Thank you for this detailed and instructive feedback. We have made several changes to the manuscript to address these comments; changes to the manuscript are noted below in our response.

The manuscript describes an analytical set-up to measure D17O f water y continuous flow analysis on ice core samples. The performances of the system are evaluated by comparing the results on parallel ice core barrels taken at the same depth. The authors also describe the different effects that influence the stability of the results and hence the final uncertainty on the D17O measurements. This technical paper is useful even if it is not clear how it can directly be applied to routine measurements of D17O by CFA because the measurement time would be very long and it seems that many adjustments (or cleaning) should be performed during the period of measurements.

I detail my comments below:

- l.71: You mention that diffusion and mixing should be reduced for D17O but it is not the case for all isotopes. Why is it more important for D17O ?

As noted in the text, the magnitude of variability of D17O is substantially smaller than that of other isotope quantities (in this ice, 45 per meg versus several per mil). Additionally, D17O (by definition) has a nonlinear response to changes in d17O and d18O. Both are reasons to attempt to prevent mixing within a system when measuring D17O. Mixing should be limited when possible to achieve the highest fidelity measurements for any isotope analysis. We have clarified that reducing mixing is important for all ice core measurement systems.

- Paragraph from l. 92: from this reading, it seems that the authors need to continuously adjust the system during measurements which makes it quite complicated and it is really difficult to understand what is done exactly. Could the authors be more precise on how they detect the problem and what action they take. Some concrete examples would be helpful. Also if these adjustments should be done continuously, how is it possible to make long runs with a good stability ?

Typically, system adjustments are made several hours in advance of measuring an ice core, so that the reference water measurements that are closest in time to the ice core analysis are measured under the same conditions. It is possible to make long runs with good stability for days to weeks, though the exact timing of interventions is dependent upon the performance of system components (filtration, debubbler) and the quality of the water passing through them (i.e. due to particulate loading or mineral precipitation within system components). We have clarified this in the text.

- What is the difference between the experimental set-up presented here and the one used in the previous study (Steig et al., 2021) ?

The system used in this study has not previously been utilized or published and was designed for this study. The system used by Steig et al., 2021 is the system whose details are published in Jones et al., 2017, as discussed in Section 2.1.

- l.137: is there any mixing linked to the melting of the ice on the melt head ?

Though it is likely that some amount of mixing occurs at the melt head, we do not observe any mixing effects at the averaging times or representative depth intervals used for this study.

- "Uemera" should rather be "Uemura" (several occurrences)

Thank you for noting this; we have corrected all instances of this name in our manuscript.

- l.175: It is not clear how the system components are cleaned ? Which components ? How is the cleaning done ? It seems that the cleaning occurs very often and I am wondering how this can be done without affecting the continuity of the measurements. It is very important that the authors explain there cleaning procedure and especially how it is done while the measurements are being performed continuously.

We have clarified this in the text; vaporizer fittings were cleaned four times during the measurement window with soap and water. Cleaning occurred between measurements of reference waters, not during or adjacent to ice core measurements.

- l.188: similarly, what is meant by "routinely clean" ? As this seems to be an essential aspect of the measurement technique, this should be detailed and explained to be able to understand and evaluate this technique.

As above, we have better explained this method and frequency in the text.

- l.210 and following paragraph: The sequence of measurements of reference waters is not given. When are these waters measured ? How is the measurement of these waters organized with respect to the measurement of the ice cores ? Do you measure the 3 waters every day ? Every week ? A table explaining the sequence of measurements (reference water, ice cores) over the 7 weeks should be given.

The three reference waters are measured repeatedly in a continuous sequence over the period of seven weeks, as described in Section 3.2. The only times that reference water measurements are not being made are when 1) an ice core is being measured or when 2) after a period of multiple days, the vaporizer operations destabilized and personnel were not in the laboratory due to COVID-19. This has been clarified in the text, and we have included new figure 5 to make this sequencing clear.

- l.221 and following paragraph: I understand that the authors did an alignment of the d18O seasonal cycles for this study to match parallel records but for real CFA measurements, when there is a need to have access to the real depth, this technique is not adapted and I do not see how you can avoid measuring the evolution of the height of the melting ice barrels. In this paragraph and in general in the manuscript, it should be made clear what is done specifically for this study with the aim to estimate uncertainty from compilation of record at the same depth (where absolute depth record is not needed) and what is done for a routine CFA measurement (where absolute depth record is needed).

We agree that routine measurements of unknown ice samples require automated depth registration and we recommend strategies similar to Jones et al. (2017) or Bigler et al. (2011). For this study, we use visual observations of melt rate during analysis and later assign high-resolution depth assignments based on the cross-correlation of d18O with the discrete d18O data (of known absolute depth). We estimate uncertainty based on the variance of depth data at inflection points in the timeseries, which we have added to the manuscript. This paper does not describe a routine CFA methodology, but here we reiterate the importance of electronic level sensing equipment to establish a high-resolution depth record for routine operations.

- Section 2.6: again the calibration sequence is not clear. A table is needed to explain what is done every day and over the 7 week period. What is exactly done during the 3 hour measurements of reference water ? Also the range of acceptable mixing ratios is very large (20,000 – 50,000 ppmv). Do you really observe so large variations in a short time ? over what time period is estimated the sigma_18O of 0.5 permil ?

We have revised the calibration section to clarify this process, and have added the complete sequence of measurements in Fig. 4. We clarify that system water vapor concentrations typically only fall below 20,000 ppm or above 50,000 ppm during maintenance downtime or when an operator was not available to restabilize the system during the COVID-19 pandemic. The screening criteria of 0.5 per mil is applied to each three-hour measurement window for the reference waters.

- Section 3.1: It would be very useful to see the raw data instead of only the average and envelopes of the records. This would help understanding the correlation (r=0.52) which I do actually not find as "high" as written by the authors. It would also help to understand the difference between the different sections of the core.

We have clarified that the r=0.52 correlation is significant with 99% confidence. We choose not to show the individual measurements at 1.39-cm resolution in the paper because, as expected, the noise at this resolution is too high for this to be useful (~25 per meg from the Allan variance). We include here a plot of all individual CFA measurements to demonstrate the observable seasonality within this population.

In this figure, we show the mean of all measurements and the standard error envelope, as shown in the manuscript. The grey dots are all discretized CFA measurements, each representing approximately 270 s of data.

$\Delta^{17}O$ for JEMS2#89
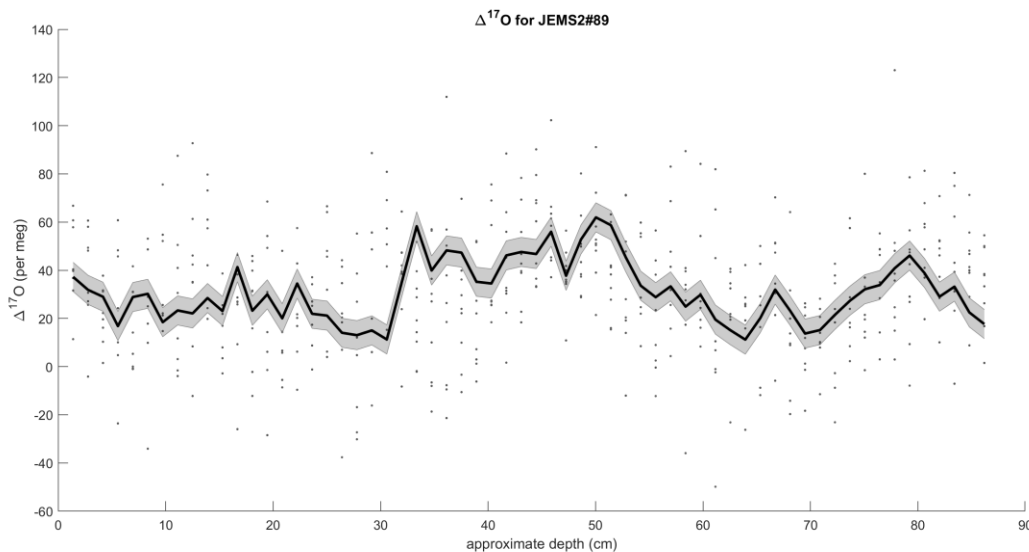
- l.292: it is impossible to understand exactly what you mean by "disproportionate drift in d17O and d18O" since no number is given nor any example. What amplitude of the drift ? Over which period ? What should be done to avoid this ?

We have reworded this section for clarity and have added all calibration information to Fig. 4 to demonstrate the magnitude and timing of the calibration drift. We have added recommendations for establishing a calibration strategy to the discussion section.

- Paragraph starting from l. 303: again this is very difficult to follow if we do not see the raw data.

We have added a new section to more fully explain the calibration process, and have better explained the treatment of this data by relating it to the calibration process.

- l.326 and following paragraph: Please show the raw data and then the different treatments + explain exactly what you mean by calibration so that the reader can understand what was done. Additionnal figures showing the different step as well as the measurements (raw data) of reference waters used are needed to understand what you mean by calibration.

We have added a new Figure 5 and Equation 3 to visualize the calibration components and better explain the calibration process. We have also rewritten this section with more specific language to emphasize the impact of the calibration intercept. All CFA-CRDS data is provided in Figure 5.

- At the end, we expect clear recommendations on how to perform routine CFA measurements to achieve a good D17O signal. So we would need a recommended sequence of measurements and calibration with numbers given – a table is recommended. Also please explain how you suggest that calibration should be done. In the present state, the manuscript is not really useful for the reader who wants to repeat this set up.

Our work shows that D17O can be detected by CFA-CRDS, but that a robust calibration strategy and stable system operations are both important to achieving good data. We have added clearer recommendations to the discussion section for those who are interested in measuring D17O by CFA-CRDS, though we cannot recommend a specific formula for CFA methodology because, in practice, it will depend on the full suite of analyses that are desired. For example, while slower and/or repeated measurements are advantageous for D17O, for gas measurements like CH4 the opposite is the case (more rapid melt rates are better and duplicate measurements are generally not possible because of large volume requirements). This particular configuration performs well for D17O, but it is likely that other established methods for measuring water isotopes by CFA are also sufficient as long as the calibration is accounted for.