Comment and Answers on hess-2022-37

Rachel Havranek (Referee)

Referee comment on "Technical note: Conservative storage of water vapour: a key to practical measurements of water stable isotopes in tree stems and soils" by Ruth Magh et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-37-RC1, 2022

General comments:

In this paper, the authors test a flexible, cost effective way to sample water vapor from trees for stable isotope geochemistry. This kind of system fills a strong need for the stable isotope community, and will be very useful for many different applications. To test that their system was reliable, the authors performed storage tests both in the lab and in the field. The authors found that there was systemic storage bias in oxygen isotopes over time, and that bias was only present for 'crazy heavy' waters for hydrogen isotopes over time. Below, I suggest the addition of one simple experiment to the manuscript to demonstrate that the vial cleaning protocol is sufficient, and that the vials are sufficiently resistant to atmospheric intrusion over the proposed storage timescale (3 days or less). Broadly, I think this is an excellent paper and the system will be widely used by the community. I strongly support the publication of this paper in HESS.

Rachel Havranek

We thank Rachel Havranek for this very comprehensive and insightful review and hope to answer all her comments/questions sufficiently and revise the manuscript accordingly.

As for her main concern about the cleaning and storage sufficiency, we agree to add data to the revised manuscript, which we already obtained for the original draft. We will include those in the revised version.

Specific Comments:

Atmospheric intrusion & vial cleaning protocol: My largest comment on this paper is that the authors did not sufficiently address the issue of atmospheric intrusion, nor did they demonstrate that they sufficiently eliminated an atmospheric signal from their vials prior to sampling.

We were recently made aware that atmospheric intrusion in the soil community is referring to the mixing of atmospheric air into the subsurface (see e.g. (Lowrey et al. 2016)). To avoid confusion between the communities and because this method can be used for soil vapor sampling, we will rephrase the respective section in the revised manuscript and will from now on use the term "diffusive exchange" when talking about the exchange between vapor inside the vial with the atmosphere.

As mentioned above, we do have the data showing that cleaning protocol and diffusive exchange in and out of the vial are highly unlikely and will include those in the revised version of the manuscript.

Briefly, we sampled dry air from the desiccation tower into three vials and stored them for 14 days. We then measured the vials the same way as the other samples and analyzed the data accordingly. We did measure water vapor contents ranging in the same magnitude as when

measuring the desiccation tower directly. However, we need to emphasize here (and will do so in the revised manuscript) that the "dryness" of the desiccation tower varies with temperature (which is to be expected) and can thus only give a range between 300 and 600 ppmV for that value. The samples we measured lay within this range, which lets us conclude that the signature within the vial prior to sampling is sufficiently "overwritten" (making the cleaning protocol sufficient) and supports our previous argument that we cannot observe atmospheric intrusion into the vial even during 14 days of storage.

With regards to the cleaning protocol: I appreciated the discussion of vial cleaning protocol and I think that baking the glass vials at 65°C for 24H is likely sufficient. However, I have concerns that the PTFE caps were sufficiently dried (since PTFE is SO 'sticky'), and suspect that might the source of *some* of the observed drift. When I have played with PTFE fittings in the SWISS system, I've been disappointed by how much PTFE exchanges. I am also curious what gas the vials were purged with prior to sampling? If there is atmosphere in the vials when they are crimped, and then they cool post heating, I would expect atmosphere to stick to walls of the vial, which could ultimately exchange with sample vapor, even after so many vapor 'turns' during sampling.

We are a bit surprised about the "stickiness" comment regarding PTFE. This is not what we, nor other authors working in the same field observed when using PTFE tubing or lids. Picarro recommends Teflon coated lids to store liquid samples because of its properties (see here: https://www.enviscid.com/uploads/5/6/3/8/56382687/l2130-i_users_manual_rev_a.pdf). We will extend the discussion of the revised manuscript a little, to show that PTFE is used because it is the most diffusion tight material available on the market, it is chemically inert and best option when stainless steel/glass is unavailable or impractical. Properties say it's hydrophobic therefore stickiness would not be easily explicable.

As mentioned above, we can also show with the newly available data that the sample turnover does sufficiently overwrite whatever leftover atmosphere is in the vial. Also, the vials are not purged with any gas prior to sampling and kept in a desiccator until cooled off. This information will be added to the protocol in the revised version as it is currently missing.

I also think that the authors need to do a little more work to demonstrate to the readers that atmospheric intrusion is not a source of error for this system. The authors didn't include data in this initial submission to allow readers to evaluate if that was a source of error. This would be a simple and convincing addition.

We agree and will therefore add the samples that contained dry air only, and additionally data on the atmospheric isotope composition, towards which the samples should have drifted had there been diffusion into the vial from the atmosphere.

Encouragingly, for most natural waters the VSVS system was within error of direct measurement. What worries me is that 0-day did not overlap within uncertainty for either oxygen or hydrogen for the light standard – which is a pretty typical high altitude and/or high latitude value. It's not clear to me that method precision is truly accounted for. To hit two birds with one stone, I suggest a very simple, short timescale experiment where the authors fill a set of vials (perhaps 10 to sufficiently catch crimping variability?) with just dry air from the drierite system, and then do a storage test, just as they did with the rest of their lab tests. Given the authors' recommendations in the discussion, I think a 3 – 5 day experiment should be sufficient. I would also recommend that the authors measure the dry air the day they fill the vials so that small changes in water

concentration can be detected. With our drierite system, I know that there can be some variability and so it's nice to have that baseline value recorded.

See above answers.

Other protocol questions: Given that this paper is directed towards an audience of potential future users, I have a few small questions about lab protocols that could likely be answered either through some supplemental text or the addition of a few short sentences into the main body of the manuscript

Crimping: I am unfamiliar with how the crimping process worked, I think a very short (a few sentences at most) discussion of how to know that a cap has been sufficiently crimped or has been over-crimped (and therefore leaky) would be very helpful for the target audience. Alternatively, is there a way the authors imagine they could screen for that during sample measurement?

Very good idea, thanks. Yes, the crimping process is a potential error source e.g. when the crimping tool (which works similarly to pliers) pinches too hard or not enough. The handling person is then able to turn the lid around the bottle neck, which should be avoided. We did not have a problem with this but acknowledge this could be a problem.

We will add this to the revised manuscript.

Did the vials re-cool between heating and measurement or were they measured warm? (did they have the hot plate under them as your measured them?).

Yes, the latter. We will add it to the revised manuscript.

Your total flushing time is somewhat based on flushing volume 'turns'. I noticed on figure one you cite an inlet rate of 35 ml/min. On our 2130 we actually only pull ~25 ml/min. So, I wonder if you have double checked that rate? It might be nice to put a note on line 103 that says something like "time to one full volume can vary Picarro to Picarro".

We will add this in the revised version of the manuscript. As far as we know the picarro does not aim for a certain flow rate per se, it rather aims at keeping the cavity pressure constant (50 torr) by locking the cavity inlet valve to a certain setting and adjusting the cavity outlet valve accordingly. Therefore, the effective flowrate seems to be a function of e.g. ambient air pressure (or whichever pressure is effective/applied at the analyzer's inlet port) or pump performance and will likely be different for different instruments. We have tested 4 different Picarros and found that the flow rates ranged from 27 to 33ml/min, with the mean resulting in ~30ml/min. We will add this information to the revised manuscript, and correct our proposed 35ml/min, which was a typo.

How do the authors identify spurious vials?

We did not exclude any data, and we did not use broken vials. Upon doublechecking whether the vials were properly crimped, we excluded some vials with twistable lids (see above explanation) prior to sampling (i.e. those were never filled or measured). Because we left enough time between measuring two samples, we did not observe any memory effect and therefor did not need to exclude data.

If we theoretically had encountered readings that did not at all fit the expected isotope signature, we would first have checked for the water vapor concentration to see whether we had missed a badly crimped vial/vial with a defective lid/vial with cracked glass. If

that had not been the case, we would have doublechecked in the protocol whether there had been a sampling error/any unusual occurrence during sampling. We will add a brief paragraph on such problems to the revised manuscript to give guidance to the reader.

Storage time correction: I think more explanation of your choice to use a generalized storage correction is needed here. From what I can tell from your data, the offset between ambient air and the measured isotopes should dictate how much it moves. For example, the storage correction for d18O from the 'light' isotopes is very different than the one predicted by the 'crazy' heavy. It's relatively easy to imagine a scenario where the ambient air isotopes are very similar to those sampled for an experiment and so just using a light or medium correction would be more appropriate. If the scale of correction is indeed not very different across isotopes, it would be helpful to demonstrate that some way in the supplement.

Storage time correction relates the change between the samples measured on the sampling day (i.e. "0-day samples) to samples measured on subsequent days (i.e. 1,3,4,7,14). For the oxygen isotopes this is only regarding natural abundance ranges as we did not label oxygen. The different isotopic signatures are not considered here since we relate the isotopic change to the storage time. We assumed this change to be similar for each natural abundance isotopic composition. We will give a table in the appendix including the slope and intercept for this relation for each single isotopic composition in the revised version.

These papers are really hard to do, and I applaud the authors for the effort. But given that they are going to apply a linear regression to correct data in the future, I question whether or not they have gotten enough data to truly say that they are representing real variability. Some further discussion of sample size, as it relates to creating a correction factor would be helpful. Further, do you think each lab should create their own correction line or do you think that this is more universally applicable?

We would argue that we provided a large enough data set (i.e. 5 standards *10 replicates per standard =50 data points per storage time) to provide an estimate of the storage effect for our lab conditions.

For this method to be sufficient under the premises that each picarro/each lab uses their own "storage effect time regression" each time they are measuring a large set of samples. This would be the minimum requirement from our point of view for our objective.

Generally, the approximate size of the storage effect would be expected to be similar between labs, however, to obtain an exact storage effect on the respective instruments, we do recommend providing their own correction of the storage effect. Our values can be used as a guideline.

We will add this to the section 4.3 of the revised manuscript.

I appreciated that the authors included a preceding works section – it demonstrated the motivation for their work and helped show context. I also appreciated the discussion of how current system constraints have introduced location and social biasing into the scientific literature.

Thanks ☺

Technical Corrections: In this section, I have labeled my correction by line

32-33: I think it would be appropriate to significantly expand this citation list to showcase the variety of kinds of in situ work that is being done. For example, it would also be appropriate to cite Maria Quade's, or T.H.M Volkmann's work here. Beyer et al., 2020 (HESS) would also be a nice addition here.

We will expand the list in the revised manuscript

43: I think it's also fair to cite Orlowski et al., 2016 here

We will add this reference to the revised manuscript

48: Beyer et al., (2020) *HESS* would also be good to cite here

Will be added in the revised version

49: The word choice interferences doesn't sit well with me, I wonder if this sentence could be reworded to make your meaning clearer. Perhaps ".... Direct equilibration between liquid and vapor water in the soil ..."

The word we chose was "inference" not "interference". Does this change the way you feel about the wording? Technically, this is what an in-situ approach does, so we feel the choice of word is appropriate in the context.

91: Were you able to dry the PTFE tubing that goes between the sample vials and the CRDS to eliminate any memory effect from that part of the system? I imagine that could be done quite simply by just having a 'dry' vial that you flush through between samples.

See above answers related to PTFE stickiness and measuring dry vials.

Figure 1: These kinds of figures are very challenging to make well. I appreciate that the photo demonstrates practical complexity in the lab setting. I think this figure could be improved with the addition of a small, simplified cartoon to the side showing all the components. This would help readers hone in on the important components without getting too distracted by all of the real-world lab complexity. Or, another way to make the figure more readable to be to add a small white box behind the text boxes, I had a hard time with the red text in particular.

Thanks, we will use your suggestions and edit the figure accordingly in the revised manuscript.

103: You cite a 35 ml /min pull rate from the Picarro, with a 50 ml container & 10 minutes of flushing that should only be 7 turns (35*10/50 = 7). I'm not sure that nit-pickiness really matters for the scope of this experiment given that the authors observed signal stabilization. But, I think given that this is a methods development paper its most helpful to the community to be hyper-specific about some of these details.

See above answer regarding the picarro flow rate and our calculation. As mentioned, the typo 35ml/min will be changed to 30. However, in the line you refer to here, we cannot see any such calculation. When calculating with 30ml/min suction rate, we get 6 turns. We will add this to the revised version to clarify that the sample will be exchanged several times.

156: A huge advantage of this system over the SWISS is the size and therefore ease of transport (e.g. 50 ml vials vs. 650 ml flasks), so I think one selling point that could be an estimate of total size & weight (just as the authors did with the battery). The SWISS

also requires quite a bit of time consuming construction and plumbing and so some sentence to that effect, and an advantage of this system is that it is easy to set up.

We will add an estimate of weight and size for 100 crimped vials in the revised manuscript. It is 2.5 kg weight and 30x13x50cm is the packaging when re-using the original vial box.

180: Is your data reduction code widely available (e.g. github)? This development paper would be even more helpful to the community if we can also see the data handling process.

The code is available upon request and will be added to a repository once it's completely cleaned.

Figure 3:

I'm not sure if it's the file the authors provided, or a formatting issue from HESS, but it would be great if figure 3 was the full width of the page. If it is an author-side issue, using ggsave you can set the figure width to 6.5 inches - ggsave(plot, "plot.pdf", width = 6.5, units = c("in")).

We will provide a figure with these dimensions for the revision.

Where does ambient air sit in isotopic space relative to the standards measured? I think for the "crazy heavy" water it's easy to see that its trending towards room values, but it'd be nice to have a sense for how far it got towards that value.

We will provide a figure where we can show that at least for the oxygen isotopes a drift towards ambient air is not the case. For the heavily enriched waters this would be easy to assume as it trends toward depletion, but the atmosphere lies within the natural abundance range of the measured standards, and it is there we can see that also for the hydrogen isotopes the drift is not towards ambient air. Which we see, in addition to the dry air samples, as an argument that if there was diffusive exchange with the atmosphere it was a one way process out of the vial into the atmosphere but not vice versa.

190: Please expand on why a wilcox test is appropriate here, and what you hope to learn through it. Unfortunately, many people reading your paper might be unfamiliar with that statistical test – and a short 1 -2 sentence explanation of its use and limitations would help your reader assess suitability.

A Wilcoxon (the typo will be corrected in the revised manuscript) test is a non-parametric approach to detect differences between two groups of data. We chose it to be able to compare the data since they were not normally distributed which would be a requirement for using a parametric test (students t-test). We will add this information to the revised version of the manuscript.

220 – this paragraph as written is a little confusing. I think this could be solved with just a quick additional introduction sentence that says something along the lines of "We observe two different patterns between hydrogen and oxygen isotopes, we first address storage effect on hydrogen isotopes and then oxygen isotopes."

We will add this to the revised version of the manuscript.

243: Ahh, very clever. I initially didn't like that choice, but with the explanation, it makes sense.

Thanks ☺

336 – My feeling is that this contradicts what was stated in the results

We can rephrase this sentence in case this makes it clearer to: "This is the best supported scenario, as the data supports that there is no diffusion…" We want to emphasize that the exact physical process behind this remains unsolved beyond the arguments we deliver in our discussion.