Reply to Reviewer 3

In the following please find the corrections and comments to the referee's response.

General comments

This manuscript reports the results of an intercomparison exercise that aimed at testing the consistency of cryogenic water extractions for the analysis of stable isotope of hydrogen and oxygen in soil water among worldwide-distributed laboratories. In the last few years, the ecohydrological and soil science international communities have shown a strong and increasing interest in better understanding the functional interrelationships between soil and vegetation based on the use of stable water isotope data. The cryogenic water extraction technique has so become a golden standard for sampling water from the unsaturated zone. Very recent studies, often conducted by the first author of this manuscript and colleagues, already showed some potentials and limitations of this technique, and provided helpful information for users. However, a worldwide interlaboratory comparison among several cryogenic extraction facilities was still missing, thus this work it is certainly welcome. Indeed, I believe that this manuscript is timely and of great interest for the readers of this journal. The manuscript is very well written, logically structured, nicely illustrated, and the conceptual steps can be followed very well. The working hypothesis and the specific objectives are well posed, following a substantial introduction, the statistical analysis is correct, and the results and interpretation are well supported by the data. The methodological approach leading to the comparison exercise was solidly defined and clearly presented. As noted by the Authors themselves, the large differences in performance found among the labs included in the exercise are somehow worrisome and pose questions on the possible adoption of cryogenic water extraction as a standard method for soil water sampling. However, these results are very relevant to the scientific community because implicitly suggest cautions in comparing isotope soil water data extracted by different facilities, and indicate that much technical work is still needed to test possible further controls on these differences and develop new techniques able to return more consistent results.

We thank Daniele Penna for taking the time to review our manuscript and providing this very positive feedback.

Specific comments

1. P4 L11. Here, and throughout the rest of the manuscript (e.g., P7 L16; P9 L12-14; P11 L15. . .), it is not immediately clear what "isotope results" are, and I suggest to replace this term with "values" or "data".

Response: We will change it to "data" or "values" or "composition" as suggested.

2. P5 L14. The authors reported that the soil was homogenized before shipping. However, as noted by the other reviewers, I wonder if possible heterogeneities in the analysed soil samples (especially for small volumes) could have been present and could have affected the results.

Response: For our intercomparison, we used "standard lab soils" from the LUFA Speyer (https://www.lufa-speyer.de/index.php/dienstleistungen/standardboeden/8-dienstleistungen/artikel/57-standard-soils). In guidelines of the German JKI (Julius-Kuehn-Institute) and other relating guidelines (OECD), soils with certain characteristics are recommended for such studies. The LUFA lab already provided dried, homogenized and sieved soils (with a 2 mm screen)

according to “Good Laboratory Practice”. However, we repeated the homogenization and oven-drying process to ensure that all subsamples that we shipped to the respective labs were equal.

3. P5 L27. I find the definition of “water-water” cryogenic extraction a bit confusing. I suggest to use, throughout the manuscript, simply the terms “water extraction” vs. “soil water extraction”, or something similar.

Response: We will follow the reviewer’s suggestion.

4. P7 L1-2. Although intuitive, I suggest to add a short explanation about the choice of applying different extraction times for the silty sand and the clayey loam soils.

Response: We will add the following to the revised version of the manuscript: ”Since different extraction times and temperatures were applied in past studies, we decided that participating laboratories should follow two different extraction approaches...” and further “For comparison, in past studies extraction times from 2.5 min (Koeniger et al., 2011) over 30 min (West et al., 2006) to 7 h (Araguás-Araguás et al., 1995) for sandy soils and from 30 min (Goebel & Lascano, 2012) over 40 min (West et al., 2006) to 8 h (Araguás-Araguás et al., 1995) for clayey soils were reported.”

5. P8 L28, and P12 L6-20, and Figs. 5-6. It seems to me that these are not “true evaporation lines” describing the progressive isotopic enrichment of an individual source water (see Benettin et al., 2018 who reported the often misused term and concept of evaporation line). This might not be a critical point in the interpretation of the results and the overall meaning of the research. However, for the sake of accuracy, I suggest to check this and in case change the terminology (eg, simply calling them regression lines) and slightly re-interpret the results reported at P12 L6-20. Moreover, it’s not very clear to me why in the left panels of Figs. 5 and 6 (8% WC) one regression line (for 8% samples) is reported in addition to the GMWL whereas in right panels of Figs. 5 and 6 (20% WC) two regression lines are shown (both for 8% and 20% WC). Please, fix this or explain.

Response: We will change the term to “regression lines”. We will further add the following to the respective section on P12: “Benettin et al. (2018) recently revised the widely used concept of evaporation lines. The authors question that the trend line passing through fractionated soil water samples correctly identifies their source water and emphasis that trend lines through evaporated samples can differ widely from true evaporation lines.”
For clarity, we deleted regression lines for the 8% WC versions from the right panels of Figures 5 and 6.

6. P9 L23-24. Do the Authors have any idea about the reason for recovery rates higher than 100%? Could this somehow affect the results? Perhaps a sentence could be added here (trying to avoid the risk of speculation).

Response: We already included a potential explanation for this phenomenon in the discussion section of our previous manuscript version: “This could be attributed either to leaky vacuum systems (which might allow atmospheric water vapor to enter the system) or to a remoistening of the oven-dried soil samples before water extraction.”

7. P11 L27. In addition to the statistical results, I wonder whether it might be appropriate to show OA-ICOS and IRMS data as boxplots to graphically stress the difference between values returned by the two techniques.

Response: Figures 3 and 4 show the mean of the three replicates with respect to the reference DI water and y-error bars stand for the isotopic variation of the replicates. Since our sample size (n=3) is relatively small, we decided to display only the data's central tendency by using traditional mean-and-error scatter plots. The decision follows recommendations by Nature methods (2014) and we hope that the reviewer will concur with this assessment (a great suggestion though, if we had a larger n).

8. P17 L23. The reference of "Orlowsky et al., 2018" is missing from the reference list.

Response: We will add the missing reference.

9. Fig. 1. I suggest to increase the size of the axis labels.

Response: We will follow the reviewer's suggestion.

10. Fig. 3 and Fig. 4. In the caption: was the mean computed among the three replicates? If so, I suggest to specify this for the sake of clarity.

Response: We will edit the sentence as follows: "Symbols represent the mean of the three replicates and y-error bars stand for the isotopic variation of the replicates."

11. Fig. 5 and Fig. 6. I suggest to add in the caption that the legend includes explanation for the symbols used for the 16 labs and the two extraction approaches.

Response: We will add the following: "…for OA-ICOS and IRMS data (upper and lower panels, respectively) from the 16 participating labs (different colors represent different labs) and both extraction methods (lab-procedure: I and pre-defined: II). For reference, plots include the Global Meteoric Water Line (GMWL, solid red line) and soil water regression lines for 8% and 20% WC (solid green and orange lines, respectively)."

References

Araguás-Araguás, L., Rozanski, K., Gonfiantini, R., & Louvat, D. (1995). Isotope effects accompanying vacuum extraction of soil water for stable isotope analyses. *Journal of Hydrology*, *168*(1–4), 159–171. https://doi.org/10.1016/0022-1694(94)02636-P

Benettin, P., Volkmann, T. H. M., von Freyberg, J., Frentress, J., Penna, D., Dawson, T. E., & Kirchner, J. W. (2018). Effects of climatic seasonality on the isotopic composition of evaporating soil waters. *Hydrol. Earth Syst. Sci.*, *22*(5), 2881–2890. https://doi.org/10.5194/hess-22-2881-2018

Goebel, T. S., & Lascano, R. J. (2012). System for high throughput water extraction from soil material for stable isotope analysis of water. *Journal of Analytical Sciences, Methods and Instrumentation*, *02*(04), 203–207. https://doi.org/10.4236/jasmi.2012.24031

Koeniger, P., Marshall, J. D., Link, T., & Mulch, A. (2011). An inexpensive, fast, and reliable method for vacuum extraction of soil and plant water for stable isotope analyses by mass spectrometry. *Rapid Communications in Mass Spectrometry*, *25*(20), 3041–3048. https://doi.org/10.1002/rcm.5198

Nature methods. (2014, February). Visualizing samples with box plots. Retrieved May 22, 2018, from https://www.nature.com/nmeth/volumes/11/issues/2

West, A. G., Patrickson, S. J., & Ehleringer, J. R. (2006). Water extraction times for plant and soil materials used in stable isotope analysis. *Rapid Communications in Mass Spectrometry*, *20*(8), 1317–1321. https://doi.org/10.1002/rcm.2456