-Behrens et al, reports on a recent commercial instrument for measurements of stable CO_2 isotopes (δ^{13}C -CO_2 and δ^{18}O -CO_2) and its application in a field study. This work is relevant, however it is lacking many essential elements and it is not written carefully enough. The manuscript will thus need a major revision to be considered for publication in AMT.

This review focuses mainly on the performance evaluation of the instrument. It is important to also carefully review the hypothesis and the conclusions drawn from the field study. Most remarks and suggestions are added to the original manuscript. Some mayor aspects are discussed below.

Author's response: We thank the anonymous referee for the detailed feedback and suggestions to our manuscript, below we answer the referee's comments, starting with the major aspects mentioned in the review and followed by the additional comments in the supplement.

p3/22 The exact wavelength should be given together with a measured and fitted spectrum. The spectral range and the spectral resolution are important elements to judge the analytical performance, also in the context of gas matrix effects. One may assume that the frequencies are as in (Geldern 2014), but the latter does not show a measured spectrum.

Author's response: Thanks for pointing this out. We added the information about the spectral region and the used absorption lines to the manuscript. We also added more details about the drying of the air sample and the spectral fit to the description of the instrument in the introduction.
Changes to the manuscript: We added the following to the introduction.
"The instrument scans a spectral region from 4.3293 µm to 4.3275 µm (Geldern, 2014), containing four CO_2 absorption lines: at 4.3277 µm and 4.3280 µm (both for ^{16}O^{12}C^{16}O), 4.3283 µm (for ^{16}O^{13}C^{16}O), and 4.3286 µm (for ^{16}O^{12}C^{18}O). A measured and a fitted spectrum is shown in Figure R1. The fitting procedure is based on a Voigt-Profile fit, that relates the isotopologue-specific absorption lines to their respective concentrations (information from the manufacturer, Thermo Fisher Scientific)"
We added a chapter about the spectrometer setup to the methods:
"Spectrometer setup
We set up the spectrometer to use the absorption lines at 4.3277 µm (for ^{16}O^{12}C^{16}O), 4.3283 µm (for ^{16}O^{13}C^{16}O), and 4.3286 µm (for ^{18}O^{12}C^{16}O). Thus, only three of the four absorption lines in the instrument's measured spectra (Figure R1), were used for the spectral fit. In particular, for ^{16}O^{12}C^{16}O, we did not use the strong absorption line at 4.3280 µm. The corresponding mode of operation is called "high concentration mode" in the instrument's operational software QTEGRA. Additionally, the sample was dried before it entered the measurement cell with the (instrument's internal) Nafion drier."
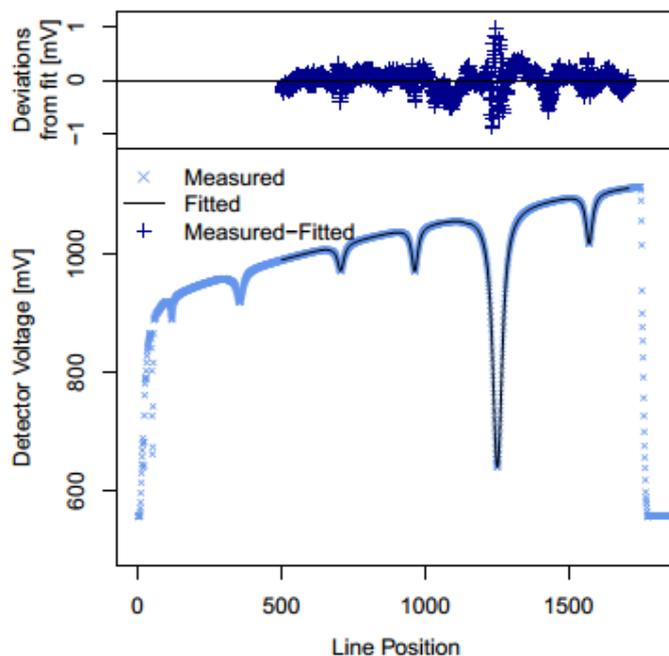
Figure R1 Measured and fitted spectrum, as exported from the instrument's operational software QTEGRA.

Water vapor may significantly impact the retrieved δ-values, either through spectral interference or through changes in absorption line characteristics (pressure broadening). If my understanding of the setup is correct, then humid samples were measured spectroscopically. Since this paper aims at validating a new spectrometer, it is vital to discuss and quantify the effect of changes in humidity.

Author's response/ Changes to the manuscript: The sample was dried with a Nafion drier before it was measured. This information was added to the manuscript, c.f. comment to p3/22.

(Along the same line) p3/23 describes the Delta Ray having "an internal calibration procedure that automatically includes two point calibrations for concentration c and both δ values as well as corrections for the concentration dependency of the measured d-values". This concept is interesting and a key feature of the Delta Ray. However, since this publication evaluates a commercial instrument, it should clearly describe the way concentration dependency is corrected (and how large it is) and to validate the procedure (accuracy, see above). This has not been achieved or is not presented.

Author's response/changes to the manuscript: We addressed this question by adding the chapter 'Evaluation of the calibration strategy', c.f. our comment to page p6/20, in especially Figures R2 and R3 below. Additionally, we changed the chapter about the calibration procedure to provide more detailed information. (We changed the order or your comments here, because we refer to this chapter later.)

2.6 Instrument internal calibration

The Delta Ray analyzer is equipped with three different internal calibration routines (Thermo Fisher Scientific, 2014). We performed these routines at the field site (in situ) each time the analyzer had to be restarted e.g. after power supply failures, instrument issues or when we manually turned off the analyzer for other reasons. All three instrument internal calibration procedures were usually done one day after restarting the analyzer, thus the instrument was in thermal equilibrium during calibration. The three different instrument internal calibration procedures are described below:

- **Correction of concentration dependency (called 'linearity calibration' in the instrument's documentation and operational software)**
  This calibration routine evaluates the concentration dependency of δ value measurements (Thermo Fisher Scientific, 2014). Mathematically, an experimentally derived correction factor $f_{correct}$ (c_{raw}) is multiplied with the raw isotopic ratio R (information from the manufacturer, Thermo Fisher Scientific)

  $$R_{c-corrected} = f_{correct}(C_{raw}) \times R_{raw}$$
  (Equation R1)

  This factor as a function of concentration is determined via a natural spline fit of measurements of a gas tank with constant δ value at different concentrations (information from the manufacturer, Thermo Fisher Scientific). This is implemented by mixing pure CO_2 with CO_2.-free air, yielding concentrations between 200 to 3500 ppm. In our setup we used the pure CO_2 with near to ambient δ values (tank 'ambient CO_2 ', c.f. Table 3) and synthetic air for this calibration.

  […]

The instrument's internal calibration procedure is based on the measurement of these calibration curves after the instrument is started in combination with repeated measurements of a known gas, so called 'referencing' (see below). As the different calibrations are only performed once after the instrument is restarted, the accuracy and repeatability of measurements is further based on the assumption that, these relationships remain sufficiently constant, and temporal changes are corrected by 'referencing'.

- *Referencing*
  This procedure applies an offset correction of the calibrated δ values using a gas with known δ values that is measured at a freely selectable concentration in regular intervals (information from the manufacturer, Thermo Fisher Scientific). In our experimental setup, referencing is carried out every 30 minutes for 80 s after the tubes have been purged for 60 s using the pure CO_2 standard ('ambient CO_2', c.f. Table 3) diluted with synthetic air. We chose the reference concentration to be the same as in the highest inlet in the adjacent cycle, because most of the measurement inlets had concentrations close to those at the highest inlet and the temporal variability of the measured concentrations generally decreased with height. Thus, we performed the 'Referencing' as close as possible to as many height measurements as possible by these settings."

Thus, the calibration procedure for δ values can be expressed with the following formula with the correction factor $f_{correct}$ (c_{raw}) as determined from the concentration dependency correction, and the slope $m_{\delta scale}$ derived from the δ scale calibration (information from the manufacturer, Thermo Fisher Scientific).

$$\delta_{calibrated}(R_{raw}; C_{raw}; t) = m_{\delta scale} \times \underbrace{\left(\frac{f_{correct}(C_{raw})R_{raw}}{R_{std}} - 1 \right)}_{\delta_{c-corrected}} + \delta_{Offset}(t)$$
(Equation R2)

p4/12 physically different samples: There is no indication that the instrument was used in a batch mode configuration. In continuous flow mode (as the text suggests), mixing in the cell (and to some

extend in the tubing) corresponds to a low-pass filter, which is fundamentally different to "physically different samples".

Author's response: We wanted to make sure, that we do not measure air samples that are majorly composed of the air masses in the previous measurement, thus we chose an averaging time that is larger than tau$_{5\%}$, yielding a situation in which less than 5% of the previous sample is mixed into the new sample (as $\tau_{5\%} = \tau_{10\%}$ ln(0.05)/ln(0.1) $\approx$14s, c.f. section 3.1.4). We agree that the formulation is misleading and changed the sentence to:
Changes to the manuscript: We changed this to "consisting of four measurements each averaged for 20 s - thus the averaging time is longer than the instrument internal cell response time $\tau_{10\%}$ c.f. section 3.1.4"

p4/~~1~~220 "temporal stability" is not standard terminology and only used once in this paper. I suggest using "repeatability", following the international vocabulary of metrology (VIM) throughout the text.

Author's response/changes to the manuscript: We changed this terminology and use 'repeatability' throughout the text.

p6/20 Accuracy was tested by comparing with one (1) gas tank which was measured using an Aerodyne spectrometer. This part of the study is a key element and completely insufficient. The main challenge in laser spectroscopy is currently not (any more) precision but rather accuracy. There is no reason why anyone should trust another spectrometer (here Aerodyne) without a very detailed description of how the latter achieves traceability. Furthermore, accuracy will depend on at least two calibration scales, i.e. δ values and concentration. Therefore, the evaluation must (!) include measurements of traceable (likely IRMS) gases at different δ values and concentrations; otherwise it is an insufficient and somewhat random exercise. If this is not possible, then an alternative may be to use traceable standards and (!) field samples that are quantified in a traceable way. This is easily possible for δ^{13}C -CO$_2$, but more difficult for δ^{18}O -CO$_2$ because of the limited stability of the samples (see e.g. Tuzson 2007, DOI: 10.1007/s00340-008-3085-4).

Author's response:
Concerning the general concerns about our accuracy measurement with N=1: We changed the chapter about accuracy in the manuscript and included measurements with gas tanks at different concentrations (N=4) and δ values (N=4 for ^{13}C and N=5 for ^{18}O), see description below. Here, we also evaluate 'potential accuracy' as defined by (Tuzson 2007, DOI: 10.1007/s00340-008-3085-4).
Concerning the comparison to the Aerodyne instrument: We agree that it is problematic to use another laser spectrometer for comparison here. We additionally analyzed this tank with a Picarro (for CO$_2$ concentration) and IRMS (for δ values) at Max Planck Institute for Biogeochemistry in Jena. In the revised manuscript, we use only gas tanks that were measured at MPI in Jena, both: for calibration as well as measuring potential accuracy (c.f. Table 3).
Changes to the manuscript: We rewrote chapter 2.5 to include additional measurements:
2.6 *Instrument characterization measurements*
We carried out additional measurement in the field and in the lab to quantify precision, evaluate the calibration strategy and quantify the instrument's response time and repeatability. These measurements involved changes in the analyzers plumbing. For all measurements that required connecting different gas tanks to the analyzer, they were either connected directly to the analyzer's internal ports ('CRef1' and 'CRef2') or the plumbing was equivalent to the plumbing of the target gas (Fig.1).

1) Lab measurements to quantify precision and evaluate the calibration strategy
   - We measured the Allan deviation by connecting pressurized air at atmospheric δ values to the analyzer and took measurements at the analyzer's maximum data acquisition rate of 1 Hz for two hours.
   - We diluted pure CO_2 with synthetic air over a CO_2 concentration range of 200 to 1500 ppm to measure the concentration dependency of the measured (raw) δ values. This dilution experiment was carried out for three different tanks with pure CO_2 at different δ values. Each gas tank was measured twice. (Used gas tanks: "ambient", "bio1" and "bio2", c.f. Table 3.)
   - We measured the concentration c and the isotopic compositions $\delta^{13}C$ and $\delta^{18}O$ of gases with concentrations ranging from (350 to 450 ppm) and isotopic compositions ranging from -37 to -9.7 ‰ for $\delta^{13}C$ and from -35 to -5 ‰ for $\delta^{18}O$. Each of these measurements was performed three times. (Used gas tanks: "ambient", "bio1","bio2", "PA-tank", SACO$_2$ -350, SACO$_2$ -450, SACO$_2$ -500, c.f. Table 3.)
   - We performed measurements of two pure CO_2 gas tanks at different δ values (diluted to different concentrations between 200 and 3000 ppm) as well as measurements of two gas tanks at different CO_2 concentrations (350 and 500 ppm). These measurements were repeated every six hours for a period of one week. (Used gas tanks: "ambient", "bio", ('SA-CO$_2$-350' and 'SA-CO$_2$-500', c.f. table 3.)"

2) Field measurements to quantify the setup's response time and repeatability
   - The response time of the tubing and the analyzer was measured by using the automatic switching unit (Figure 1) to switch from ambient air (height 1) to the target standard. We superimposed the measurements of four switching events to observe the adjacent turnover processes.
   - The analyzer's repeatability under field conditions was quantified by the half hourly target measurements described in Sect. 2.5.

We removed the chapter "Accuracy" and replaced it by the following:
3.1.2 Evaluation of the calibration strategy

The instrument's internal calibration strategy (described in section 2.7.1) is based on:

- A nonlinear relationship between raw δ values and concentrations (Figure R2).
- A linear relationship between calibrated δ value (measured with IRMS) and the concentration-corrected δ value - $\delta_{c\text{-corrected}}$ in Equation R1 (Fig. R3, left panel).
- A linear relationship between measured (raw) and real concentrations (Figure R3, middle and right panel).
- The repeatability of the calibration curves – for δ values modulo the Offset correction, that is applied by the instrument's internal 'Referencing' (Figure R4 and Table R1).

Raw δ values show a nonlinear dependency from raw concentrations (Fig. R2). This nonlinear relationship deviates from the concentration-dependency correction applied by the instrument ($\delta_{c\text{-corrected}}$ in equation R2). In Fig. R2, this function is shown for the used gas tank 'ambient' after an Offset correction at a concentration at 400 ppm, which is similar to the instrument's internal 'referencing'. Thus, the deviations of the measured δ values from the concentration-dependency correction (top panel of Fig. R2) give an estimate of the uncertainty of measurements that is related to the deviation from the reference concentration. For referencing at 400 ppm, these deviations were below 0.2 ‰ for ^{13}C and 0.4 ‰ for ^{18}O.
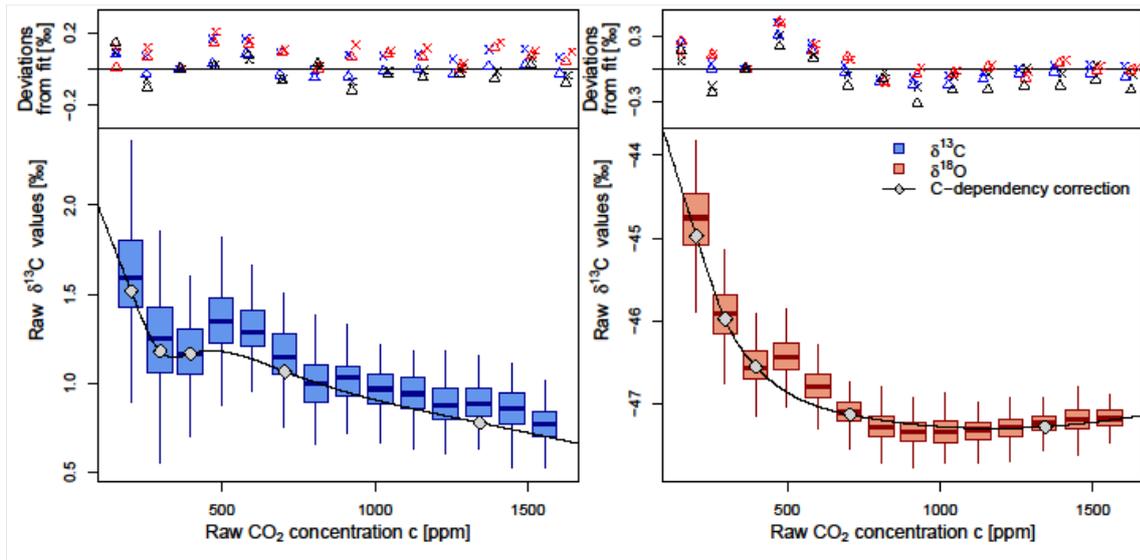
Figure R2 Box whiskers plots showing the nonlinear concentration-dependency of raw δ values for ^{13}C and ^{18}O respectively, here as an example for the CO_2 tank 'ambient'. This measured c-dependency is compared to the respective concentration-dependency correction (black line, with grey symbols marking the data points used during the respective calibration measurement). The c-dependency correction is Offset-corrected to match the raw δ values at 400ppm and the mean deviation from the fit is shown in the top panel for two measurements (different symbols) with three different gas tanks ('ambient' in blue, 'bio' in black and 'bio2' in red).

The measured linear relationships for concentration and δ scale calibration (Fig. R3) have R^2 values of above 0.9999 for concentration, above 0.999 for δ ^{13}C, and above 0.998 for δ ^{18}O. The linearity and potential accuracy, as defined by (Tuzson et. al., 2008) can be quantified as the 1σ standard deviation from the linear fits. The so defined potential accuracy of the instrument internal calibration is 0.45 ppm for CO_2 concentration; 0.24 ‰ for δ^{13}C and 0.3 ‰ for δ^{18}O. For both δ values, this is comparable to the uncertainty related to the nonlinear concentration calibration that varies with δ and c as discussed above.
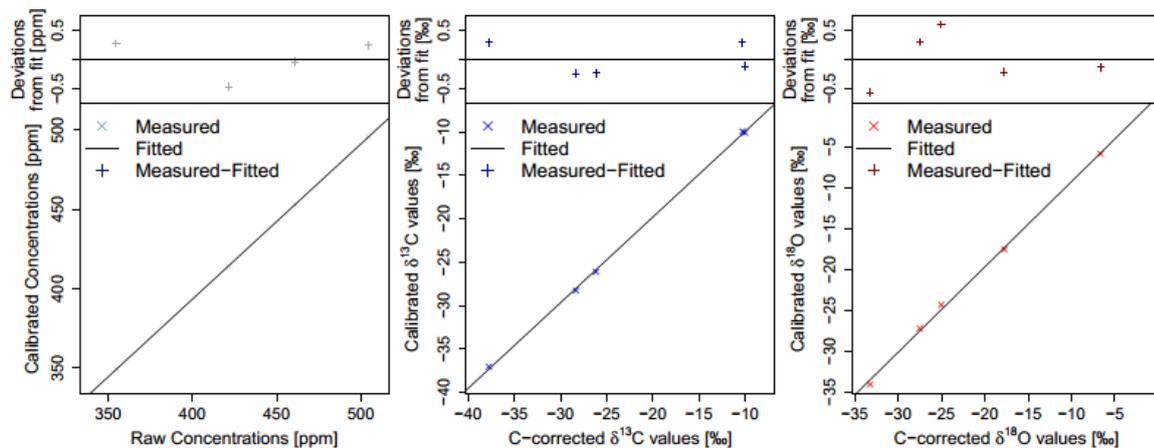

Figure R3 Linear calibrations for concentration (left panel), $δ^{13}C$ (middle panel) and $δ^{18}O$ (right panel).

The repeatability of the calibration curves is discussed here based on measurements of the nonlinear concentration dependency (Figure R2), and repeated measurements of gas tanks with two different c

and δ values to evaluate temporal changes in the respective linear relationships (Figure R3). These measurements were taken every six hours for a period of nine days. The standard deviation of the different measurement is below 0.2 ppm for concentrations and below 0.05 and 0.1 ‰ for ^{13}C and [18]O respectively. Thus the uncertainty related to the repeatability of the linear calibrations is smaller than the potential accuracy discussed above. For δ values, these values are comparable to the repeatability reported by several authors measured with other laser spectrometers (e.g. Sturm et al 2011; 2013; Vogel et al 2013). For concentrations on the other hand, Sturm et al 2013 reported a much smaller value of 0.03 ppm, based on more frequent calibration. In our setup, the concentration calibration is only performed once after the instrument is restarted, thus there might be a potential for better repeatability in concentration measurements by more frequent concentration calibration. For δ values, the repeatability that is related to deviations from the reference concentration depends on concentration (Table R1). Repeated measurements of these deviations have standard deviations of below 0.15 ‰ for concentrations between 200 and 1600 ppm.
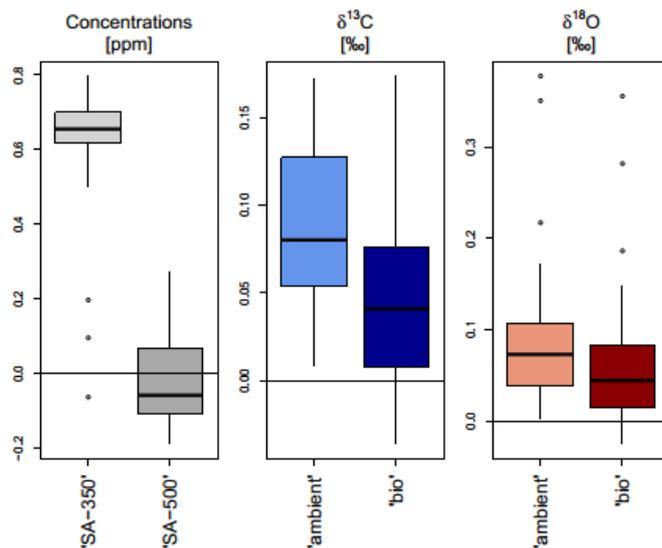

Figure R4 Box whiskers plots for the deviations of calibrated concentrations and δ values from laboratory measurements (at MPI in Jena) for repeated measurements of different calibration tanks (c.f. Table 3 for c and δ values of the gas tanks) over a period of 9 days (N=36). Delta values were measured at 400 ppm and 'referencing' was done app. Every 30 minutes at 380ppm to simulate conditions during a measurement campaign.

	tank ‚ambient'		tank ‚bio'	
Concentrations	σ (δ^{13}C- δ^{13}C$_{tank}$)	σ (δ^{18}O- δ^{13}C$_{tank}$)	σ (δ^{13}C- δ^{13}C$_{tank}$)	σ (δ^{18}O- δ^{18}O$_{tank}$)
202	0,07	0,14	0,09	0,13
396	0,04	0,05	0,08	0,08
600	0,09	0,08	0,12	0,12
807	0,08	0,08	0,11	0,11
1018	0,10	0,08	0,13	0,11
1232	0,12	0,09	0,13	0,11
1450	0,14	0,11	0,15	0,12
1664	0,14	0,11	0,14	0,12
3145	0,17	0,15	0,17	0,15

Table R1 Standard deviations σ of the differences between the calibrated δ values and the known values of used tanks 'ambient' and 'bio' over a large concentration range.

p6/26 Measurement of the Allan plot was done in the lab because of limited gas supply in the field. This is not sufficient, because the goal of this study is characterization under field conditions. The argument of limited gas in the field is not convincing because at 80 ml/min, it would easily be possible to have many corresponding measurements of about 10 - 30 minutes, which, given an Allan Minimum at around 100 s, would be sufficient. A minimal approach would be to evaluate the 80 s target gas measurements. Alternatively, or in addition, one may use ambient conditions that are sufficiently stable (e.g. well mixed, afternoon, highest sampling port) to obtain at least a conservative estimate for the precision in the field. Finally, data from the PA tank measurement also give an indication of precision in the field.

Author's response: Thanks for these suggestions. We used the field measurements with the PA-tank to calculate Allan Deviations under field conditions at an averaging time of $\tau =1$ s, yielding comparable values, c.f. table R2.

σ_A ^{13}C [ppm]		σ_A ^{18}O [ppm]		σ_A c [ppm]	
Lab	Field	Lab	Field	Lab	Field
0.29	0.34	0.40	0.44	0.09	0.09

Table R2: Comparison of Allan Deviations at 1 s averaging time based on field and lab measurements.

However, based on your questions about the calibration strategy (see comment to p6/20 above) and your comment about instrument characteristics under field conditions (see your comment to p1/4 below), we decided to add more lab measurements to this manuscript (e.g. measurements to evaluate the calibration scheme). Thus in the revised manuscript, we generally focus more on lab measurements to characterize the instrument.
Changes to the manuscript: We removed "under field conditions" in the abstract, and rewrote chapter '2.5 Instrument characterization measurements' see our answer to your comment to p6/20.

p7/30 Referencing was done at the concentration of the highest sampling port. Discuss the uncertainty resulting from the fact that some height had other concentrations, taking into account the "linearity calibration" (which does not test linearity but dependence of the retrieved δ values on c; a terminology that should be improved).

Author's response:
Concerning the terminology: We agree that the term "linearity calibration" is not very clear. We used this term because this is the name of the corresponding calibration procedure that can be found in the Delta Ray's manual as well as in the operational software. Thus, we think we should keep this terminology, to be consistent with the manual. To avoid misunderstandings, we replaced "linearity calibration" by "Correction of c-dependency (called 'linearity calibration' in the instrument's documentation and operational software)" at first occurrence and by "Correction of c-dependency ('linearity calibration')" for the following occurrences.
Concerning the uncertainties related to the referencing: We addressed this question by adding a chapter 'Evaluation of the calibration strategy', c.f. our comment to page p6/20, in particular Table R1 and Figure R2.

p8/1 This whole chapter is badly written and should be revised with respect to language. In addition, the arguments are not convincing. The concentration range of HS and LS is not any larger than the standards used in the first calibration (300 and 430 ppm). Choosing two out of five standards, that were meant to evaluate accuracy for calibration, leads to only three remaining standards that are perfectly bracketed. The mean and uncertainty at N=3 becomes then statistically very weak. Furthermore, the results for c also illustrate why using just one tank to assess the accuracy of the δ values is not sufficient and somewhat arbitrary (see p6/20 above).

Author's response: We originally introduced this post calibration because we found a large jump in the concentration measured with the target standard. No such jump occurred in δ values. The jump in the target concentration could be removed replacing the instrument internal calibration with the applied post calibration. We agree with you, that it is not convincing that this is related to the concentration range of the instrument internal calibration. We think that during this period there was a problem with the instrument internal concentration calibration. The reason is not very clear to us; it might be that we have a problem with target gas flow during this particular concentration calibration. After replacing this particular concentration calibration by the linear post calibration, the corresponding jump in the target standard disappeared.

Changes to the manuscript: We rewrote the chapter about the post calibration and applied it only for a time period in which we observed a jump in the target concentration. "For the time period from the 15th of October to 15th of November, we replaced the instrument's internal concentration calibration by a manual linear calibration, based on manual measurements with five different gas tanks in the field. This was necessary, because measurements with five different gas tanks (including the target standard) showed a consistent linear relationship between raw and known concentrations, that deviated from the linear relationship that was used in the instrument's internal calibration. Thus, we conclude that during this period there was a problem with the instrument's internal concentration calibration which might be related to gas flow or a leak during this particular concentration calibration."

p9/15: The authors state that they chose an averaging time of 20 s as compromise between number of measurements and precision. This is misleading or not clear enough. If there are no measurements of standards between 20 s intervals, then the precision does not mean much because the next mean value for 20s may have an excellent precision (given as SD) but may have drifted significantly, thus the two values with good precision cannot be compared at the level of their individual precision (it then becomes an issue of repeatability or accuracy, depending on the context).

Author's response: We removed this misleading description.

p9/20 "the mean deviation of N=300 field measurements of a tank with pressurized air" is a suitable way to quantify repeatability and should be compared (or moved) to the results found in the corresponding chapter 3.1.3. Unfortunately, the values are only given graphically in Fig. 3. However, looking at the difference of one (!) sample, one cannot determine accuracy of the spectrometer. Especially not for an analytical technique which is known to be strongly dependent on concentration and gas matrix. The test is thus not suited for its aim. This chapter and the next can be combined to determine repeatability (preferred terminology), and which - at least in the title - may be called long-term stability. However, it is critical to find a way to reliably determine accuracy.

Author's response/changes to the manuscript: See comment to p6/20 about the accuracy measurement.

p9/25 "sum of uncertainties". What the authors likely mean is the combined uncertainty or an uncertainty budget. However, this is not achieved by simple addition of the uncertainties, as suggested in the text. It is necessary to know what the authors consider for the individual uncertainty contributions (and why), what distribution they assume and – if the contributions are independent – how they calculate the combined uncertainty, and at what level of confidence they then express this combined uncertainty.

Author's response/changes to the manuscript: This section was removed, instead we discuss 'potential uncertainty' as defined by Tuzson 2008**.**

P10/7 The standard deviations of repeated measurements (0.2 ppm for CO_2 concentration and below 0.3‰ δ values) should be compared to literature values. For example, (Sturm 2013, amt-6-1659-2013) found repeated measurements of the same gas tank with a standard deviation which is a factor 4-7 better than the results shown here.

Author's response/changes to the manuscript: We added more lab measurements and discuss the repeatability in chapter 'Evaluation of the calibration strategy', including the comparison to literature data, please see our comment to your question p6/20. Here we added the following to the discussion – please note that these values slightly changed, because we removed two periods with known instrument problems (c.f additional footnote and new figure 4) and recalculated the post-calibration only for the time period in which we observed a problem with our target measurements.

"Repeatability during the field campaign

For concentration, the measured repeatability of 0.3 ppm is slightly larger than the repeatability of the concentration calibration discussed above but still below the potential accuracy discussed in section 3.1.2. In the case of δ values, the obtained repeatability of app. 0.2‰ for ^{13}C and 0.25 ‰ for ^{18}O is larger than the repeatability of the linear calibration parameters obtained during lab measurements (0.05 ‰ for ^{13}C and 0.1 ‰ for ^{18}O). The measured repeatability during the field campaign also exceeds the repeatability of the measurements of the concentration dependency (below 0.15 ‰ for both δ values over a large concentration range) c.f. section 3.1.2. This could be related to the fact, that the δ values of our target standard were out of the calibration range, leading to an enhancement of fluctuations in the calibration parameters."

We added the following footnote*: "In the case of ^{13}C, we excluded the target measurements between 23rd of September till 29th of September, because we obtained a problem with the ^{13}C calibration that lead to a large jump in the delta ^{13}C value of the (very depleted) target standard, but did not occur in the height measurements, probably because they were much closer to the reference delta value.

Fig 5 What is the slope of the linear decay, and what process does it represent?

Author's response: We derived this linear relationship from a first order approximation for the theoretical (and unrealistic) assumption that no mixing occurs in the measurement cell. We added the missing relevant information to the manuscript.
Changes to the manuscript: We added the following paragraph to the discussion.
"The linear fit shown here describes a first order approximation of the theoretical instrument response for the (unrealistic) assumption that there is no mixing in the measurement cell. From this assumption, it can be derived that the δ values would show a dominantly linear decay with the slope $m = (\delta_{new} - \delta_{old})/\tau_{theoretical}$ with the theoretical instrument cell response time $\tau_{theoretical} = p*V/\Phi$, with pressure p, Volume V and flow rate Φ. In our case $\delta^{13}C_{new} - \delta^{13}C_{old} = -29$ ‰ , $\delta^{18}O_{new} - \delta^{18}O_{old} = -36.7$ ‰ and $\tau_{theoretical} = 5.9$ s."

ADDITIONAL COMMENTS IN THE SUPPLEMENT OF RC1 (except typos and grammar mistakes, for which we directly include the referee's correction into the revised manuscript)

P1/9 "field conditions" This is ok, but only if sufficiently exhaustive to replace lab characterization.
Author's response/changes to the manuscript: We changed this and included in general more lab measurements, c.f our comment to p6/20

P1/9 "accuracy of 0.1 ‰ for δ ^{13}C" how can this be smaller than repeated measurements?
Author's response/changes to the manuscript: We changed the way how we quantified accuracy, c.f. our comment to p6/20. We changed the abstract to: "The potential accuracy (defined as the 1σ deviation from the respective linear regression that was used for calibration) was approximately 0.45ppm for c, 0.24‰ for ^{13}C and 0.3‰ for ^{18}O."

P1/14 "became insignificant" Explanation needed (or explicit statement that no explanation found).
Author's response: We are not sure which explanation is needed here a) the correlation itself or b) the change in the correlation coefficient from significant to insignificant? However, in the abstract, we just summarize the observed correlation and shifted the explanation into the discussion, because the discussion of both, a) and b) is a bit long.
Changes to the manuscript: "This correlation became insignificant (p>0.1) for the period after the first snow, indicating a decoupling of δ^{13}C of respiration from recent assimilates."

p2/25 and 25: "isotopologues" isotopocules, or isotopologues and isotomers, or remove bracket.
Author's response: We think it might be a bit confusing for the reader to use the term 'isotopocule' or add 'isotopomer' here, because the latter is irrelevant for CO_2- The Hitran database and many other authors use the the term isotopologue in this context (e.g. Kerstel and Giafriani 2008, Barbour 2011, Ellekoj 2013, Wehr2013 ,Oikawa 2017 , Mohn 2007 ,Affek and Yakir 2014, Vardag 2014). However, we added the information about isotopomers in the footnote and tried to give a clearer definition of isotopoloues.
Changes to the manuscript: "These rotational and vibrational transitions are characteristically different for isotopologues* (defined e.g. by Coplen 2007 as 'molecular species that differ only in isotopic composition'), see e.g. (Varadag 2014, Esler 2000, Kerstel and Giafrani 2008)."
We added this footnote*: "In general this is also true for isotopemers (defined by Coplen 2007 as 'Molecular species having the same number of each isotopic atom […] but differing in their positions.', (e.g. Mohn et al, 2008)".

P2/32 Since this is already cited, check the references and cite them directly.
Author's response/changes to the manuscript: We agree, but this part was removed anyway to shorten the introduction.

P3/9 The classification does not work for this instrument because it combines mid-IR with enhanced effective optical path length.
Author's response: Thanks for pointing that out. We added this information to the manuscript by adding a third category of instruments that combine the approaches of category a) and b)
Changes to the manuscript: We changed this classification in the introduction:
"A slightly modified categorization can be made that differs three classes of laser spectrometers (a) laser based direct absorption spectrometers in the mid infrared where strong absorption features are available (b) laser absorption spectrometers in the near infrared that compensate the weaker absorption in the near infrared by a strongly enhanced effective optical path length and (c) path length enhanced- absorption spectrometers in mid infrared. […] (Guillon et al…). An example for an instrument of class category 2c) is the CCIA-48 (Los Gatos Research. Inc, *San Jose*, USA) that combines a mid-infrared quantum cascade laser with off-axis integrated cavity output spectroscopy (Oikawa et al., 2017)."

P3/20 "direct laser absorption spectrometer" direct absorption, not direct laser.
Author's response/changes to the manuscript: We changed this to "laser based direct absorption spectrometer" throughout the text to be clearer.

P5/33 "purging pump to avoid condensation in the tubes" purging is ok, but why should it avoid condensation (except because of pressure drop).
Author's response/changes to the manuscript: This was misleading, we changed this to: "We purged the main tube to reduce the time the air masses spend in the tubing. To avoid condensation, we heated the valve box (at which we expect a pressure drop) and the adjacent tubing."
P6/5 "the tubes with this small flow rate and" Please explain why condensation is linked to flow rate.
Author's response/changes to the manuscript: We removed this sentence.

P7/13 „linearity calibration" It's not really about linearity but about concentration dependence of the retrieved d values.
Author's response/changes to the manuscript: We changed the terminology, see our answer to your comment to p7/30.

P9/14 Put in relation. Is this better/worse/comparable?
Author's response/changes to the manuscript: We changed this to: "we measured a comparable (slightly better) Allan Deviation below 0.03‰ (c.f. Table 5)."

Multiple comments to chapter 3.1.2
- there is no such thing as "expected uncertainty".
- "measured uncertainty" unsuited terminology
- in the context of the evaluation of a new analyzer you have to make sure that this is not a coincidence. (N=1).
Author's response/changes to the manuscript: We replaced this chapter by a chapter about the evaluation of the calibration strategy. Please see our answer to your comment about p6/20

P10/10 "instrument drift" It would be very interesting to know what the instrument drift is. However, the data shown here is the drift of the retrieved data after all (drift) corrections.
Author's response/changes to the manuscript: Thanks for pointing this out. As the remaining drift after all drift corrections does not seem a meaningful quantity, we removed this part of the data evaluation.

P10/13 "Turnover time" This implies that it is the "turnover" of a perfectly mixed reactor (cell). However, what you then determine are several elements; I suggest calling this "response time".
Author's response/changes to the manuscript: We changed "turnover time" to "response time".

P10/20 „to mixing of gas" please state whether the gas flow is turbulent under the given conditions.
Author's response/changes to the manuscript: The gas flow in all tubes is laminar with Reynolds numbers below 100 for all tubing (6mm and 1/16'). We added this information to chapter 2.5.

P13/3 "As soil respiration has been measured to account for around 80% of total respiration in an old beech forest in below 30 km distance to our field site (Knohl et al., 2008), we further focus on soil respiration and discuss the following hypothesis:" Please check if this is really the line of thought that you want to communicate.
Author's response: We changed this to "Because soil respiration has been measured to account for around 80% of ecosystem respiration in an old beech forest in below 30 km distance to our field site (Knohl et al., 2008), we assume that soil respiration dominates ecosystem respiration and thus we further focus on soil respiration and discuss the following hypothesis:"

P30/table 5 "not necessary; Figure with 1 s and minimum values is sufficient."
Author's response/changes to the manuscript: We would like to keep this table for the readers convenience.

P30/table 6 „not necessary; can be described in one sentence."
Author's response/changes to the manuscript: We removed this table and added the numbers directly into the text: "The analyzer's power consumption of approximately 220W was slightly smaller than the power consumption of the basic infrastructure of the setup that included the pump to purge the 9 inlet tubes and the heated valve box (330W)."

P30/table 7 "not necessary; text and Fig. 4 are sufficient."
Author's response/changes to the manuscript: We would like to keep this table for the readers convenience.

P33/21 "review language of this paragraph"
Author's response/changes to the manuscript: We rewrote this paragraph: "Additionally, the calculation of Keeling-Plot intercepts based on longer timescales increased the number of Keeling-Plot intercepts within reasonable ranges. For Keeling-Plots that were averaged over 2h (5h), a fraction of 97% (99%) of the Keeling Plot intercepts were between -33 and -25‰. Because the range of the Keeling-Plot intercepts should not depend on the chosen timescale, we considered the Keeling-Plot intercepts that were outside of this range as outliers and removed them (also for Keeling-Plot Intercepts that were based on shorter timescales)."