Dear Mrs Magh,

Thank you very much for your extensive and very helpful feedback.

We are now working with a native speaker to improve the language and grammar of this paper. Thank you very much for pointing out a few necessary corrections already, we will make sure to see to the other language mistakes you did not comment on.

The MM section will be revised and improved as per your feedback. We will make sure to change my perspective from “expert” to “reader with little to no expert knowledge” when rewriting not just this section but the entire paper. The legend for the figures, in particular figure (1) will be added and figures will be clarified/improved as per your feedback.

Regarding the grammar we are now working with the already mentioned native speaker.

Regarding the figures, we agree that there are too many and they will be merged where possible. The scale in Figure 13 will be adapted and the other figures for similar improvements crosschecked. We really appreciate you pointing out that the current state of the figure is misleading; we will improve it as recommended.

After rereading the paper with a fresh mind, we have also noticed the repetition in the discussion and will crosscheck this section with the results to ensure no repetition remains while also shortening this part.

We are very grateful and appreciative for the grammatical, lingual and especially the comments showing your topical expertise on this draft.

Responses to the specific comments to the text:

Abstract:

L 16: specify flow rates; e.g. sampling flow rates or measuring?

Yes, this will be added.

L 19: delete “with” and then “deviating”

Thank you.

Keywords:

Replace “tightness” with “air-tightness” and “container” with “storage container”?

Thank you for the suggestions!

Introduction:

L 49: wrong citation style

This will be edited.

LL 52 ff: redundant, delete

This will be deleted.

LL 81 ff: this paragraph is missing the original publication of Havranek's 2020 publication

(**https://doi.org/10.1002/rcm.8783**)about the SWISS system, please add for completeness.

Thank you for pointing this out and providing a direct link to the paper. This will be added.

L 87: add information about the length of the storage period.

Will be added, thank you.

LL 109 ff: Please rewrite and structure this section more carefully, so that the RQ's gain visibility and it becomes clear to the reader which of these you aim to actually address with this work. Please also rewrite the next paragraph, it reads awkward and partially redundant.

Structure the objectives and number them to increase visibility. Also, the last paragraph: "Innovation" does either need a proper header or rewriting to integrate the Innovation.

Thank you, we will rewrite and double-check this with the native speaker.

Material and Methods:

Generally, it would be a good idea to replace the section headers with experiment and number? This way every experiment is described with every step needed to carry it out, and nothing is left out or to imagination as it is now.

Great idea, thank you! We will modify this chapter accordingly.

Figure 1 and Paragraphs 2.1.1/2.1.2: I don't understand why the 4. Point (ND20 and 1L Tedlar bags) is in parentheses and which the tested material is. Please clarify.

We will double check this and clarify.

Also, the figure could use a proper legend, probably making it easier to differentiate between material and test. Now it looks like there's 2 main experiments, while from the text it seems there should be 4, so maybe a better color-code and symbol legend would enhance understanding at first glance without having to read everything twice?

Thank you, we will work on this and make sure color-coding for better clarity and a proper legend are added.

Also, I understand the decision criteria, but what if the container that was tightest didn't score as high on the isotope test?

The sum of all the criteria provided an indicator of which container is most optimal. We would not always achieve the best results for every criterion, but the overall total would determine the selection.

How do you proceed from there?

We need to determine which container is most suitable by evaluating all the selected criteria. We will carefully review this section of the decision criteria and, if necessary, restructure it to provide more precise details.

Also, what kind of criterion is 3. "time to analysis"?

The analysis time was considered essential, as the objective was to obtain data under conditions comparable to in situ measurements, that is, in real time, with a maximum interval of 0 to 1 days in between sampling and isotopic concentration measurement. We consider this interval ideal for reducing the effects on isotopic concentrations.

How are the criteria ranked, is the number 1 the most important and decreasing importance? This should be clarified please. - Each criterion was rated on a scale from 1 to 5 (1 = poor, 5 = excellent) and assigned a relative importance: isotopic stability (50%), leak tightness (20%), analysis time (10%), reusability (10%), portability and transport (5%), and cost (5%). The sum of the weighted scores enabled an objective comparison of the containers evaluated. The supplementary information details the scoring process. Section 2.1.7 of the methodology explains each decision criterion. There are two groups: critical and non-critical. Criteria involving leak tightness, isotopic stability, and analysis time are considered critical, while the others are non-critical. We will be adding this into the paper.

Also, this section is missing the reference to how long the storage is and what was stored?

We will also carefully review this section of the manuscript and include the requested information.

Section 2.1.3:

The rationale for syringe size is not explained. Please elaborate.

We will further explain the reasons for selecting syringe sizes and their role in the experimental design. The reasoning behind testing it was that a small syringe might potentially constrain the flow rate to the analyzer and cause rapid pressure drops/increases that could lead to isotopic effects. We will address this in the revised version.

L 198: which method was developed in this study?

(L 120) Our goal was to develop a better lab method for sampling water vapor in glass bottles to fill technical gaps in current isotopic analysis protocols. By evaluating factors like container type, flow rate, and storage temperature, the study identifies the conditions needed to ensure isotopic stability and high data quality. This multifactor approach provides an accessible and robust framework for investigating water

isotopes in plants and soil-atmosphere interactions. The method enables the creation of continuous, comparable data sets and is designed for future use in the field.

L 200: after the first day of sampling? So 7 + 1 day?

This means that after 24 hours of sampling (one day), we will correct this drafting error for clarity.

L 201: I don't understand, did you measure the same sample after 3 days and again after 7? Or is this a replica? And if so, why is this only mentioned for 3 and 7 days, did you not also measure after 6h and 1 day?

Thank you for identifying this point of confusion. We will revise the paragraph to clarify when the sampling and measurements occurred. We aim to explain that vapor sampling was performed, and the samples (three replicates for each storage time) were stored for 6 hours, 24 hours, 3 days, and 7 days. After each established storage period, each sample was measured.

L 203: yes, I get it, but again, readers do not know this yet.

Thank you for pointing this out.

L 206: but this is not a repetition, this is a new experiment with the syringe size. Please be careful with wording.

We will review and carefully edit this to ensure the reader understand that a new experiment was undertaken for this section of the paper.

L 209: so, the dry air test was not conducted using different flow rates?

Our goal was only to determine whether the diameter of the syringe used to make the hole for sampling affected the tightness, thank you for pointing out your confusion so we can ensure to clarify and describe this section more carefully.

L 209 ff: which are the isotopic references?

They have been explained in section 2.1.8.2 (table 1), but they will be added to this section to improve clarity.

L 214: please stick to either days or hours, this is getting confusing.

We will edit to use the same unit always.

L 216: this flow rate depends on analyzer and type it is not a predefined setting for all Picarros.

Correct, in this case, it corresponds to our Picarro. We will add the specific flow rate of our analyzer in order to clarify this.

L 218: So the reference material was stored at these temperatures, or the samples were after sampling?

The samples were, we will clarify this.

Section 2.1.4 should be logically before section 2.1.2

We will swap the sections according to your suggestion.

L228: are those the same bottles as the ND32?

They are the same and we will clarify this.

L235/236: how were those bags treated before reuse?

Only the 1L FlexFoil sample bags were reused and rinsed three times with dry air through the valves before sampling to minimize memory effects. The 500ml aluminum bags did not get reused. We will elaborate on this aspect in the revised version.

Section 2.1.5 again, I think assigning each section to a specific experiment would be beneficial to not have to go back and forth between the sections.

This is a great idea which we will aim to implement in the revised ms.

LL 242 ff: this is confusing, I would imagine that you used different flow rates for sampling not for measuring? If so, why did you use larger syringes for measuring, I think the flowrate here would be determined by the laser and therefore be low? Pls elaborate and rewrite.

We will carefully review the description of sampling and measurement procedures to improve clarity.

LL246 ff: so all this is only relevant for experiment number 1? Or 3,4?

Only for experiment 1. We will clarify this in the text.

Section 2.1.7

L 287: I don't see how (3) "time to analysis" would be a critical condition, it is however related to points 1 and 2 but not a point itself, imo. You can decide to sample more often when aiming for an in situ like approach, but you do not need the data in real time, which makes this approach so beautifully suitable for remote regions, where you might find a great experimental site but no infrastructure. If I misunderstand, I would ask you to clarify in the text.

Thank you for sharing your expertise on this. Regarding the explanation of analysis time, we agree that it can be improved by taking samples more frequently if real-time data are desired. However, our method aims to explain that time until the analysis of the sample can be an important factor to consider in water vapor sampling because of the risk of contaminating the sample or have other isotope effects in the sampling vessel. Temperature changes during the time between sample collection and analysis, for instance, could have a great impact on the water vapor stored in the sampling bottle/bag. We will review and discuss the use of this criterion again and evaluate its importance within the criteria for selecting the most optimal container for water vapor sampling.

LL 296 ff: I would argue that isotopic stability should have more weight than 50% since it is the most important criterion in defining the usability of the methods. If all other criteria in um add to 50% but the sample is not isotopically stable, it is not useful at all and should not be considered.

Taking your very valid argument about isotopic stability into consideration, we will reevaluate the decision criteria to reconsider the level of importance of each one. Thank you for pointing this out.

Section 2.1.8 and subsections, sorry but this is too many subsections, again, I think restructure would enormously benefit the readability here.

We will restructure this section or reduce the number of subsections to improve readability.

L 310: and a reference to the respective section would be nice here.

Great point, we will make sure to add the requested reference.

LL 311 ff: I think that is not entirely correct. The test conditions only provide an assessment for the tightness of the entire storage container, not the syringe size per se. E.g. with the crimped vials the leakage might be due to the lid not fitting properly while you assess it as a syringe size problem. Please either elaborate if I misunderstood or rephrase so that this is clear.

The reviewer's assumption is correct and we will revise the text to clarify that the test conditions evaluate the overall tightness of the storage container system rather than syringe size alone.

L350 ff: Is this not redundant? I feel you have described this before in the treatment section

We will review this section and remove redundant descriptions where appropriate.

2.1.4 or is this a different treatment?

No, it is not and therefore the descriptions will be edited and, where appropriate removed.

L348: mas flow controller (MFC)

This abbreviation has been defined in L273 and therefore would not need additional defining. I have now, however, noticed that I have added an additional definition with this term in L394, which I will remove.

L353: drawn not withdrawn

Thank you.

L369: this should be section 2.1.9 not 2.1.8 please check the remaining sections for correction after this point.

Thank you for pointing out this oversight, we will revise the numbering of the following sections accordingly.

Section 2.2

L435: the liquid values, not fluid?

You are correct, it should be liquid values.

Section 2.3

L459: Furthermore…?

This misplaced word will be deleted.

Results:

Figure 5: the resolution of the figure is very poor, please resubmit with a better resolution and larger axis text etc.

We will improve our figures by enhancing their resolution and revising labelling to ensure an easier readability and understanding.

Section 3.1.1 What were the initial sampled H_2O concentrations? Did you test whether the bottle truly contained 0 ppm H_2O?

The in section 2.1.4 described process, confirmed that the bottles were completely dry. The dry-air rinse (L334) was used as an additional measure to ensure no H_2O residue remained in the bottles. A final check was conducted with the Picarro which showed that the bottles had an approximate concentration of 200 ppm of H_2O.

Section 3.1.2

L 494 ff: If you do not show the data I think it would be beneficial to the reader if you did not mention the tests at all. Also it would save some time when this is also deleted from the MM section.

After consideration we agree and will remove it during the editing.

Figure 6 also needs a re-do in resolution and text size etc

All figures will be revised according to your feedback thank you.

Figure 8 also needs a re-do in resolution and text size etc

We will produce figures with higher resolution, better axis labelling and greater legibility.

Section 3.2.2.1

The text does not really reflect the results shown in the figure: for example the statement in line 619 ff in the text you combine flow rate, storage time and condition, while the figure pools all samples into the respective condition. So, the present presentation does not support your conclusion. Please either adapt the text or the figure 9. Also, just from looking at the storage time, there is no real effect even at 72 hours storage (i.e. the box of the boxplots as well as the interquartile range does not look so different from the 6 hour storage..)

We will rewrite the explanation of the results shown in Figure 9 to ensure the descriptions give a clear picture of the presentation and findings.

Figure 9 s.a. and additionally horizontal lines would benefit readability

Horizontal lines will be added and the paragraph will be rewritten for improved clarity.

Section 3.2.1

Either I overread it or the information about how many replicas per flow rate and standard you have is missing for this section. It would be nice if you could add this information to the boxes of the boxplots and the tables you show in this section.

Thank you for pointing out the missing information, we will add this into the section.

Section 3.2.2.2

It is interesting to see that the low flow rate influences the measurements so much, one would assume the opposite, do you have an idea why?

We believe that when using a low flow rate, the sampling time should be longer than what we used as our results showed a slow but steady increase. We have a theory that this could be due to connecting three samples simultaneously. We are planning to look at a single sample to see if this would increase the process in the future as well as keeping the sampling with the three connected samples over a longer measuring period to see if the levels plateau at some point and when.

Can you please reference the Figure 10 sooner? It is much easier to follow the text with the figure already seen.

Great idea, we will implement this in our edits.

Figure 10: s.a. and add horizontal lines, also the y axis should read either δ2H or δD but not δ2D (the same is true for Figure 11 btw)

We will address all the points mentioned to improve the visibility of the data in the figures.

Section 3.2.3

What is the corresponding analysis you mean here? The rest of the paragraph reads fuzzy and confusing. Please clarify which analysis is the basis for this conclusion and then draw the conclusions. E.g. it is not obvious that flow rates of 125 ml/min truly is the best suited as the results look very similar for all flow rates except the lowest.

We will rewrite this section to enhance its clarity and ensure careful clarification of the conclusions and corresponding analysis.

Figure 11 seems redundant and considering that the paper already has too many figures, maybe it would be good to delete this one.

This will be deleted.

Figure 12 can be turned into a table imo

This is an excellent idea which we will implement for our edits.

Figure 13: I feel this representation is misleading, 2H as a fact always has a larger uncertainty in measurements. Since it is roughly 8 times as sensitive you could adapt the sclae by that factor and therewith enhance the readability and comparativeness between the two isotopes. You could then also plot both in one plot and in different hues of the same storage time. This is also true for the previous figures and would save you from having to delete too many figures.

We will implement your suggested corrections to the figure.

Discussion:

In the revised manuscript, we will substantially shorten and simplify the discussion, focusing on the interpretation of the results and avoiding repetition of results presented earlier.

The discussion is very long and many stretches are simply a repetition of the results (e.g. LL780 ff). It would greatly benefit from being rewritten and shortened. The results should not be repeated, if necessary, reference the respective section.

We will carefully revise the discussion section and remove repetitions and redundancies.

LL 833 ff: Conclusion i: then the argument would be that you should sample the bags longer and use an outlet so that the bag is flushed instead of simply filled. Is that something you did?

Our interpretation of rinsing the bags with steam while allowing an outlet, rather than simply filling them, comes from the approach we used with the bottles, which produced good results. Based on this, we believe that rapid filling of the bags could affect the isotopic composition of the collected sample. However, this procedure was not applied in the present study. The bags were filled as described in the methodology, without using an outlet to rinse them. Therefore, this suggestion should be considered a possible improvement to test in future studies.

LL 893 ff: This paragraph can be shortened into its main message, which is lines 904-906. The rest is repetition of results.

Any repetition will be removed while leaving the main message with improved language.

LL 914 ff: I have trouble with that conclusion; your results show stability within the first 72 hours as well as did mine in 2022. Why do you then recommend measurements have to happen in the first 24 h?

Our conclusion is based on the idea that performing measurements as soon as possible (24 hours) reduces the risk of additional variations in results caused by changes in factors such as temperature. The results presented in Table 2 describe considerations including flow rate, time, and storage conditions, and indicate the expected possible errors in the results. While the comment of the reviewer speaks about stability alone, our analysis includes other factors as well. Considering all of these factors, our recommendation is to analyze samples as soon as possible, within 24 hours. However, 72 hours are possible, but likely at the cost of precision.

The reviewers' comment is very valuable, as it provides a new perspective that will help us restructure the conclusion and clearly integrate the arguments from our results.

LL935 ff: Again, this can be boiled down to the main message which is a recommendation based on the results. It does not need to be this long.

We agree that this needs to be shortened and we will concentrate in our edits on the main message while removing any unnecessary information.

Section 4.3 should be shortened and integrated into section 5 Conclusions.

This will also be considered during editing with a strong focus on shortening and combining sections of this chapter.

Section 5 is not a conclusion but a nice summary, which could be used in the abstract. Otherwise, it may be redundant and could be removed.

That is a great idea for this summary and will be used for the restructuring and editing of the abstract. Thank you for this.