Thank you to two anonymous reviewers and the editor for the detailed and helpful comments listed below in black. Our responses to all feedback are listed in red, below.

First review:

The author's response do not fully address some of the raised questions which makes it difficult to re-review all aspects of the revised manuscript. Therefore, only a few technical aspects are listed here: Please specify the resolution in ice core depth that corresponds to the 5 permeg precision stated in the abstract. (i.e. 5 permeg only at d >= 30 cm, according to Fig. 6) as well as the number of repeated measurements required.

We updated the abstract to specify that 5 per meg is attained only after 3000 s of analysis time. The depth resolution depends on the melt rate utilized, so we choose to report by analysis time instead.

Please consider improving the readability of the manuscript by more concise statements and by removing obsolete statements such as "..., it would be ideal to identify and eliminate sources of calibration error so that such an adjustment is not necessary".

We have made this statement and others throughout the text more concise for clarity.

Fig2: improve resolution

The resolution of the figure should be sufficient when taken from the .png and not from the .docx file.

Fig4: add labelling A, B, ... to the graph. Consider adding the D17O values to the figure.

We have removed figure 4 while attending to another reviewers comments; this information is now provided in Table 2.

Fig7: consider adding the SW2 as derived from Fig4, i.e. compare continuous with sequential (discrete) measurements of reference water.

We appreciate the suggestion but believe this would add no new information and would make the figure unnecessarily cluttered. The relationship between integration time and measurement precision for D17O for SW2 in Figure 7 is similar to previously published work using similar instruments (see e.g. Steig et al. 2014, Schauer et al. 2016, Steig et al. 2021) so this response seems to be characteristic of the instrument and not limited by this sample preparation method. Over long timescales, the continuous injection performs better than discrete injections due to the loss of information between injections and the propensity for instrumental drift over time; it is also a faster measurement because it eliminates downtime between discrete pulses into the commercial vaporizer.

In the conclusion specify the required time for CFA-CRDS measurements to obtain the cm-scale resolution.

We have added to the conclusion section that it still requires >1000 s to achieve meaningful CRDS data, and that it therefore requires >1000 s per cm to measure at the cm-scale.

Second review:

The authors did a good job in addressing the different comments. Still, I feel that some comments were not really addressed in a useful way for the reader.

1- Figure 4 does not really show the sequence of measurements. It should be possible to provide a table with date and depth/samples analysed so that we see when only standards were measured and when ice core samples have been analysed (with relevant standards for the calibration). If you really want to stick with Figure 4, focus should be done on periods when ice core barrels have been analysed (including also the standard measurements performed for calibration).

We have removed Figure 4 and added this information to the new Table 2 as suggested. Except where noted, the continuous sequence of reference waters was measured repeatedly between the ice core measurements as discussed in the text.

2- I am surprised by the answer concerning the cleaning procedure. Only a cleaning with water and soap of the fittings permit to remove problem of air and water flow restriction.

During the measurement study, we used only soap, water, and physical agitation to clean the vaporizer tee. We have since adopted a simpler procedure that includes soaking the tee in lactic acid, but we choose not to describe this in the manuscript since we did not use that procedure during the study period.

3- I understand that showing the 9 individual series will display a high noise and this is fine. Still, it would be very useful to show them (in appendix ?) because there is no way that CFA measurements can ever be done in 9 replicated barrels which hence questions the usefulness of this analysis. Do you think that measuring 2 or 3 replicate barrels may be enough to have a useful stacked D17O signal ? THis should be added somewhere in the discussion of the manuscript.

We do find that stacking measurements of Δ17O is advantageous both for increasing analysis time (and therefore improving precision and/or resolution) and for reducing the calibration error. We have added a section about this to the discussion. We have also included the spread of data for three averaging times in the new figure 5; figure 5 shows the spread of D17O_adjusted and D17O_calibrated so that the seasonality and total variability at each resolution are clear.

Editor's comments:

Comments to the author:
Dear Authors,

Thank you for submitting to AMT. Both reviewers agree that your revised manuscript provides significant science, but still judge that clarity (presentation and scientific) of the paper is only fair. Following the reviewers' recommendation, I hereby decide that the article can be published subject to minor revisions, but require that the reviewers' comments must be taken into account in order to improve the clarity of the article. In addition to these suggestions, additional minor and technical issues aiming at improving the clarity of the paper are listed below. This concerns particularly the offset correction or mean-subtraction-method, which is difficult to understand. I therefore suggest adding an appendix on this subject. Please mention changes to the manuscript not only in the 'tracked version', but also in your comments on the reviewers' and editor's remarks. This will greatly facilitate the task of following up your corrections. Please also check style guidelines for AMT (https://www.atmospheric-measurement-techniques.net/submission.html#manuscriptcomposition). In particular, please follow the house standard not to hyphenate modifiers containing abbreviated units (e.g. "3-m stick" should be "3 m stick"). This also applies to the other side of the hyphenated term (e.g. "3 m long rope", not "3-m-long rope"). e.g., in lines 110, 125, 126, 127 and many more.)

Hyphens have been removed from modifiers containing abbreviated units. The calibration adjustment method has been formalized within the manuscript text.

l. 12. Please add the resolution required to reach the 5 per meg or give another uncertainty-resolution pair that you find more useful for the reader.

We updated this statement to indicate that 5 per meg precision is attainable after 3000 s of analysis time.

l. 116-117. I suggest to remove parentheses here

This text has been updated as suggested.

l. 155-157. Please check punctuation to enhance the logic of the phrase. I suggest replacing the first comma by a full stop and the semicolon by a comma sign.

This statement was rewritten for clarity.

l. 196-198. It seems that pulsating flow is causing the observed anti correlation and insufficient back pressure might cause/facilitate these pulsations.

This section has been reworded for clarity.

l. 202. '10s to 100s' may be confused with 10 s to 100 s. Therefore, the notation '10s per meg', etc. for 'several 10 per meg' should be avoided.

This has been rephrased to avoid confusion.

l. 261 etc. One referee has asked for pointing out the fraction of rejected data and your response has indicated that about 50 % of the analysis time corresponds to acceptable measurement conditions. It seems, however, that the corresponding number is missing in the changed manuscript.
This detail has been added to the text in line 277.

l. 330. replace 'suboptimal' by 'eventually biased'

We changed ‘suboptimal’ to ‘biased’ as suggested.

l. 343 and before. Until here, calibration issues may lead to biases in intercept and slope. It is only shown much later that intercept biases are likely the culprit. You should explain why discussing the slope is neglected in the paragraph beginning at l. 343.

Calibration issues might be attributed to the slope or intercept (and likely both), though a correction to just the intercept can account for the offset. It is important to consider the influence of all calibration information, so we have instead removed references to the calibration intercept, and have added text to clarify the possible causes of this error.

l. 343 ff and 630. Can you clearly demonstrate that the shift technique (removal of the CFA-CRDS mean value) removes only offset b effects and not m etc ? I suggest to formalize the procedure mathematically and present it in an appendix.

Thanks for this comment. We have formalized the procedure mathematically within the manuscript text, which has greatly simplified all references to the two data treatments throughout the manuscript. Though the shift can be achieved by changing only b for both delta values, the error itself is inherently caused by effects of both m and b (and specifically the relation between m or b for d17O and d18O), so we have reverted to calling this the “calibration offset error” and removed any attributions to the intercept alone.

l. 343-346. This paragraph is confusing and the description does not make clear how the data has been processed. One possible source of confusion is using the term intercept when talking about the $\delta17O$ and $\delta18O$ data, but the mean value removal likely concerns the $\Delta17O$ data. This should be stated more clearly.

We have clarified how these two things are related by defining the $\Delta17O$ calibration adjustment based only on D17O values and also by defining it similarly to how it has been treated by others (i.e. as a function of d17O and d18O). Adjusting the final value to an accepted value of $\Delta17O$ can be achieved by shifting $\Delta17O$ directly or by shifting both component values of $\delta17O$ and $\delta18O$ to their accepted values. Because the accepted values for D17O have corresponding accepted values of d18O and d17O, it does not matter how the offset is defined – nudging d17O and d18O to their accepted values will also correct D17O, but it is not necessary to involve d17O and d18O to make this adjustment. Therefore, we have defined the simpler mean adjustment on D17O only and use this definition in our discussion of the offset error.

l. 600 etc. Please use lowercase a, b, etc in the figure caption to comply with figure content. Understanding could be enhanced by indicating that scales in subfigures are not always the same.

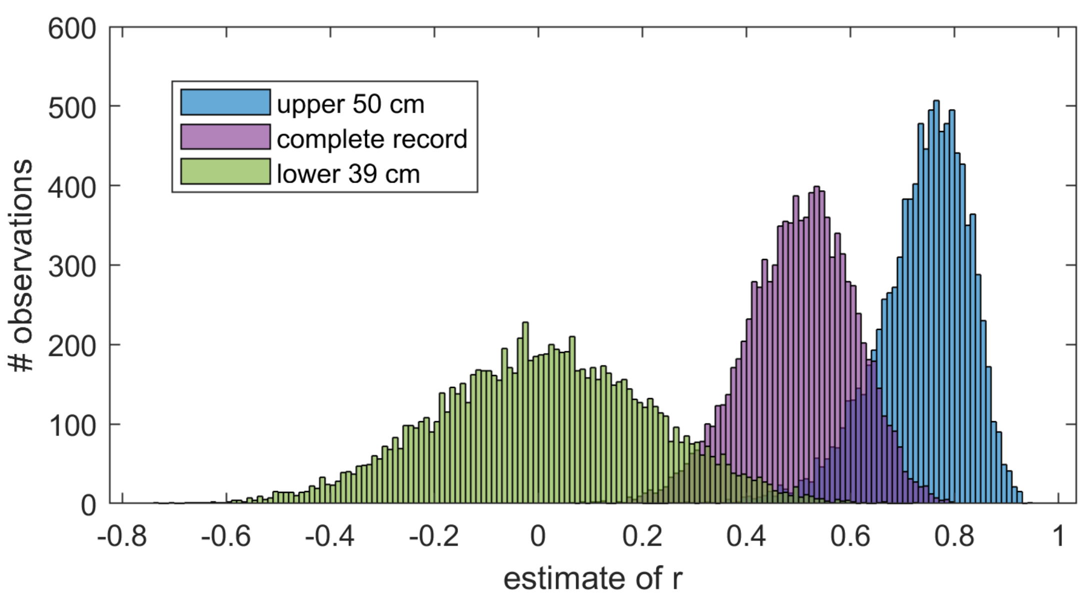
All figure captions that include labeled panels have been updated to use lowercase letters, and we have noted when subpanel scales differ.

l. 608 etc. As before, please harmonize capitalization of labels between legend and figure.

All figure captions that include labeled panels have been updated to use lowercase letters.

l. 318 + 319. The statements on the correlation coefficient (r=0.52 with 99% confidence) or (r=0.70 with 99% confidence) are confusing. Since r has been estimated from the sample, it has been determined with an uncertainty and some range of r-values should be associated to the 99% confidence interval. But no such range is given -- or do r-values given with two significant digits mean that the range is smaller than ±0.005 ? The uncertainty of r is particularly interesting for assessing the significance of the difference in r between the upper 50 cm of the core and the rest of the data.

Thank you for pointing this out; we agree that this was confusing. We have updated these statistics with the range of r estimates from a resampling experiment and listed them at a 95% confidence interval in the text. Additionally, the figure below captures the full range of estimated r values and demonstrates that in the upper portion of the core where the seasonal signal is most prominent, the range of r values is substantially different from the lower section of the core where the D17O signal is relatively flat in both discrete and CFA datasets.



l. 396. After calibration, this should likely be accuracy and not precision.

We updated all language after the data is calibrated to refer to the accuracy, error, or variability instead of the precision.

l. 515 and Table 1. All instances of 'dxs' should be replaced by the symbol d for deuterium excess, both in the axis label and in the figure caption.

All instances of dxs have been replaced by the symbol d.

Table 1. It is very surprising that δ17O is given with more significant digits than δ18O. Usually, measurements of δ18O are more accurate and thus require more digits than corresponding δ17O values. Note that a better than 3-digit precision is necessary to obtain a 1-digit precision in Δ17.

Our data have been normalized to the VSMOW-SLAP scale using reference waters that have been analyzed against VSMOW, SLAP, and GISP, as described by the table caption. Our data for δ17O and δ18O are presented to the number of decimal places suggested by Schoenemann et al., 2013 to accompany measurements of Δ17O.

Figure 4. The definition of m differs from that in equation (3). Please explain, or better, use m-1 in per meg as ordinate axis label.

The definition of m in Fig. 4 (now Table 2) is the same as in Eq. 3. m is the slope of the calibration equation and is therefore unitless (carries units of ‰/‰). We replaced Fig. 4 with Table 2, but the units of m were not changed.

l. 618 Figure 6. indicates that the calibration intercept error (CIE) increases with increasing measurement resolution. Is there a physical reason behind this ? Has this been discussed in the text ? Better understanding the behavior of the CIE probably allows to improve the technique further.

Thanks for this comment. We have added a section to the text to address the apparent change in calibration error with depth resolution. Definitionally, the calibration offset error is determined only by the mean value of the CFA measurement and its offset from the accepted value and therefore it should not depend on the depth resolution. The grey shading in Figure 6 indicates the full range of possibilities for the total error in the dataset and also for the total error in the calibration-adjusted dataset. Additionally, at very small averaging times, the calibration offset noise is essentially indistinguishable from the instrumental noise (see new figure 4), so accounting for the offset does not improve the data quality as notably as it does when other sources of error have been minimized.

l. 619. Replace 'by' by 'as a function of'.

Changed as recommended.