Review for „Data treatment and corrections for estimating H2O and CO2 isotope fluxes from high-frequency observations" by Robbert P.J. Moonen et al.

General comments:

The authors present H2O and CO2 flux data from a 6-day long field campaign in North Eastern Spain above a crop field with special focus on measuring the isotopic signal of the fluxes with the aim to gain information about ecosystem scale flux partitioning. The authors discuss the feasibility and complications arising from the combination of a high frequency eddy-covariance (EC) system with slower laser spectrometer measurements in their set-up. By addressing the isotopic composition of the fluxes, the authors tackle an underexplored way to separate the different contributors to an "easily measurable" net flux signal which are essential to validate earth system models and to identify potential model bias sources. The general reason for the current lag of establishing this method in the wider agricultural, ecological, and hydrological scientific community is the lack of high-frequency isotope measurements and the consequential need for post-processing methods to correct for the signal loss in the high-frequency domain. Adding on to a few already existing published methods, the authors present a spectral post-processing method that uses only the information stored in the lowest frequencies to correct for the loss of information in the high frequencies. This results in very high correction factors. The authors also mention some potential general set-up errors that have led to sub-optimal data quality during their campaign. The manuscript is well written and the presented data is nicely visualized. A data availability statement is missing.

With their paper the authors contribute to developing a state-of-the art method for measuring isotope fluxes which is timely and interesting for scientists from different communities. However, I am missing a general discussion on the limitations of the presented method and an analysis of a performance difference to previously published methods, an analysis on the correction factors including magnitude and dependency on environmental conditions, an evaluation of the associated uncertainty, a recommendation for "best practices" when measuring isofluxes and the mentioned but never shown supporting meteorological data; all of which would help other scientists to decide whether to use this method for their campaigns. Therefore, further analysis is needed which I will detail in the specific comments and list the few technical corrections.

Specific comments:

L. 12-13: The discussed "offset" in isotope and humidity measurements is commonly referred to as "memory" in isotopic studies, which is stronger for H_2O than CO_2 measurements due to the structure of the molecules. It is also more prominent for D than for ^{18}O and has to be considered when designing the set-up. Please discuss what could have been done better to minimize the memory effects as much as possible (no/different filter, shorter tubes, less surface)

L. 17: As mentioned above, for stating that the method works "reliably" I am missing an uncertainty evaluation as well as a CF analysis and/or a different quality control procedure.

L. 26: Are the presented data actually interpretable as "ecosystem scale flux measurements"? More details on fetch, and the field site are needed to evaluate this.

L. 37&38: Please describe what "framework" you are referring to and it is not clear at this point which parameters are needed for measuring the "turbulent iso-flux" (what does "iso-flux" mean here)

L. 80: In section 2.2 or in the Introduction you should add more literature on the different methods used for measuring iso-fluxes and discuss their limitations and give reasons why you chose your set-up. Literature missing: Yakir & Wang 1996, Keeling 1958, Welp 2008, Wei 2015, and specifically Good et al. 2012 and 2014.

L. 110: For a general understanding of the environmental conditions during the field campaign and the discussion of the limitations of the methodology it is essential that you show the meteorological data, if not in the main text you can decide to do so in the appendix. As a minimum I would like to see the magnitude of surface fluxes, the humidity and CO_2 level, the atmospheric stability and a wind rose. All these parameters influence the measurements and I assume, also the data quality.

L. 125: Setup: Have you performed a dedicated lag time test while in operation to see how long it takes from the inlet until you see a signal in the spectrometers that is purley due to the pump rate end tube length. A detailed setup scheme or diagram would also be of help to identify the reason for the comparably prominent memory effects. How have you tested your setup for leaks? At a pump rate of 30 l/min the risk for leaks is high and could also explain why EC mole fraction and spectrometer data did not always align.

L. 130: A close up picture of the actual inlet setup would be beneficial for understanding the location of the filter and the inlet compared to the EC system.

L. 132: specify inner diameter and material of inlet tubing

L. 133: Was the air flow rate of 30L/min measured or just estimated? Specify where the lines were separated. A diagram might help here.

L. 161: Data treatment: The processing steps of the isotope fluxes are detailed in the manuscript but there is no mentioning of corrections, problems, data quality or similar of the original EC IRGASON data which in itself requires careful consideration. See e.g. Mauder et al. 2006. What software was used to compute the net humidity and CO_2 fluxes, how did you downsample the original 20Hz EC wind data to match to the spectrometer dataset?, What despiking technique was used?, Was data excluded due to unfavourable wind conditions?, How did you rotate the wind data?, What percentage of the full EC dataset can be regarded as high-quality data?

L. 162: Calibrations: There is essential information missing in the description of the calibration procedure. Specifically, you should detail what criteria were used to generate the dataset used for the calibration: In what humidity/CO_2 range were you calibrating, how long were the intervals averaged for the individual calibration pulses, what was the maximum accepted mole fraction fluctuation within a pulse, how many datapoints were used for the calibrations, what is the uncertainty on these pulses, what was the uncertainty associated with the standards you used… (can go in appendix)

L. 170-173: The "however" is irritating in this sentence structure. Also, the humidity-dependency or mole fraction dependency is critical especially at low humidity levels. Therefore, it is essential that you show the humidity level during your field campaign.

Table 1 & 2: Consider merging the two tables and importantly giving all the information for both CO_2 and H_2O calibrations, i.e. standard values used, associated uncertainty, differences in calibrations before and after campaign

L. 182 & L. 189,190: What makes you think the cross-dependency between mole fraction dependency and isotopic signal is NOT what Weng et al. 2020 observed but a leak in the system?

Figure 3: Consider taking out this figure and adding the information in the Table. It does not contain a lot of information but uses a lot of space. Instead, presenting the fit to the mole fraction data and the residuals of that fit (before and after the campaign) contains more information.

L. 198: Consider restructuring the manuscript in a way so that both "time shift corrections" (one due to clock offset and one due to memory effects) are described in the same paragraph since the same (?) method was applied to identify the time lags and it is otherwise a bit confusing to introduce the second time shift correction step later in the results section.

L. 201: When correlating two timeseries, one usually calculates the correlation coefficient r [-1,1] while R2 is a statistical parameter within linear regression that defines the goodness-of-fit. If you calculated the correlation coefficient, the adequate symbol is "r".

L. 202: "after the two timeseries are coarsely aligned". Please specify which observed atmospheric parameters you have correlated to find the specific lag times.

L. 206: Generally, how have you sub sampled? Averaged or sliced?

L. 212: Why did you choose to correlated and shift 10min data sections when you later compute the fluxes for 30min averaging periods? Does that lead to inconsistencies?

L. 198-214 & Figure 4: The whole section including Figure 4 on time shift corrections is too detailed. I assume that "lag-time correlations" is an established enough method to identify time offsets that does not require such in-detail description. Instead of showing the concept of a lag-time correlation in Figure 4, you could show the identified lag times over the period of the measurement campaign, which would visualize the "hour of day" dependency you later refer to and the linear drift.

L.224: The w timeseries for the cospectra with the slower sensors was downsampled for this. Mention this specifically and detail how?

Fig. 5: Consider introducing subplot labels instead of "left" and "right".

L. 228: As said earlier, also the OPGA needs corrections due to path averaging, detrending, wind rotations..

L. 236 ff: I encourage the authors to separately outline the different spectral correction methods used previously in the literature that they cite here and how their approach is different.

L. 243: I encourage the authors to show the figure Appendix A3 in the main manuscript next to the Figure 6 since these figures visualize the methodology explained in this manuscript. I therefore think A3 is significantly more relevant in the main article compared to e.g. the time lag figure

L. 246: What is the highest frequency your system can resolve for CO2 and H2O respectively? Is this different for different isotopologues?

L. 255: the phrase "instead of using the measured d18O covariance" is confusing here since you ARE using the measured d18O covariance for the fitting procedure. Reword please.

L. 265: How was the fitting done? By minimizing what function?

L. 258: Considering that the atmospheric state influences which are the dominating eddy sizes one could think that a varying LPF frequency is beneficial to increase the

"fittable" part of the spectra whenever the atmospheric conditions are favourable for the system.

From a very practical approach, how many datapoints were available for fitting with this LPF and an integration time of 30min?

Figure 6: Please add a line at the LPF frequency to visualize in which range you are fitting and please add the unscaled wH_2O co-spectrum for comparison

Eq. 8 and Line 269: For supporting your statement that the correction is "physically sound" you should present analysis and statistics on the computed correction factors and identify potential physical reasons for variations in the CF magnitude. Further, as explained earlier, an error or uncertainty estimate is needed.

L. 273: How did the CF between d18O and dD for H_2O compare? Were there strong differences? Also, judging from Fig 8, panel 3, the CF for d18O in CO_2 is often >1. Show analysis of CF for both CO_2 and H_2O systems

L. 283: How did you calculate the time lag between isotopes and mole fraction timeseries? Was there a difference between the different isotopes? This would directly influences your cospectra calculations

L. 283 & L. 290: Can you give a reason for the diurnal cycle pattern in the time lag / memory? Does it have to do with the absolute humidity level? Why is lag smallest when temperatures are high?

L. 295: But did you correct for these found offsets or not?

L. 298: What is the reason you were not able to match the timeseries during the night? Too low variability in the signal? Too high lag times? What was the humidity level and the atmospheric stratification during the night times?

Fig. 8:

    -Panel labels are missing,

    -I notice some few positive Delta d13C values in panel D. Do you trust these values?

    -What quality control screening have you implemented for the CO_2 isotope dataset?

    -Top panel: Could you show agreement between Aerodyne and OPGA as scatter plot and give correlation coefficient? What software did you use to calculate the net fluxes shown in this panel?

    - panel 2 and 3: since the re-scaled data points are your final data product, consider swapping the markers for raw and re-scaled, since the light gray crosses are very subtle

    - What is the uncertainty on the data you present here?

    - There is significantly more noise visible on the signal in panel 3 compared to 2, and scaling factor and thus CF seem to be >1.

L. 302: So you did spectrally correct the CO_2 isotope fluxes? This sentence is in disagreement with L. 309 and L. 337? Please clarify.

L 323-327: Have you analyzed a sample of the irrigation water for its isotopic composition?

L 236: How do the results form the H_2O isotopes support this hypothesis? If the difference in Irrigation vs Precipitation isotopic composition is indeed the reason, you would expect to see a prominent difference in the ET isotopic composition after the Precipitation event compared to other days.

Fig 9:

- Top panel: a scatter plot between OPGA and Picarro LvE would be nice to see.
- Top panel: Do you have an explanation for the "W" shape of the OPGA signal when looking at one day from noon-noon?
- Also L. 345-348: Which formula/method was used to calculate OPGA scaling? Could you compare the performance of the different methods and analyze differences and explain them with meteorological conditions – the bias seems to be quite significant → maybe do this in discussion

L. 360: The paper would benefit from a clear best-practice recommendation from the authors concerning the system set-up for future field campaigns to allow for more widespread implementation.

L. 370-373: I would suggest to either show the results of these phase spectra or remove the sentences relating to it.

L 375: To judge "reliability and accuracy" an uncertainty or error analysis is missing.

L. 387: I encourage the authors to expand on the discussion between the two different correction methods, present detailed analysis and discuss causes for the differences.

L 405: In other words, you assume that turbulence is not fractionating.

L. 413: The authors have to show an analysis of these correction factors in magnitude, variability and timing and discuss possible explanations for it.

L. 417-425: This is an interesting thought but I would assume the performance of this method is highly dependent on atmospheric stability and the nature of the turbulent transport (i.e. better suitable for unstable atmospheric stratification). Please elaborate on the limitations of the presented method concerning different atmospheric and general field site conditions. If the atmospheric stability requires shorter flux integration periods (very stable atmospheres might require integration times of a few minutes only) and using the constant LPF, there could be cases where essentially no data points are available for the fitting procedure of your presented method.

Additionally: Could you somehow manipulate an existing timeseries to represent a timelag and smoothing impacted timeseries and therefore show the feasibility of your method?, i.e. no signal loss even if you only use frequencies up to the LPF.

Section 6.1: I find the presentation on the quadrant analyses very interesting and would encourage an additional figure (maybe included in Figure 10) with a complementary time window (e.g. night time) to see the differences in the iso flux behavior and the added information of such an analysis technique.

L. 458: I would argue that a time lag correlation is no new concept, please remove

Conclusions section: The whole section is missing impact and the highlighting of why this method is superior to other methods. Further, an uncertainty statement and a best practice recommendation should be given for any reader that would like to implement a similar technique in future field studies.

Figure A1: instead of showing the fitting lines color coded it would be better to show the fitting parameters as timeseries to see drift and/or variability. Details on how the data points that are the basis for these fits were generated are missing. (accepted variability, length of calibration pulses,…)

Figure A2: From line 337 I got the impression that the CO_2 flux data did not need additional time lag corrections? How does this compare to a lag analysis of d13C – CO_2 mole data?

Technical corrections:

L. 10: rewrite ".. analyse the measured values of central ∂-flux variable". What does central mean here?

L. 19: ET and NEE need to be introduced first

L. 44 and other places: The "Wahl et al." reference is missing details. Update reference to the published version https://doi.org/10.1029/2020JD034400

L. 44: Delete the Griffis 2013 citation here.

Eq. 1.: Usually, the R is defined with the isotopologue ratio instead of the isotope ratio (Mook 2000) which is also what the laser spectrometer instruments measure.

L. 66: Evaporation should not be used as exemplary process to explain "equilibrium fractionation" since it is NOT an equilibrium process alone and kinetic fractionation contributes to total fractionation. Rewrite that paragraph.

L. 119: 26th

L. 127: mention sampling frequency of Irgason

L. 131: correct "inst"

L. 135: correct grammar

L. 138: delete "m"

L. 314: insert what delta values you are giving here? D18O? of CO2 or H2O?

L. 316: Specify everywhere in the script whether you talk about d18O of CO2 or H2O.

Fig. 9 caption: The panel labels are mixed up.

L. 350: what is meant by "evaporative fractionation related to the transpiration" -reword

L. 381: Exchange "A possible cause" with "The cause"

Figure A3: Which time period is presented?, use "re-scaled CO2" for consistency in legend

Literature:

Good, Stephen P., Keir Soderberg, Lixin Wang, and Kelly K. Caylor. "Uncertainties in the Assessment of the Isotopic Composition of Surface Fluxes: A Direct Comparison of Techniques Using Laser-Based Water Vapor Isotope Analyzers." Journal of Geophysical Research: Atmospheres 117, no. D15 (2012). https://doi.org/10.1029/2011JD017168.

Good, Stephen P., Keir Soderberg, Kaiyu Guan, Elizabeth G. King, Todd M. Scanlon, and Kelly K. Caylor. "δ 2 H Isotopic Flux Partitioning of Evapotranspiration over a Grass Field Following a Water Pulse and Subsequent Dry Down." Water Resources Research 50, no. 2 (2014): 1410–32. https://doi.org/10.1002/2013WR014333.

Keeling, Charles D. "The Concentration and Isotopic Abundances of Atmospheric Carbon Dioxide in Rural Areas." Geochimica et Cosmochimica Acta 13 (1958): 322–34.

Mauder, Matthias, Claudia Liebethal, Mathias Göckede, Jens-Peter Leps, Frank Beyrich, and Thomas Foken. "Processing and Quality Control of Flux Data during LITFASS-2003." Boundary-Layer Meteorology 121, no. 1 (2006): 67–88. https://doi.org/10.1007/s10546-006-9094-0.

Wahl, S., H. C. Steen-Larsen, J. Reuder, and M. Hörhold. "Quantifying the Stable Water Isotopologue Exchange Between the Snow Surface and Lower Atmosphere by Direct Flux Measurements." Journal of Geophysical Research: Atmospheres 126, no. 13 (July 16, 2021): 1–24. https://doi.org/10.1029/2020JD034400.

Welp, Lisa R., Xuhui Lee, Kyounghee Kim, Timothy J. GRriffis, Kaycie A. Billmark, and John M. Baker. "δ 18 O of Water Vapour, Evapotranspiration and the Sites of Leaf Water Evaporation in a Soybean Canopy." Plant, Cell \& Environment 31, no. 9 (2008): 1214–28. https://doi.org/10.1111/j.1365-3040.2008.01826.x.

Wei, Zhongwang, Kei Yoshimura, Atsushi Okazaki, Wonsik Kim, Zhongfang Liu, and Masaharu Yokoi. "Partitioning of Evapotranspiration Using High-Frequency Water Vapor Isotopic Measurement over a Rice Paddy Field." Water Resources Research 51, no. 5 (2015): 3716–29. https://doi.org/10.1002/2014WR016737.

Yakir, Dan, and Xue Feng Wang. "Fluxes of CO2 and Water between Terrestrial Vegetation and the Atmosphere Estimated from Isotope Measurements." Nature 380, no. 6574 (April 11, 1996): 515–17. https://doi.org/10.1038/380515a0.