

# Rebuttal 2

*Data treatment and corrections for estimating H<sub>2</sub>O and CO<sub>2</sub> isotope fluxes from high frequency observations.*

Robbert P.J. Moonen, Getachew A. Adnew, Oscar K. Hartogensis, Jordi Vilà -Guerau de Arellano, David J. Bonell Fontas, and Thomas Röckmann

We again thank the referee and the editor for their valuable comments and encouraging the resubmission. Our replies are given in blue, whereas changes to the final manuscript are given in red.

## General comments

### Referee #1

The authors have successfully included the feedback from the first review round and the manuscript has improved in clarity. I also thank the authors for the extensive replies to my comments, which resolved most of the questions I had. Some general questions remain which I will outline below.

In general, I think the manuscript is a valuable contribution to the field of measuring stable isotope fluxes and isofluxes and will help in designing future field campaigns and postprocessing procedures.

## Specific comments

### ***Mole-fraction calibration for H<sub>2</sub>O***

*Thank you for outlining the calibration protocol in Table A1 which resolved many questions I had. I admit though that I still don't fully understand what was done and why a leak would explain the cross dependency between mole fraction dependence and standard water isotopic composition considering that the leak/ambient isotopic composition before and after the campaign when the calibrations were done was probably not identical in itself. I also still think that a visual representation of the calibration coefficients in Figure 3 is not very helpful and should at least be accompanied by the final values that you used in a table format (possibly together with the CO<sub>2</sub> coefficients in Table 2) and a graph showing the fitted quadratic curves to the datapoints from the calibration (in the Appendix). Could you also outline what mole fraction level you used as reference for the correction function? A priori I would assume that a fit to "3 humidity levels with 6 and 4 datapoints each" (did I understand correctly?) in the range 7000-24000ppmv is not very robust and might be the reason for differences in calibration coefficients before and after the campaign rather than a leak. Similarly, you are attributing the differences in span calibration (0.4‰ for 18O is rather high) to drift alone but this could also just be due to noise in your calibration pulses and should be phrased more cautiously in my opinion. As you say though, the accuracy of your measurements is arguably less important than the precision. That being said, I wonder if the mole-fraction calibration is indeed as crucial as you say since the fluctuations within the 30min averaging window will be small, even if the mole fraction sensitivity of your instrument seems very high if I read off the coefficient values correctly from Figure 3. And lastly, I don't understand the uncertainties of the standard waters you give in Table 1 and A1 in terms of significant figures/decimal places and origin of these values. Does the standard error come from a separate analysis of these standard waters?*

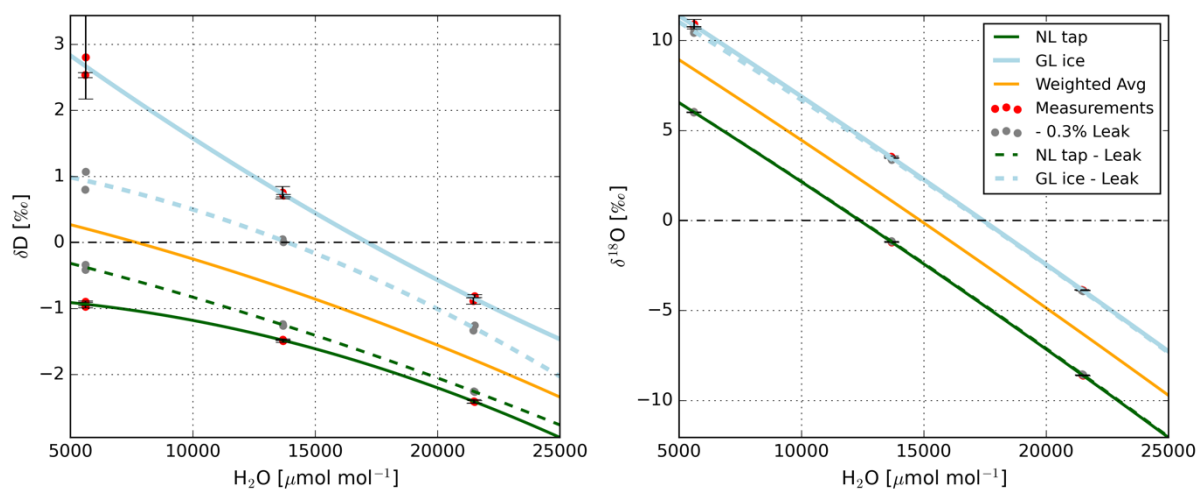
We thought it important to specify our mole fraction calibration approach and correction procedure as we observed large differences in the 30-minute isofluxes derived with or without applying mole fraction dependence corrections. This difference arises from artificial fluctuations in the delta-value signal which correlate with  $w'$  as the underlying mole fraction fluctuations correlate with  $w'$ . As the reviewer indicates, our instrument does have a large mole fraction dependence, resulting in a large artificial flux when not using dependency corrections. We agree that a rough dependency estimate would already result in reasonable isofluxes, and that the correction for a possible leak is only a small improvement.

As the reviewer observes, atmospheric isotopic compositions were likely different before and after the campaign. However, natural atmospheric variations are much smaller compared to the differences

between the standards we used (Table 1). The approximate atmospheric background composition is also reflected in the rough 3/1 ratio between the two standards to reflect the atmospheric isotopic composition.

We created the figure below, which we expect will clear up the mole fraction calibrations and leak correction we applied. It is an example where the calibration coefficients are fitted to the references measured after the campaign. Note that the measurements are offset to be approximately zero on the y-axis to be able to display both standards in one figure. The solid green and blue lines indicate the fitted dependencies of each standard. For  $\delta D$ , both lines depart from linear behaviour at low mole fractions. The GL ice standard, which is naturally depleted, tends towards higher  $\delta D$  values at low mole fractions, while the NL tap water standards, which is naturally enriched (see Table 1 in the manuscript), is moved towards lower  $\delta D$  values at low mole fractions. To show that this effect is likely caused by a leak of ambient air, we simulated a 0.3% leak in the inlet during the calibrations. We observe that at low mole fractions, the impact of the simulated leak on the delta values is larger due to the low water content in the calibration gas stream. More importantly, the leak corrected data points and calibration lines are consistent with the Weighted Avg mole fraction dependence we derived, which indirectly corroborates that our correction approach makes sense. Note that the effect of a leak in  $\delta^{18}O$ - $H_2O$  is much less pronounced due to the larger mole fraction dependence in  $\delta^{18}O$ , combined with a smaller absolute difference between the delta values of the calibration standards and ambient air.

We double checked if the same leak could have affected the span calibration measurements. In that case, a scale contraction should be observed. In addition, this scale contraction should be larger during the calibrations after the campaign, as a larger leak was present there (see Figure 3). We observe neither of these effects in our calibration data. We recall that a leak of ambient air during a span calibration affects the entire calibration scale, except at the isotopic composition of said ambient air, which is exactly what we measured during the campaign.



The error bars of the measurements plotted in the figure indicate the standard error of the mean during each 20min measurement plateau, after averaging the high frequency data in 1-minute bins. While our calibration approach is simple, the error bars indicate the method is robust. In line with that, repeat measurements were mostly indistinguishable from the initial measurement.

The standard errors of the standards in Table 1 were also derived as the standard error of the mean during the crosscalibration procedure. However, we did not bin the data to 1-minute averages in the previous version of the manuscript, causing the  $\sqrt{n}$  term to explode due to the frequency of the measurements and thereby generating unrealistically small uncertainties. In addition, we realize that the deuterium error given in table 1 was 10x too small due to the accidental addition of a decimal 0, causing it to be extra unrealistic. We now applied 1-minute binning which we assume to result in independent samples and adjusted the errors accordingly.

Finally, Table 2 (below) was updated to include the weighted average mole fraction calibration coefficients.

**Table 2.** The dependence of isotopic composition measurements ( $\delta$ -values) on the mole fraction of the respective molecule, H<sub>2</sub>O or CO<sub>2</sub>, expressed in  $\mu\text{mol mol}^{-1}$ . The CO<sub>2</sub> dependencies were derived after the campaign while the H<sub>2</sub>O dependencies represent the weighted average coefficients during the campaign (see Fig. 3).

Mole fraction dependence	$\delta^{18}\text{O} - \text{CO}_2$	$\delta^{13}\text{C} - \text{CO}_2$	$\delta\text{D} - \text{H}_2\text{O}$	$\delta^{18}\text{O} - \text{H}_2\text{O}$
Linear	$-1.27\text{e}^{-2}$	$-3.64\text{e}^{-2}$	$-7.71\text{e}^{-5}$	$-8.54\text{e}^{-5}$
Quadratic	$1.80\text{e}^{-6}$	$1.16\text{e}^{-5}$	$-1.78\text{e}^{-9}$	$-2.60\text{e}^{-9}$

Besides adding the figure to the appendix, updating table 2, and including in-text references, we adjusted the following:

L. 202; Differences in the span calibrations performed before and after the campaign were +/- 0.4‰ for  $\delta^{18}\text{O}$  and +/- 0.3‰ for  $\delta\text{D}$  at atmospheric isotopic compositions. We suspect that instrument drift is the cause given the small uncertainty in the (re)calibrations (0.02‰ for both d13C and d18O). When ignoring drift inducing events like instrument rebooting and transportation, the interpolated drift during the measurement period is still below 0.1‰ for both species. ~~Drift during the measurement period was below 0.1‰ for both species when assuming drift to behave linearly. Likely, the drift during field measurements was even smaller given that instrument rebooting and transportation cause more drift than continuous operation.~~

Adjusted errors in table 1 and appendix table A1:

**Table 1.** The isotopic compositions of the H<sub>2</sub>O calibration standards including the standard error of the 1-min binned data during cross-calibrations of the references, and the average atmospheric isotopic composition during the measurement period including its approximate range

	$\delta^{18}\text{O}$	$\delta\text{D}$
NL tap water	$-6.98 \pm 0.02\text{‰}$	$-47.12 \pm 0.04\text{‰}$
GL icecore	$-30.80 \pm 0.02\text{‰}$	$-240.86 \pm 0.04\text{‰}$
LIAISE atm	$-13.4 \pm 1\text{‰}$	$-94 \pm 8\text{‰}$

L. 213; We suggest that the similarity in the Weighted Avg (yellow) calibration coefficients "Before" and "After" the campaign in Fig. 3 is no coincidence, but a feature of an ambient air leak of variable magnitude. In Appendix A2, we give an example where the calibration coefficients in Figure 3 are fitted to the measurements. We also show simulations where we assume a small (counter) leak of ambient air (of 0.3%), which is able to explain the observed mole fraction dependencies. An instrument related cross dependency of the isotopic composition on the mole fraction dependence, as described by Weng et al. (2020), is not expected to average out like this. Ultimately, the dependencies were eliminated using 15000  $\mu\text{mol mol}^{-1}$  as the reference H<sub>2</sub>O mole fraction.

#### **CF factor analysis**

*I still think it would be interesting to see or discuss the CF factor values. The statement of Line 415 that the authors were willing to include (comparison between CF for dD and d18O) is very limited and does not discuss the "difficulty" of correction values that are VERY high for either isotope species. I think this aspect of correcting a measured signal by a very high factor (at least for H<sub>2</sub>O isotopes) should be discussed in the section 6 as a general problem of measuring isotope or delta fluxes. It could be discussed also in the context of the intriguing thought of precision vs. sampling frequency in new instrument development (L. 489-492).*

*Additionally, I think it is worth to include a discussion of the CF factors for CO<sub>2</sub> that are  $\approx 1$  and thus in line with previous studies which I think is supporting your new correction technique. I therefore recommend to include a short CO<sub>2</sub> CF factor discussion in section 6.*

Thanks for these useful suggestions which we implemented in the text. The second point was already partially mentioned from L. 479 to L. 483, but that we extended this argument.

We made the following changes to the text;

L. 489; However, isotopic ecosystem flux partitioning is impacted by our findings.

Be aware that in contrast to most correction approaches, where the order of magnitude of the correction is 10%, the spectral scaling approach leads to correction factors of the order of 100%. This means that most of the signal in the d-flux does not originate from the actual measurements of the delta-flux, but from the correction method, which is not desirable. Consequently, the errors of the corrected delta fluxes (see Sec. 4.4) can best be based on the uncertainty in fit of the correction, and not on the uncertainty of the measured delta-flux. As long as errors are properly quantified by propagating this fit error to the flux, we do believe that using the spectral correction approach is valuable for deriving delta-fluxes of the correct magnitude.

The implications of the spectral scaling principle are broader than using it to find adequate corrections.

L. 479; To prove this hypothesis we can investigate the cospectra of  $\delta$ -fluxes and mole fraction fluxes measured with a setup in which all turbulent scales are well represented. In Fig. 6 we show that the cospectral density for the  $\delta_{13}\text{C}$  and  $\text{CO}_2$  observations is generally very similar, supporting the validity of our hypothesis. While comparing the corrected and non-corrected d-fluxes of both  $\delta_{13}\text{C}$  and  $\delta_{18}\text{O}$  in Fig. 8 we see the same pattern. The corrected d-fluxes are mostly similar to the non-corrected d-fluxes, even though all high frequency fluctuations were eliminated. Note that the noise and large uncertainties that are present in the corrected delta fluxes can be partially attributed to the instrument drift in the LS- $\text{CO}_2$  on long timescales and a relatively low signal to noise ratio in the d18O-fluxes.

While the principle seems to hold, the  $\delta_{13}\text{C}$  signal in Fig. 6 is impacted by instrument drift on long timescales and a relatively low signal to noise ratio. More precise experimental measurements of the net ecosystem exchange and net isotopic exchange of trace gasses affected by various fractionation processes should increase confidence in this hypothesis.

**Data availability statement:**

*I do support the general community effort of making data publicly available with a doi upon publication in line with the FAIR principles:*

[https://www.atmospheric-measurement-techniques.net/policies/data\\_policy.html](https://www.atmospheric-measurement-techniques.net/policies/data_policy.html)

[https://www.atmospheric-measurementtechniques.net/policies/data\\_policy.html#data\\_availability](https://www.atmospheric-measurementtechniques.net/policies/data_policy.html#data_availability)

We now made the 30-minute flux data available online using *figshare*, using the following DOI: 10.6084/m9.figshare.23828514. We added the DOI to the data availability statement, including our contact details if other researchers want to use our high frequency data.

## Technical corrections

L. 16-18: <i>It sounds as if you have used your results to validate models in this manuscript. Maybe rephrase.</i>	Incorporated
Only after such corrections and verifications are made, ecosystem scale fluxes can be partitioned using isotopic fluxes as constraints, which in turn allows for conceptual land-atmosphere exchange models to be validated.	
L. 21: <i>missing space after CO2</i>	Incorporated
L. 56 & 60: <i>"<math>\delta</math>-flux" term used before its introduced/defined/explained in section 2.2</i>	incorporated
L. 76: <i>change "isotope fractionation" to "isotopic fractionation"</i>	Incorporated
L. 129: <i>correct "we included figure is included"</i>	incorporated
L. 154: <i>correct "and an scroll additional scroll pump"</i>	incorporated
L. 164: <i>repetition of "which"</i>	incorporated
L. 190: <i>something missing: "the calibration of the isotopic against a reference standard"</i>	incorporated

<p>Section 4.2, L. 229: I think you should include a short explanation here why you chose to time shift 10minutes instead of 30 minutes.</p> <p>After the two time series are coarsely aligned using known clock offsets, the longer time series was divided into short data intervals. We choose to use 10-minute intervals instead of the 30-minute intervals used for flux calculations to increase sensitivity to data series which drift fast and irregularly with respect to each other. One of the data series is subsequently cropped by a minute on each side which makes it possible to shift it in time with respect to the other data series within this (two) minute window.</p>	incorporated
L. 287: repetition of “that”	incorporated
L. 301, 452, 456, 457: I think the term “isotopic exchange in the inlet line” is misleading. I suggest to use “isotopic retention” or, as you use in Line. 562, “inlet line attenuation” instead.	incorporated
L. 321, L. 329: How much of the data (30min windows) did you eventually filter/exclude from the analysis? This is interesting for the reader to know.	incorporated
After these filtering steps, 81% of the 30-min intervals of co2 isotope fluxes, and 80% of the 30-min intervals of h2o isotope fluxes were still complete.	
Figure 7 caption, L. 342, L. 346, Figure A3: replace “R2” with “r”	incorporated
L. 343: I don’t see why a linear interpolation is the “likely pattern” during the night given that the variability in periods of similar length during day time is much more variable than linear. Since you exclude night time periods from your analysis (L. 353) I recommend to not show linear interpolations in Figure 7 and remove the “likely pattern” statement.	incorporated
Updated Fig 7 to remove night-time interpolation shown in dashed green, after which the likely pattern remark about the dashed green line is correct.	
L. 383: delete “=”	incorporated
L. 493: Replace “still” with “however”	incorporated
L. 495: include here a statement on the choice of LPF frequency	incorporated
L. 547: delete “in”	incorporated
<p>Section 7: I would also include best practice recommendations not only for the technical set-up but also for the post-processing procedure since this is what you are newly proposing in this manuscript (LPF, averaging time, clock synchronization...)</p> <p>We extended the best practice recommendations to include post-processing procedures:</p> <p>Additionally, we provide the following best practice recommendations for post-processing the isotope flux measurements.</p> <ul style="list-style-type: none"> <li>■ Apply outlier filtering to the detrended scalar data before flux calculations, especially for the measured isotopic compositions. This prevents errors in the flux estimates due to fluctuations that were not induced by turbulence. We used an Inter-Quartile Range (IQR) filter for outlier filtering, as described in Sec. 4.4.</li> <li>■ Time alignment of the high frequency isotope analyser(s) and the high frequency anemometer measurements is a prerequisite for calculating reliable iso-fluxes. The alignment strategy we used is described in Sec. 4.2 and uses the mole fraction signals of the target molecule, measured by both the OPGA and high frequency isotope analyser, as a reference. It corrects for both the effects of instrument clock drift and inlet delays.</li> </ul>	incorporated

<ul style="list-style-type: none"> <li>■ Test if there is a lag between the mole fraction and <math>\delta</math>-value signals of the isotope analyser, especially when measuring H<sub>2</sub>O isotope fluxes. This can also be done using the alignment strategy described in Sec. 4.2. If the lag is not zero, isotopic inlet attenuation likely occurred, and exchange spectra are probably affected.</li> <li>■ Compare the cospectra of the net exchange flux with the cospectra of the isotope exchange flux to identify if high frequency isotope flux signal was missed, as is shown in Fig. 5. If flux signal was missed, the spectral scaling method can be used to correct for the missing high frequency signal.</li> <li>■ Quantify the uncertainty inherent to the correction method in case a major contribution to the <math>\delta</math>-fluxes comes from the spectral scaling correction. In Sec. 4.4 we indicate how this uncertainty can be derived and propagates to the <math>\delta</math> fluxes and isotopic source compositions.</li> </ul>	
<p><i>L. 567: Consider moving the sentence “The attenuation of the isotopic signal...” to Line 564 after “reported previously.”</i></p>	<p>incorporated</p>

## Changes from authors

In figure 3 the units of linear and quadratic terms were swapped in the old manuscript. This has now been corrected.