The paper by Davidge et al. presents a detailed and systematic investigation of the continuous flow analysis (CFA) method in terms of the reproducibility of \( \Delta^{17}O \) measurements of ice cores using CRDS. They address remaining open questions related to observed fluctuations and offsets, and propose a recipe for an optimal CFA operation. The topic is highly relevant for the ice core research community, especially for those deploying the CFA-CRDS method. Therefore, I recommend it for publication after some clarifications and corrections are made.

**General comments:**

The abstract needs some adjustments to better summarize the main topic of this work, i.e. optimization of the CFA method, intercomparison of CFA-CRDS with discrete sample CRDS measurements, error analysis and noise contribution, etc. The authors should refrain to use strong wording such as "routine measurements", because if that would be the case then their current work would not be relevant. Also the claim of high-precision measurements of \( \Delta^{17}O < 5 \) per meg is questionable. Considering the reported 0.3 cm/min melt rate, the 1.4 cm resolution would correspond to about 270 s measurement time on CRDS that would result in about 25 per meg precision, in best case (see Allan deviation plot). The precision of ~5 per meg can eventually be achieved at averaging times longer than 3000 s, which corresponds to depth averages of ~15 cm. Obviously, more clarification is needed here.

Another aspect that requires further discussion is the "calibration errors". First, it needs a clear definition and second, the interpretation of the data should be reconsidered. The authors present a large number (>40x, 3 h, over seven weeks) measurements of the reference waters (SW, SPS2, and CW), but they don't say/show anything related to the CW, although this is supposed to be used as an independent verification of the calibration. Also, the information about the two-point linear calibration curve (offset, slope and their variation across the individual measurements) is not given.
Furthermore, a detailed description of the many limitation and critical role of a constant flow across the CFA is presented, but then the calibration gases are used on much lower flows compared to the ice core measurements. It is expected that changes in the flow would have a significant impact on the droplet formation in the vaporizer and that adsorption, memory effects, and isotope fractionation can also be changed. Considering all these details, it is difficult to see any direct and easy link that would allow for the conclusion drawn by the authors about the dominating role of "calibration error" in the CFA-CRDS method.

Finally, the direct comparison of the CRDS stability (Allan variance) with the CFA reproducibility on Fig 7 is questionable. The Allan variance assumes a continuous stream of data, condition that is not fulfilled by the discrete measurements. Plots of the two-sample variance from different measurements based upon successive recordings of samples display a 'shoulder' effect of the reduced duty cycle on the system performance and the first deviation from the $1/\tau$ slope is a duty cycle effect and not an indicator for an accuracy (calibration) problem.

**Specific comments:**

Pg4, l.94. The authors should quantify the "several times more vapor".

Pg4, l101. A pressure sensor monitors pressure and not flow conditions.

Pg4, l111. Is there no issue (adsorption, memory) with having water carried over PFA tubing? Why do not use electro-polished stainless-steel tubing with inert coating instead?

Pg5, l130. The expression "analytical cavity" is not correct. I suggest using "optical cavity". Check for all instances across the manuscript. Also the "required sample volume" is not appropriate. The volume of the optical cavity is fixed as well as the pressure at which the CRDS operates.

Pg5, l131. What is meant by "measured volume"?

Pg5, l134. Replace "all system instrumentation" by "sample handling system"

Pg5, l137. Consider simplifying the wording and replace "system instrumentation volume and system tubing diameter were minimized" with "the overall sample handling system volume was minimized"
Pg5, l140. Quantify the "excess liquid". How much compared to the measured sample?

Pg5, l143. Quantify "several times more"

Pg.5, l145. The pump rate is minimized when measuring reference water. The authors should specify by how much is the flow reduced and comment on the expected effects due to the change in flow especially in terms of droplet formation in the vaporizer, adsorption, and isotope fractionation (see also my general comments).

Pg.7, l198. The anticorrelation between water vapor and δ18O is explained by two different effects: 1) incomplete vaporization, and 2) insufficient backpressure. The authors should be more consistent and clearly state which one applies. Is the correlation a real physical effect, due to e.g. isotope fractionation, mixing, etc., or is it simply reflecting the δ18O dependence on water vapor amount fraction of the CRDS?

Pg.7, l201. What does the "apparent fractionation" means? An instrumental response?

Pg.7, l205. The authors associate flow inconsistencies with isotope fractionation, but it is a misleading argument, because without excluding artifacts from the CRDS measurements itself it is difficult to disentangle the observed effects. In other words, the authors should discuss the mechanism behind the isotope fractionation generated by the variations in the flow.

Pg.7, l205. There is a list of many significant interventions: adjusting the peristaltic pump rate, replacing filter screens, adjusting FV-1, replacing peristaltic pump tubing, replacing or cleaning the capillary tube, or cleaning the vaporizer. The authors should give a more detailed discussion how often are these interventions necessary, what does it mean in terms of operation down-time and what is the impact on the calibration scale. Changing so many items should definitely result in rather different system response in terms of memory and surface effects, etc.

Pg.7, l210. The authors should explain how they are able to perform the automatic measurements while routinely tune the system to maintain steady pressure.

Pg.8, l221. What is a manual observation?
What is the amount of rejected data compared to the entire set of measurements?

It would be very helpful to provide the scatter of the 47 individual 3 h measurements of the reference water to illustrate the stability of the CFA-CRDS system for this fundamental step.

This sentence is misleading. I suggest to modify it, e.g. "our ice core measurements cover of about two years period".

In contrast to the author's statement, a correlation coefficient of 0.52 could either be interpreted as a "good" or "moderate" correlation, depending on the applied rule of thumb. The observed correlation may not necessarily be a good estimate for the population correlation coefficient, because samples are inevitably affected by chance. Therefore, the observed coefficient should always be accompanied by a confidence interval (95 %), which provides the range of plausible values of the coefficient in the population from which the data were sampled. Furthermore, the correlation coefficient of 0.52 corresponds to a coefficient of determination ($R^2$) of 0.27, suggesting that only about 27 % of the variability can be "explained". Finally, since both data are observations, a Pearson correlation analysis would be more appropriate here. In this case, both variables are assumed to be subject to natural random variation.

Again, this statement is inappropriate since the deviations are comparable with the seasonal variation. It would be instructive considering a scatter plot using the CFA and discrete CRDS data shown in Fig7 (bottom).

Although, this in principle holds, the slower ice melting would result in slow response time across the CFA, which would then have an impact on the achievable resolution. In general, instead of hypothesizing what would in principle be possible, the authors should consider only those cases that are realistic for high precision and routine ice core measurements.

The term of high-frequency instrumental noise is misleading. Use 1 Hz precision instead.

The authors should explain what they mean under calibration error and what are the disproportionate drifts in $d_{17}O$ and $d_{18}O$. Is there any systematic investigation of the oxygen isotope fractionation during vaporization? If yes, it would be helpful for the reader to know its magnitude, reproducibility and dependency on various factors, such as flow, pressure, etc. Without these facts the claim cannot be proven.
The Allan variance analysis doesn't determine the theoretical variability, but the observed one.

Define "internal noise".

This statement is not clear enough. If the reference water is handled in the same way as the melt water then how is it possible that its measurement does not account for the variability from the CFA system?

Again, the 1.39 cm resolution would correspond to 270 s measurement time leading to about 25 per meg precision (1σ) according to the Allan variance. Thus, the signal-to-noise on the seasonal cycle is less than 2.

Fig. 7: The Allan variance plot has a remarkable character. It seems that the CRDS stability is extraordinary as the deviation continues to decrease even after $10^4$ second (2.8 h). The authors should comment on why do they stop at this stage and consider the 2 per meg as best precision. It would be very interesting to see how long this continues. The authors should perform even longer reference water measurements to explore the limits where drifts start to dominate. At such instrumental stability, there is no need to make any calibrations for hours. This would imply that a 50 cm ice core could be easily measured in one run. Why this is not shown in this work?

There is a "bump" at 10 s averaging time. Normally, this indicate a periodic oscillation in the system. Can the authors comment on the origin of this deviation from the white-noise at that time-scale?

Technical comments:

Abstract l.8-10: I recommend combining the two sentences and remove the "recent advances", because the CRDS technology for oxygen isotope measurements is almost 10 years old without any significant development since then. My suggestion: "... continuous-flow analysis (CFA) methods coupled to CRDS instrument allow for simultaneous measurements ...."

I also suggest deleting the last sentence in the abstract.

Pg2,l.30. Replace "routine" by "were demonstrated" since many of the cited references give a demonstration and show the feasibility of the technology rather than presenting
routine ice core measurements.

Pg2, l.32. Replace "laser spectroscopy instrument" with "a laser spectrometer"

Pg3, l.67. I suggest to write "CRDS spectrometer (L2140-I, Picarro Inc.)"

Pg3, l.84. Remove space between value and degree, i.e. 200°C (look for other instances in the manuscript as well)

Pg4, l.98. Replace "near-instantaneous analyzer output" by "CRDS values". The authors can assume that the reader is familiar with the CRDS and knows that this instruments report the results on a 1 Hz rate. I suggest removing all instances of "instantaneous".

Pg4, l.102. I suggest replacing the very vague sentence "Instantaneous instrument output that reflects the internal vaporizer conditions and monitoring of the CFA line pressures provide information that is used..." by "These information is used..."

Pg4, l.107. Replace the "x" by "×". For sake of simplicity write 30×30 mm²

Pg5, l.140. Change "ice-core" to "ice core". Check for all instances across the manuscript.