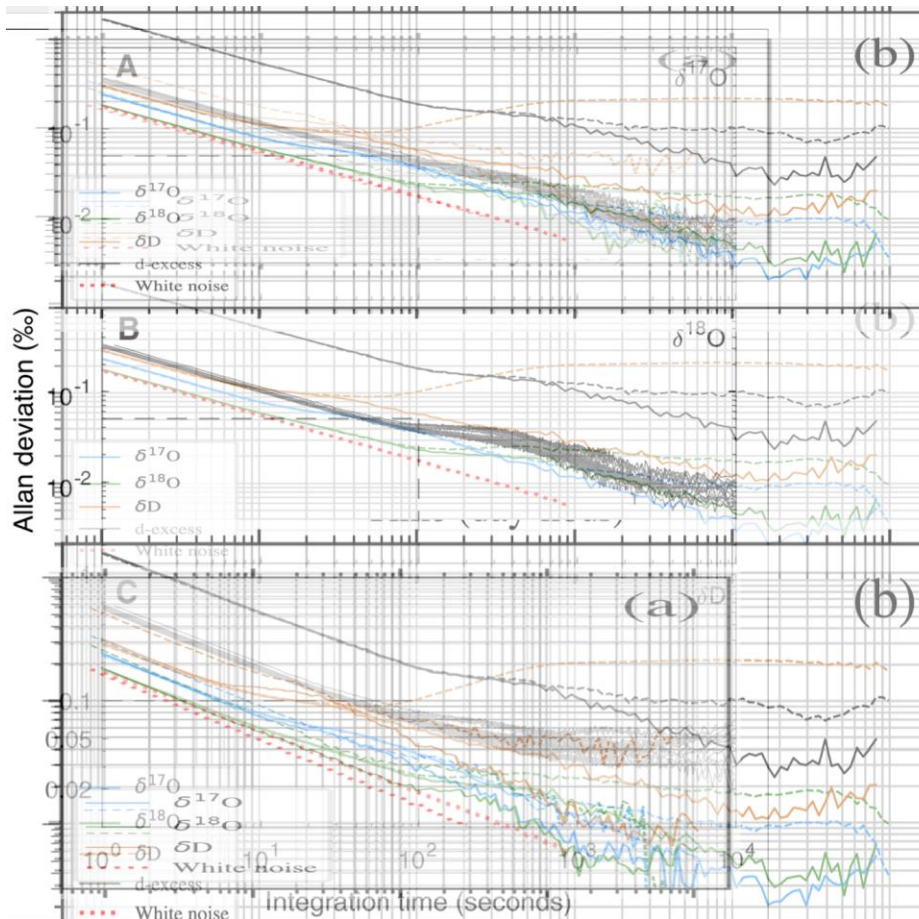
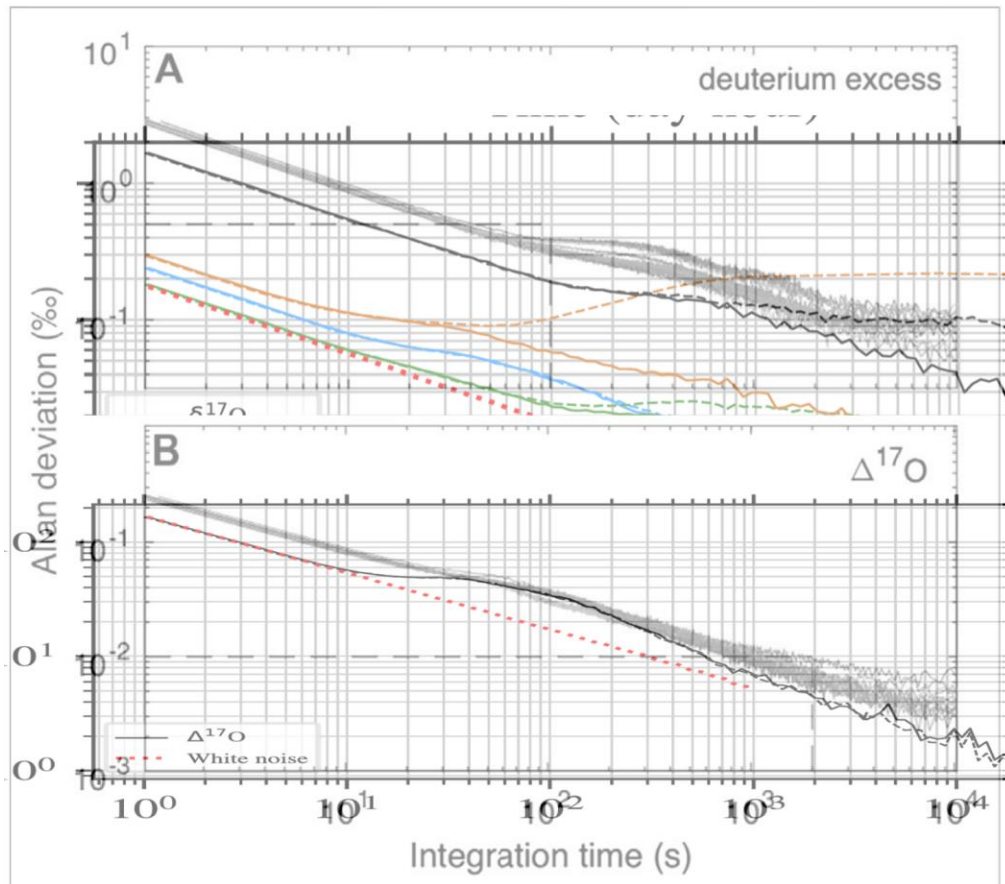


## Comments on memory and allan variance

Especially considering that this paper demonstrates that the high-frequency variability is owed to the stability of each instrument and not the vaporization module itself, these plots all fall well within the range of values published by Steig 2021, whose data are overlain on Figs 3 and 8 below and deviate notably only at high-frequencies. The Steig 2021 data are from much shorter runs of calibration standards during the South Pole ice core measurements, and they achieve similar results for all isotopes, especially when compared with the older instrument in figure 8 or the untruncated data in figure 3. The Steig 2021  $\delta D$  values are similar to the performance shown in figure 8 and far exceed the un-trimmed data, even without the extensive 16-hour trimming. This actually suggests that the memory internal to this new vaporization system is more significant than the memory of the system used in that paper. Can you speculate about why this very long data trimming is necessary or comment on your data processing workflow for routine measurements given this limitation? Please ensure that your language around stability and memory is better placed within the context of this type of similar vaporization system.





**Technical corrections and minor comments:**

In 12: I suggest rephrasing this to discuss the measurement of many samples instead of their "connection"

In 17: "measuring unknown samples shows" is specific to  $^{17}O$ , right?

In 18: the standard error is not provided -- perhaps this is a typo?

In 19: "enhancement" is not quite right here. Maybe rephrase to talk about the increase in deviation or noise level instead?

In 21: typo, should say "factor of 2"

In 27: "achieved to operate" could just say "operated"

In 41: typo after citation, should say "is classically"

In 46-47: please clean up citation formatting and italics. why not use the typical notation for the deuterium excess?

In 70-73: language is very confusing and inefficient -- please rephrase the first few sentences. A suggestion for In72 is to say "...deployment is the bubbler system, which has been used continuously..."

In 74: "and that there is minimal" could just say "and the minimal"

In 78: maybe "is not feasible for many campaigns" would be clearer

In 94: double-check the  $^{17}O$ -excess notation here and throughout for consistency

In 97-99: please rewrite this sentence as it is very confusing to follow

In 109: maybe "accuracy, we have further developed the patent application which was published in Steen-Larsen (2016)."

In 115: could say "sufficiently high accuracy for D170"

In 115-130: it might be more intuitive to understand the purpose of the two case studies if this whole section is rewritten as a brief paragraph about objectives and how you tested them

In 135 is a good example of how to refer to the patent for the first time -- consider revising In 109 to be similar

In 170: typo, should refer to 4.3. please check section references throughout.

In 170 and section 4.3 talk about decline in humidity values over long timescales, but figure 5 shows that it sometimes increases. we find that depending on the air flow and water flow, the precipitates either clog the tee itself or the capillary, and depending on which flow is decreasing (air or water) the instability can cause an increase or a decrease in humidity. is this similar? it seems consistent with the values shown in figure 5.

In 194: typo, "between 0.5 to 3.5 should" say "between 0.5 and 3.5"

In 195: why not just say "with a 1.59mm PEEK union"?

In 220: "frequent" should be "frequently"

In 222: SW and WW haven't been defined yet -- please make sure these and other acronyms are defined before use

In 265: please see comments above -- the performance shown here is excellent, but it is not significantly different from Steig et al 2021, which does not trim data as extensively as is shown here or have nearly as much time for each run. this (and the comparison with fig 8) suggests an increase in memory for this system, which should be discussed in more detail, if only to demonstrate that it can be managed for the applications promised.

section 3.2 compares the memory for this new system to the memory of the autosampler, which demonstrates that this method could be useful for lab measurements. But how is memory handled for field-calibration? What is a typical field calibration workflow?

In 318: should specify that the short-term trend is for the water vapor concentration

In 337: it seems like the variability in dD values at low-humidity ranges in Fig. 7 could also be influenced by the relatively longer memory of dD in the system – has this been investigated? Even if the liquid flow rate is the same at all levels because of the secondary mixing at TEE2, the retention time of water within the optical cavity itself should also worsen with decreased humidity which seems like it could contribute to this effect?

In 364: many readers will have difficulty understanding the statement of D170-->d170 conversion

In 377: consider "values" or something more specific than "one" for clarity

In 379: it is unclear what you mean by "error in the scale"

In 384+: this section is describing the experimental setup to examine the high-frequency noise – consider restructuring to include relevant details in the methods section of the paper instead

In 392: typo, "exceeding" should be "excess"

In 410-11: typo? – this sentence doesn't make sense to me

In 480-5: this does seem like a general problem for automated adjustments, though certainly not beyond characterization. Have you attempted to characterize the range of this effect for your system?

Ln 484/S3: which standards are used for calibration in each of the tests? It is not clear if it is the same every time and which of the standards listed in the supplement are used.

Ln 491: what is "relatively longer"?

Ln 494: is this memory limitation practical? Is it possible to shorten the tubing, or remove other dead volume from the system?

Ln 520: relative discrete autosampler injections

Ln 536: does not say what the standard error is (and looks copied from earlier section of the paper)

Ln 537-9: citations would be helpful here and elsewhere – this should also be discussed earlier in the paper

table 1: how is  $17O_{excess}$  determined for BER?  $d17O$  does not have enough significant digits to determine these values.

table 2: I'm not sure there's an advantage to showing both sets of calibrated values -- perhaps just choose one calibration method, explain it in the text, and report the values in the table?