

Author's reply to the referees comments to manuscript AMT-2017-120 - Anonymous referee 2

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 ...

This paper describes the testing and use of the new Delta Ray IRIS CO₂ isotope spectrometer during a three-month field campaign. I have serious concerns about the analytical details, as well as the conclusions regarding interpretations of the field measurements. Overall, I am not convinced that this instrument has been put through the necessary rigorous tests.

Author's response: We thank the anonymous referee for his comments and suggestions. In this response we show additional data of instrument test and more detailed data of the field measurements to answer the referee's questions.

1) The authors conclude in the abstract that "1) the new Delta Ray IRIS with its internal calibration procedure provides an opportunity to precisely and accurately measure c , $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ at field sites" I am concerned with this statement, because the internal calibration procedure in the IRIS is never actually described. How are the absorption spectra used to calculate isotope ratios, and how are these modified based on the calibration? This point appears critical for understanding whether the internal procedure is adequate and/or necessary, or for understanding what other post-hoc calibrations may be needed. This is a critical gap in the paper. One cannot simply assume that the manufacturers of the instrument have worked out the details here. There are instruments that are sold that do not necessarily function as advertised, thus it is necessary to validate every step of the way. I would like to see plots and regressions of raw vs. known values for both $\delta^{13}\text{C}$ and $[\text{CO}_2]$ for a number of different standards spanning a broad range of δ values and mole fractions of CO₂.

Author's response: We added this missing information about the spectral fit, the calibration procedure and about the evaluation of the calibration procedure (including the suggested plots) to the manuscript.

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 CO₂ absorption lines: at 4.3277 μm and 4.3280 μm (both for $^{16}\text{O}^{12}\text{C}^{16}\text{O}$), 4.3283 μm (for $^{16}\text{O}^{13}\text{C}^{16}\text{O}$), and 4.3286 μm (for $^{16}\text{O}^{12}\text{C}^{18}\text{O}$). 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 $^{16}\text{O}^{12}\text{C}^{16}\text{O}$), 4.3283 μm (for $^{16}\text{O}^{13}\text{C}^{16}\text{O}$), and 4.3286 μm (for $^{18}\text{O}^{12}\text{C}^{16}\text{O}$). 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 $^{16}\text{O}^{12}\text{C}^{16}\text{O}$, 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."

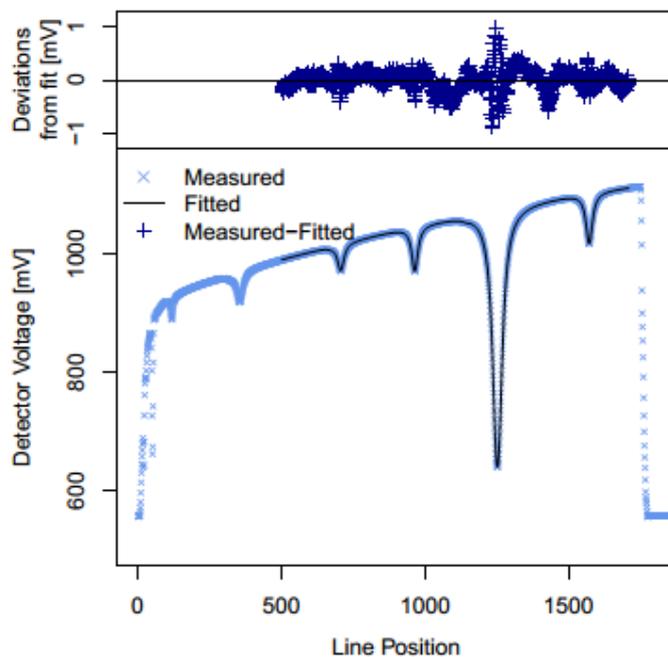


Figure R1 Measured and fitted spectrum, as exported from the instrument's operational software QTEGRA.

We added a chapter about the internal calibration

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 $f_{\text{correct}}(C_{\text{raw}})$ is multiplied with the raw isotopic ratio R (information from the manufacturer, Thermo Fisher Scientific)

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

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 CO₂ standard (‘ambient CO₂’, 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 $f_{\text{correct}}(C_{\text{raw}})$ as determined from the concentration dependency correction, and the slope $m_{\delta\text{scale}}$ derived from the δ scale calibration (information from the manufacturer, Thermo Fisher Scientific).

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

We added a chapter to describe the 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 CO₂ with synthetic air over a CO₂ 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 CO₂ 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 $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ of gases with concentrations ranging from (350 to 450 ppm) and isotopic compositions ranging from -37 to -9.7 ‰ for $\delta^{13}\text{C}$ and from -35 to -5 ‰ for $\delta^{18}\text{O}$. Each of these measurements was performed three times. (Used gas tanks: “ambient”, “bio1”, “bio2”, “PA-tank”, SACO₂ -350, SACO₂ -450, SACO₂ -500, c.f. Table 3.)
 - We performed measurements of two pure CO₂ 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 CO₂ concentrations (350 and 500 ppm). These measurements were repeated every six hours for a period of one week. (Used gas tanks: “ambient”, “bio”, (‘SA-CO₂-350’ and ‘SA-CO₂-500’, 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 ¹³C and 0.4 ‰ for ¹⁸O.

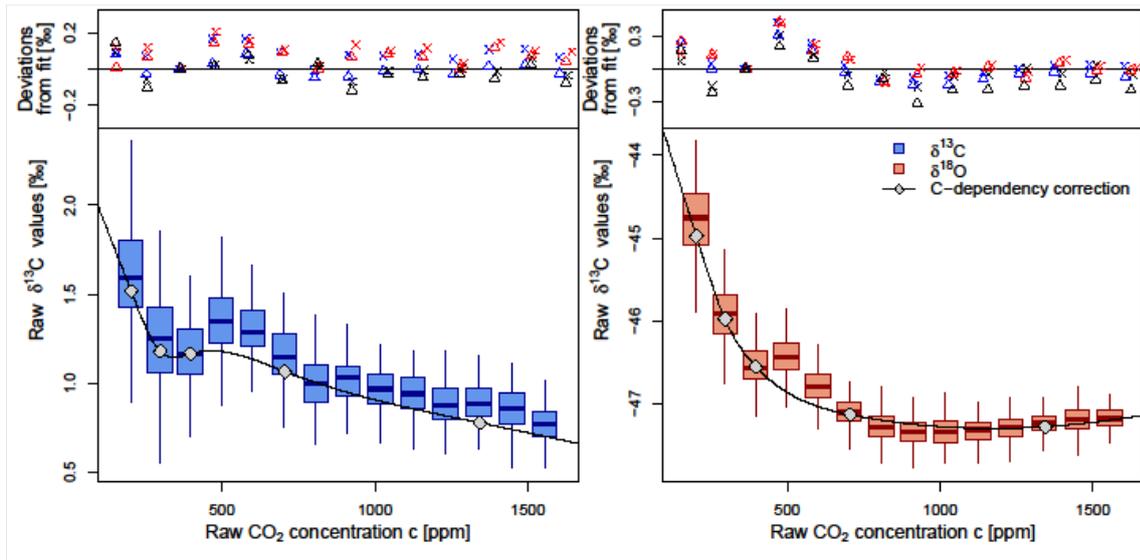


Figure R2 Box whiskers plots showing the nonlinear concentration-dependency of raw δ values for ^{13}C and ^{18}O respectively, here as an example for the CO_2 tank 'ambient'. This measured c -dependency is compared to the respective concentration-dependency correction (black line, with grey symbols 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 $\delta^{13}\text{C}$, and above 0.998 for $\delta^{18}\text{O}$. 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 $\delta^{13}\text{C}$ and 0.3 ‰ for $\delta^{18}\text{O}$. 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.

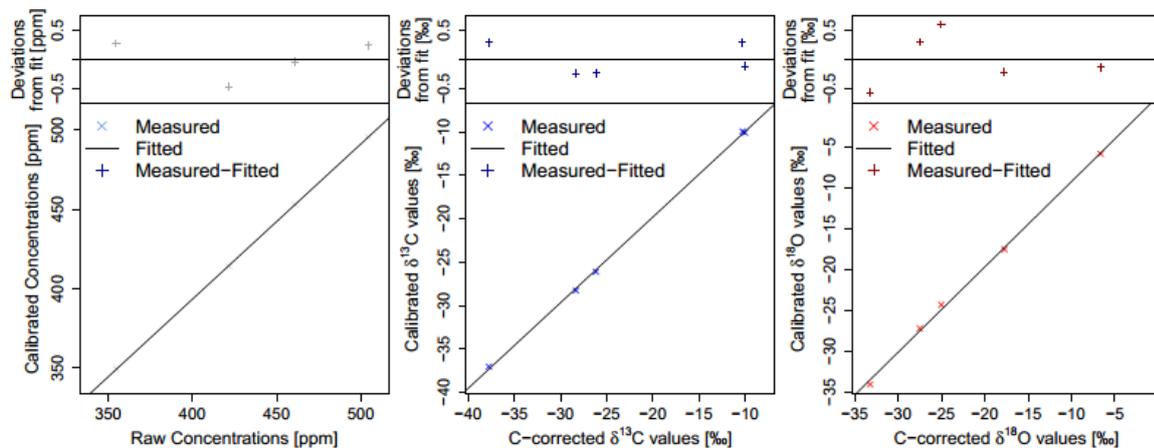


Figure R3 Linear calibrations for concentration (left panel), $\delta^{13}\text{C}$ (middle panel) and $\delta^{18}\text{O}$ (right panel).

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.

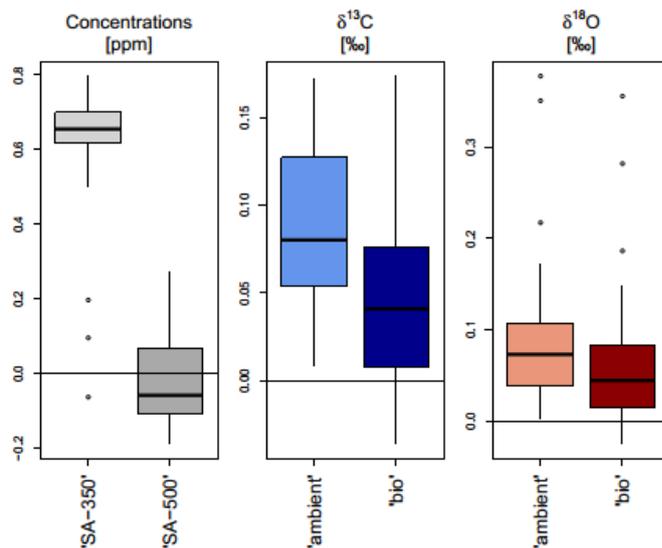


Figure R4 Box whiskers plots for the deviations of calibrated concentrations and δ values from laboratory measurements (at MPI in Jena) for repeated measurements of different calibration tanks (c.f. Table 3 for c and δ values of the gas tanks) over a period of 9 days ($N=36$). Delta values were measured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

Concentrations	tank ,ambient‘		tank ,bio‘	
	$\sigma (\delta^{13}\text{C}- \delta^{13}\text{C}_{\text{tank}})$	$\sigma (\delta^{18}\text{O}- \delta^{13}\text{C}_{\text{tank}})$	$\sigma (\delta^{13}\text{C}- \delta^{13}\text{C}_{\text{tank}})$	$\sigma (\delta^{18}\text{O}- \delta^{18}\text{O}_{\text{tank}})$
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.

2) The authors mention that they used a post-hoc CO₂ concentration calibration, but it is unclear how often the additional standards used for this were measured (once? Halfhourly?) in relation to their check standard. Note that quadratic relationships may give a better fit as employed elsewhere for other absorption-based CO₂ instruments.

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: Concerning the potential nonlinearity, Figure RC2 above evaluates the instrument's linearity and quantifies the deviations from the linear regression, please see also the chapter "Evaluating the calibration strategy" (above). 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."

3) Given that this is a methods paper, it would have been very useful to see tests using a broader range of CO₂ mole fraction and isotope compositions in the range of standards, and to see more standards tested. Without this, we cannot validate the linearity of the instrument both in concentration and isotope space. This is a critical deficit of the paper. Why was the need for a post-hoc $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ calibration not tested or described?

Author's response/changes to the manuscript: We addressed this questions by adding an additional chapter about test measurements to evaluate the calibration strategy. This chapter evaluates the concentration dependency of the δ values over a range of 200 to 1500 ppm and includes measurements of different gas tanks with concentrations ranging from (350 to 450 ppm) and isotopic compositions ranging from -37 to -9.7 ‰ for $\delta^{13}\text{C}$ and from -35 to -5 ‰ for $\delta^{18}\text{O}$. Please see our response to question 1) above.

4) Note that many of the other laser-based isotope instruments achieve much higher precision with frequent (e.g. 20 minute) isotope calibrations in the field. This need appears especially critical here given the large (1 per mil) jumps in $\delta^{13}\text{C}$ values observed in the check standards shown in Figure 4. This suggests that there are some serious stability problems that need to be addressed with more frequent isotope calibration.

Author's response:

Concerning the large jumps in observed δ values: Thanks for pointing this out. We showed Figure 4 mainly to show the repeatability of the instrument, but we agree with you, that the large (1 ‰) jumps in δ values need more discussion. For these two large jumps, we found explanations: The first of these large jumps appeared in ^{13}C after calibration on 23th of September and disappeared after calibration on 29th of September. This jump only occurs in $\delta^{13}\text{C}$ of the target measurement. In particular, this jump was not visible in the $\delta^{13}\text{C}$ in the measurements of the different heights (see figure R5 for the highest inlet as an example). Thus, we conclude that there was a problem with ^{13}C calibration. This problem might be enhanced for δ values that deviate from the 'reference' δ value, in particular for the very depleted target measurement, that was even out of the calibration range.

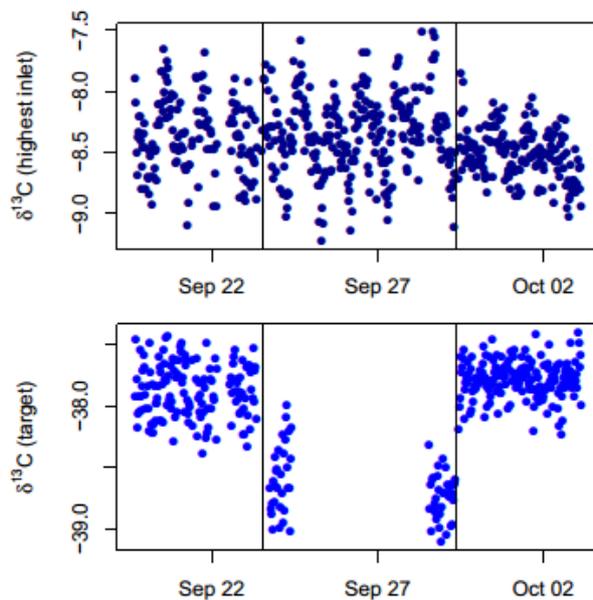


Figure R5 Time series of $\delta^{13}\text{C}$ values for the time period that shows a large jump in $\delta^{13}\text{C}$ for target measurements, but not for the height inlets, shown here as an example for the highest inlet.

The second large jump in the time series of the isotopic composition of the target gas from the 9th until the 16th of October includes the period during which we had a laser alignment problem and the laser needed to be readjusted. After calibration on 16th of October, the measured target gas value jumped back to its value before the 9th of October. We originally wanted to show all data points for completeness, but as we can relate them to a) a problem with one specific $\delta^{13}\text{C}$ calibration that occurs particularly for the very depleted target gas and b) a general laser alignment problem, we think it is more appropriate to remove the corresponding data points from further analysis. In case of the laser alignment problem we also removed the corresponding time series of the height measurements.

Concerning precision, here we quantify precision by measuring the Allan deviation of the uncalibrated δ values, like many other authors (e.g. Sturm et al 2013, van Geldern et al 2014). For this comparison, Table 2 gives an overview about the precision of the δ value measurements of different laser-based and broadband light source-based instruments. In case you refer to what we called 'long-term-stability' in the original manuscript, but call 'repeatability' in the revised manuscript, this is discussed in the chapter 'Evaluation of the calibration strategy' (See our comment

above).

Changes to the manuscript: We changed Figure 4 and the respective description to “Figure 4 Time series of target gas measurements excluding periods that show problems with target gas flow, calibration and a laser alignment problem.”

We added the following chapter: “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.

5) With respect to the second major conclusion of the abstract, “2) even short snow or frost events could have strong effects on the isotopic composition of CO_2 exchange at ecosystem scale” this finding is not new, but also not very well supported by the data (e.g. Figs 7 and 8. There are now several multi-year datasets of canopy CO_2 and $\delta^{13}\text{C}$ profiles in temperate ecosystems that have shown similar patterns.

Author’s response:

Concerning conclusion 2) Here we summarize the results concerning ^{13}C as well as ^{18}O that are discussed in detail in the results section. This statement does not only refer to Figures 7 and 8. The parts of the manuscript that support this conclusion are in particular figure 9 (top panel) for ^{13}C and Table 8 for ^{18}O (in addition to figure 7). As we discuss in section 3.2.2, for ^{13}C we do not observe a change in the $\delta^{13}\text{C}$ values, but we find indications, that the processes controlling the ^{13}C of CO_2 exchange shifted. For brevity in the abstract, we tried to stay general, but specified this in the revised manuscript.

Concerning the mentioned multi-year records: We are well aware that there are now several multi-year records of ^{13}C and ^{18}O in CO_2 profiles (e.g. Bowling et al 2002b, Wehr et al 2016, Bowling et al 2003, Shim et al 2013). However, we are not so sure if the ‘similar’ pattern that you talk about show the same change in the time lagged (and negative) correlation between Reco^{13}C and 2-d averaged radiation (not VPD), particularly in the combination with frost events. It would be very interesting for us to see which species (^{13}C or ^{18}O) and which datasets you are referring to in particular and we are happy to include the respective citation

Changes to the manuscript: We specified the abstract “2) even short snow or frost events could have strong effects on the isotopic composition (in particular ^{18}O) of CO_2 exchange at ecosystem scale.”

We added a more comprehensive list of citations, focussing on multi-year record to the introduction (page to page 2 line 12) “The temporal variability of the isotopic composition of respiration for example has been studied on timescales ranging from sub-diurnal (Barbour et al., 2011) to seasonal (Ekblad and Höglberg, 2001; Bowling et al., 2002; Knohl et al., 2005). Further, the isotopic

composition in CO₂ profiles has been studied on several sites over multiple years for ¹³C (e.g. Bowling et al 2002b, Wehr 2016) as well as for ¹⁸O (e.g. Bowling 2003, Shim et al 2013).“

We added a sentence referring to the observed peaks $R_{\text{eco}}^{18\text{O}}$ in the monsoon-dominated woodland observed by Shim et al to the discussion: “Similarly strong peaks in $R_{\text{eco}}^{18\text{O}}$ have been observed in a semi-arid woodland after precipitation in New Mexico (Shim et al 2013), but this study refers to a monsoon dominated ecosystem with comparably large variability in the ¹⁸O and does not focus on the difference of these pulses of snow and rain events.”

6) With respect to the Keeling plot intercepts, no data is shown to actually validate the approach (e.g. plots of $\delta^{13}\text{C}$ and $1/\text{CO}_2$ space), nor summary statistics presented for these regressions. This is another serious deficit given the key methodological issues the authors point out in the Appendix, but do not quantify in the text. I don't think the authors present enough information here to rigorously test the hypotheses proposed in the Results/Discussion section.

Author's response: Below we show the histogram of R^2 .

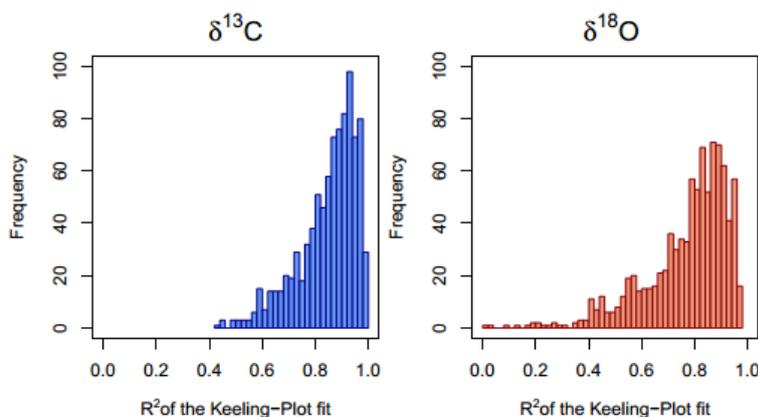


Figure R6 Histograms showing the R^2 values of accepted Keeling Plots based on data that was measured within 90 minutes during nighttime (between 20pm and 4 am).

Changes to the manuscript: We provide summary statistics about the regressions to the text:

“The filtered nighttime Keeling-Plot intercepts based on 30 minutes of data acquisition had R^2 values with a median of 0.87 and 0.81 for ¹³C and ¹⁸O with mean values of 0.85 and 0.77 and standard deviations of 0.1 and 0.16 respectively.”

We added the following Example Keeling-Plots to the supplementary material. We chose Keeling-Plots with R^2 values spanning the range of the respective mean+/- 1 standard deviation.

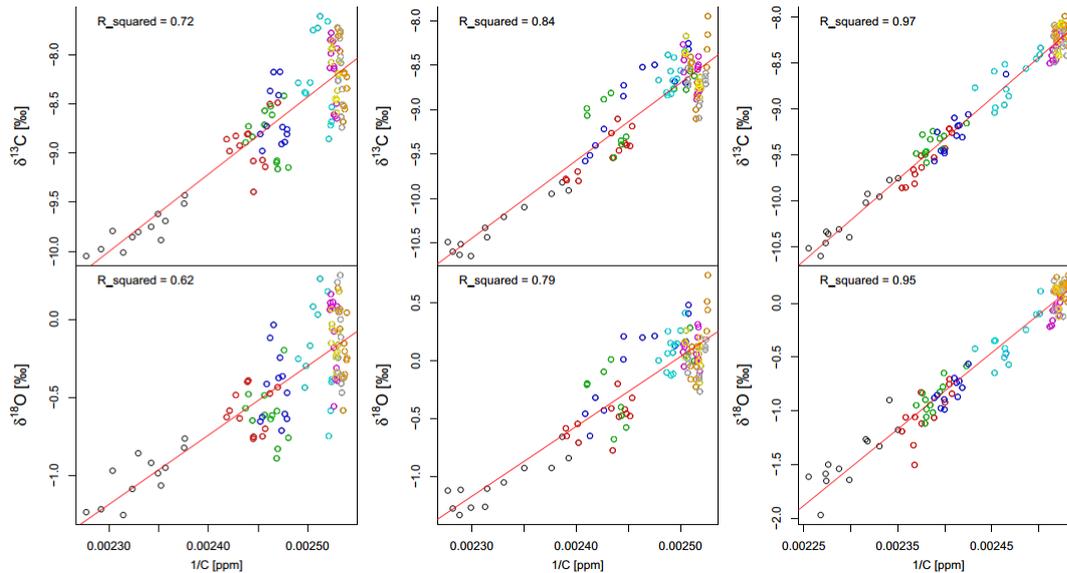


Figure 7 Example nighttime Keeling-Plots with typical R^2 -values (spanning the range of the mean $\pm 1\sigma$). Each Keeling-Plot is based on 90 min input data. Different colors represent different inlet heights.

The value of the CANVEG modeling exercise for the overall study was not terribly apparent to me, nor were the questions that it sought to address.

Author's response: We included the modelling to test the Hypothesis (a) (page 12 lines 1-14) as discussed in particular in lines 22ff. We modified this explanation to make the reason for the inclusion of the model clearer.

Changes to the manuscript: We added an additional sentence to the explanation in section 3.2.2 "Hypothesis (a): The variability of $R^{13}C_{eco}$ can be partly explained by the isotopic composition of recent assimilates $^{13}C_{Ass}$, which is controlled by meteorological drivers during photosynthesis according to the Farquhar model. Thus, the variability of $R^{13}C_{eco}$ is linked to the variability of meteorological drivers of photosynthesis and photosynthetic discrimination with a time lag that is consistent with the time lag between respiration and assimilation. [...We observed a correlation between radiation R_n and $R^{13}C_{eco}$,...] But the correlation itself cannot be directly explained by the Farquhar model of discrimination as radiation influences both, the CO_2 supply (by influencing stomatal conductance) and the CO_2 demand (by influencing assimilation) in the leaf (Farquhar and Sharkey, 1982). In particular, we did not find a significant time lagged positive correlation between $R^{13}C_{eco}$ and VPD, RH or the ratio VPD/PAR (Fig. 8), which could be directly associated with the Farquhar Model and has been found by the above mentioned studies. [this refers to (Ekblad and Högberg, 2001; Bowling et al., 2002; Knohl et al., 2005)] To test if it might be still reasonable to interpret the observed negative correlation of $R^{13}C_{eco}$ with R_n as a time lagged link between $R^{13}C_{eco}$ and isotopic composition of recently assimilated material $^{13}C_{Ass}$ on ecosystem scale, we performed a more complex calculation of $^{13}C_{Ass}$ by using the multilayer model CANVEG. The advantage of CANVEG is that it accounts for the non-linear interactions between air temperature, air humidity, radiation, stomatal conductance and photosynthesis."

To explain this thought earlier, we added/moved the following to the beginning of chapter 2.8:

"To test if the measured variability of the isotopic composition of respiration can be partly explained by the variability of the isotopic composition of recent assimilates, we used the Multi-layer model CANVEG to simulate the isotopic composition of assimilated material during our measurement campaign. In particular, we analyzed the correlation of modeled $^{13}C_{Ass}$ to net radiation R_n , a driver of photosynthesis and photosynthetic discrimination, during our measurement period in autumn 2015. We further compared the resulting relationship between R_n and $^{13}C_{Ass}$ to the observed (time lagged) relationship between R_n and the ^{13}C composition of respiration $R_{eco}^{13}C$, derived from the measured

Keeling-Plots, c.f. section 3.2.2. This analysis was performed to test the hypotheses of a link between δ values in assimilated material and respiration.”

7) More specific comments:

Introduction: there is much excessive detail here that repeats recent reviews, such as the Griffis 2013 paper. Please condense.

Author’s response/changes to the manuscript: We shortened the introduction, in particular p2 line 22 ff.

P1 18: the main constraint is low temporal resolution

Author’s response/changes to the manuscript: Thanks for pointing this out. We added this information to the manuscript.

P4 13: how are these “physically different” air samples if the pump is flowing continuously?

Author’s response/changes to the manuscript: We removed this misleading description.

P8 5: “A possible reason for this resulting deviation is the range of the gas tanks we used for the instrument-internal concentration calibration, that was approximately 300 to 430 ppm” this logic doesn’t make sense to me this is similar to your other standards

Author’s response: We agree that this might not be the reason for the observed problems with concentration calibration, please see our answer to your comment 2).

P8 6: I am having trouble understanding how your “target standard” could be stable without posthoc calibration yet your five other standards were so variable.

Author’s response: This was not the case. We found a need for post concentration calibration because the ‘target’ standard was not stable, please see our answer to your comment 2).

P8 9: “Secondly we set the IRIS analyzer’s internal referencing procedure (described in Sect. 2.7) to 1800 s which corresponds to an Allan variance of 0.03 ‰ for both δ values and 0.01 ppm for CO₂ concentration.” This is unclear to me are you measuring the standards every 1800 s? For how long?

Author’s response: Yes, we measured the reference standard every 30 minutes. We measured it for 80s after the tubes were purged for 60s. We rewrote the chapter about the calibration procedure to be clearer (see our comment above).

Where are these new Allan variance values coming from?

Author’s response: That was a typo. Thanks for finding it!

Figure 4: There are apparently large (1 per mil) jumps in measured “target gas” isotope values at several points. These are disconcerting. Are the data shown in this figure the raw values or the calibrated values? If they are the calibrated values, this suggests that the two-point calibration employed here is inadequate.

Author’s response: The figure you are referring to (figure 4 in the original manuscript) shows calibrated values. Please see our answer to your comment number 4) above and the new chapter about the evaluation of the calibration strategy.