Atmospheric
Measurement
Techniques

Open Access

EGU

Discussions

Interactive comment on "Calculating uncertainty for the RICE ice core continuous flow analysis water isotope record" *by* Elizabeth D. Keller et al.

Anonymous Referee #1

Received and published: 13 March 2018

1   Overview

The manuscript by Keller et al. presents methods to assess the uncertainty of the water isotope ratio measurement for an ice core continuous flow analysis (CFA) system used for the measurements of the RICE ice core. This is a work that builds on previously published methods by Emanuelsson et al. (2015). The work focuses on a rather special but very essential part of CFA system for water isotopic analysis that of the uncertainty characterisation. It fits very well within the scope of the Atmospheric Measurements Techniques journal. I find it very positive that the authors decide to focus a separate manuscript for assessing the uncertainty of the water isotopic measurement. This is not a usual practice and it is most welcome.

Unfortunately though i cannot recommend the manuscript for publication. I would like to let the decision to the editor (based also on the assessement of the other reviewers) on weather the authors should proceed with a major revision of the manuscript or withdraw their submission and start fresh. The second option would actually be my recomendation. The reason for this decision is that almost every aspect of the manuscript is in my view inadequately developed or/and presented. The methods -particularly those used for the assessement of the total uncertainty- of the measurement are inacurrate, while the presentation of the results lacks clarity. Additionally, it is troubling to see that based on the results the authors draw conclusions that I would argue are wrong. Lastly, nomeclatures with respect to water isotope analysis and SMOW/SLAP calibrations, some of the presented mathematical expressions as well as nearly every figure in the manuscript need modifications in order to come up to the standard of a journal like AMT. Some of these issues also indicate a possible misunderstanding of some of the concepts or tools presented in the manuscript.

In the following points I will try to explain my decision in more detail to the authors. I would also like to suggest possible approaches and techniques the authors could consider that hopefully can improve the quality of modified version of this manuscript or can be used for compiling a manuscript for a new submission. I would be happy to elaborate more on these ideas in case the authors are interested in using some of those in another version of their work.

2 General comments

2.1 Water isotopic standards

Accurate measurements of water isotope ratios require properly and accurately calibrated "local" standards with isotopic values on the SMOW/SLAP scale. Quality of

calibrations, storage and handling is essential for the quality of the measurements. I find it very concerning that in this manuscript the values of the two primary standards used for the SMOW/SLAP calibrations are signifficantly different between 2013 and 2014. The differene is more profound for δD and up to 5 ‰ difference between the two years. Looking carefully into the reported values in Table 1 it does not look like this large difference is due to evaporation (D_{xs} is roughly unchanged). No uncertainty estimation is given for the discrete measurements either. These differences are concerning and better addressing these issues with the standards is essential and describing them with clarity is necessary.

One very important point here is that there is a claim in the manuscript that the 2014 values are likely more accurate because they represent better the melting conditions of the CFA system since the calibration was carried out in continuous mode (as apposed to the 2013 standard measurements carried out in discrete mode). This is a claim that is fundamentally wrong. As long as an uncertainty estimate is obtained for a measuerement one should not expect to gain anything by calibrating the standards on the same system a measuremnt is performed. This wrongly drawn claim is given in several spots in the manuscript, including the main conclusions (P4line33, P5line3, P8line18, P9line13).

2.2 Statistically independent errors assumption

Equation 1 (equation 4 in the manuscript) is essentially a wrong approach in reaching a "total uncertainty" estimate for the system. This is the cornerstone of the uncertainty assessement in this manuscript. It is described as novel and the majority of the reaults and conclusions are based on the application of Eq. 1. Essentially Eq. 1 is the mathematical description of error propagation for the sum of three Gaussian–distributed variables. Equation 1 reads (in my review I will be using the σ notation for standard deviation and the ε notation for the variances as $\varepsilon = \sigma^2$; the manuscript mixes those

two symbols and definitions at several places; the authors should make sure they use a consistent nomeclature with respect to variances and standard deviations throughout the manuscript):

$$\sigma_{tot} = \sqrt{\sigma_{Allan}^2 + \sigma_{scatter}^2 + \sigma_{calib}^2} \tag{1}$$

Running the risk of sounding trivial this equation is based on

$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial X_i}\right)^2 \sigma_{x_i}^2 \tag{2}$$

Getting from eq.2 to eq.1 requires an important condition. This is that the variables of which the variances are used in eq. 1 are independent from each other. This can by no means be said for the variables given here. Putting in the same bucket an estimate of the variance of the isotopic composition (in some way what the $\sigma_{scatter}^2$ describes), the error of the mean (essentially what σ_{calib}^2 represents) and the Allan variance is not an option. All these three parameters are dependent on each other. Therefore I have a hard time seeing how any of the "total uncertainty" estimates in the manuscript is valid. Considering that the manuscript's goal is solely to give proper estimates of the uncertainty I hope the authors can see why I am inclined towards a fresh submission following a very different approach/method.

2.3 The calculation of $\epsilon_{scatter}$

This is a questionable calculation in the way it is performed. The quantity $\epsilon_{scatter}$ in the manuscript is essentially a standard deviation and not a variance so here I rewrite the equation using a standard deviation symbol and replacing the "mean" with what it means mathematically. As a result of this eq 8 in the manuscript becomes:

$$\sigma_{scatter} = \frac{1}{N} \sum_1^N \frac{\sigma_i}{\sqrt{n_i}} \tag{3}$$

C4

In eq. 3 σ_i is the standard deviation of every 15 sec interval. Firstly the 15 sec choice for the length of the interval is fairly arbitrary and in that sense it introduces a subjectivity in the calculation of $\sigma_{scatter}$. A very important question here that the manuscript does not touch upon at all is what is the influence of the data acquisition rate of the instruments. Calculating standard deviations over 15 s intervals for two instruments that have a different acquisition rate implies that there is a different amount of averaging in the calculation. Is then the difference in $\sigma_{scatter}$ due to actual lower measurement noise of the laser instrument or is it an artifact of the higher acquisition rate (it can also be a combination of these two parameters)? A technical paper like this should deal with these questions when its sole purpose is to characterise the uncertainty of the measurement.

The plots of the scatter error in Figure 5 look very unphysical to me. I suspect that this is due to the fact that you have calulated the scatter error based on the RICE standard blocks bracketing each run and assumed that this error should vary linearly for every data point in between. Even if your system does not have a stricktly Gaussian behavior, to assume that the error varies linearly from one RICE block to the next is by all means statistically wrong.

2.4   The interpretation of the Allan variance

The Allan variance is a great tool for accessing the stability of anaytical instrumentation. Sometimes unfortunately its meaning can be misunderstood. In the case of this manuscript it is misued. Firstly, Fig. 4 does not give any chance to the reader to get an insight on the stability performance of the system based on the Allan variance. It is a Figure with several plots all on grayscale with no legents, or caption explanations as to what is what. What am I looking at? Are these Allan variances of δD, $\delta^{18}O$ or $\delta^{17}O$? Which plot refers to the 2013 and which one to the 2014 system? The bottom black line shows an Allan variance that decreases almost indefinately. How do you fit that

one when there is no minimum? Why does this curve show such a stable behavior and what is the reason for it? In fact this is a grayscale plot directly taken from Emanuelsson et al. (2015) refering to $\delta^{18}O$ and in fact the bottom curve is from a completely different system that uses a different analyser and sample evaporation system from a different laboratory.

The interpretation and use of the Allan variance in order to calculate the error introduced due to instrumental drifts is falsely performed. The Allan variance gives an estimate of the maximum time one can decrease the error of the mean by averaging. After the optimal point, further averaging is either not helping or makes things worse. So to start with, in the system described in the manuscript there is no averaging of data for such long times (order of 600 s). I would agree that the Allan variance gives an estimate of how your accuracy may be affected by instrumental drifts when SMOW/SLAP calibrations are performed infrequently. This is though only a qualitative assessement as the Allan variance concerns the averaging of data in order to reduce the error of the mean of a variable.

Based on the way that calculation of the Allan variance fit is described and the text "where t is the time difference between the data point and the calibration" I understand that the starting point of your linear fit (where $t = 0$) is the optimal time where the minimum of the Allan variance is located. The error that takes place here is twofold. Firstly for $t = 0$ based on eq. 7 in the manuscript $\varepsilon_{Allan} = 0$. This is not possible. Secondly by starting your fit at the minimum Allan variance you have completely neglected all the first part where averaging in fact does reduce the error of the mean. With this in mind the parabola shape of the Allan error is wrong. It essentially suggests that the error increases monotonously from the time the calibration is finished until the point right in between the two calibrations, when it starts decreasing again monotonoulsy. One has to expect that all measurements that are taken sooner than the optimal time (Allan variance minimum) in fact benefit from averaging. One additional comment on Fig. 4b is that either because of a typo or a miscalculation the legend is wrong. If the

scatter noise for $\delta^{18}O$ is a factor of ≈ 3 lower than the δD signal, then it is physically very difficult (i would say impossible) for a measurement system that measures those two parameters practically simultaneously to result in an Allan variance that is lower for δD.

2.5 The explanation of the calibration protocol

The explanation is rather poor. The terms calibration and normalisation are used wrongly in the manuscript. The term scale normalisation actually refers to the slope. The term calibration refers to the intercept (the term intercept is actually never mentioned throughout the manuscript). Please give the general equation of a calibration and avoid writing terms like "slope" or "$RICE_{true}$" in a mathematical equation. It makes the reading of math formulas very difficult. "the normalisation correction is the measured mean of the RICE standard" Here you probably mean the calibration correction ie line intercept is the value of the RICE standard. This is actually wrong and surprisingly it is not even supported by eq. 3 the way you write it. For a line calibration $\delta_{cal} = a\delta_{raw} + b$ you can solve for b to get your calibration value.

In eq. 3 you define what $step$ is (use directly t_1, t_2 directly in the equation or replace $step$ with a symbol) but it is unclear what t exactly means. Yes t is the time of the measurement. What units? Is it sec? Using eq. 3 I equalised the two calibration terms from the two different calibrations and solved for t after substituting t_1 and t_2 in the equation. So the time at which both calibrations are considered equally in the weighing scheme of eq. 3 (this time should fall precisely in the middle between the two calibrations) is $t = 1/2$. Is this 0.5 sec or?? You are likely confusing rather than helping the reader with this type of errors.

2.6 The quality of the plots

Nearly all plots are of poor quality. There is no consistency with respect to line coloring (please choose one color for δD and one color for δ^{18}O and be consistent with your choice), the two parameters (δD and δ^{18}O) often share a common axes resulting in the very odd "δ" label, measurement units are sometimes in parenthesis (as they should) and sometimes they are placed right next to the δ symbol with no spacing, while font sizes vary between the different figures. All the plots where the three different errors are presented (fig 4, 5, 7) use the δ symbol on their axis when the plot actually presents something else. Only fig. 7 shows the full record while fig. 4 and 5 cover shorter section (and not the same) A plot of the calibration error (accuracy) is notably missing from the manuscript. Presentation quality is key for a manuscript of this type and I consider those issues with the manuscript's figures a major drawback.

2.7 On instrumental drifts

I wish to make a short note on the topic of instrumental drifts. Instruments can indeed drift with time sometimes in a nicely linear way that can be corrected for. However often one looks into the combined effect of the laser instrument, the sample preparation system as well as the protocol of the measurement. Sometimes in fact, measures to get a handle on the instrumental drifts can actually "create" those drifts in an artifactual way. The injection of a "check standard" or "drift standard" in frequent intervals can in theory offer insight into the nature of the instrumental drifts and possibly allow for a correction. The danger though is that the very same standards create those drifts as they are injected for a time interval that is too short thus not allowing for a stable value due to memory effects.

The most notable misconception with respect to instrumental drifts is that one calls problems as valve wear or drill fluid contamination or leaks as instrumental drifts. They

are not. If your system is leaky at some point in the line you simply have an unsystematic error for which the Allan variance for example cannot say much about, neither can frequent calibrations be of any help. Some of these claims made in the manuscript are purely speculative. Have you got any evidence that the drill fluid causes a spectroscopic interference in the wavelengths you are measuring with the spectrometer or is this something that you just mention in the absence of any other information or guess on sources of error? There is not even information on the type of drill liquid throughout the manuscript. Claims on those possible sources of error can also be found in the last four lines of the abstract. There is no single sentence, or data plot during the whole manuscript providing any (data or physics based) supporting evidence for these claims.

2.8 The depth registration

There is absolutely no comment on the uncertainty of the depth registration. Performing CFA measurements on ice cores is a tedious process. It is in fact mentioned in the manuscript that melting was at times interrupted because of sections of poorer ice quality. It is vital to at least comment on errors on the depth scale if this is to be a manuscript on proper error estimation of this analysis. Information on the melting rates used and even a rough estimate on how they vary throughout the measurement is notably missing from the text.

2.9 Results, discussion and conclusions section

Here are some more specific comments regarding the results, discussion and conclusions sections.

- It is mentioned that the calibration error is greatly reduced in 2014. No discussion can be found regarding the reasons as to why this is the case. A discussion on

the technical differences between the 2013 and 2014 systems is notably missing. Are the two instruments operating at the same wavelength region? Is there better control of cavity pressure and temperature? Are there changes in the sample transport and evaporation of the isotope line in the CFA system?

- The mean Allan error cannot be higher for $\delta^{18}O$. If the variance of δD is higher then the Allan variance follows similarly.

- I have a hard time to see how the scatter error is higher for the 2014 system. Based on the Allan variance plots in Emanuelsson et al. (2015) the noise in the data is lower for the 2014 system for both δD and $\delta^{18}O$. Why do you not plot the scatter error for the full record?

- The calibration error is of course larger than the scatter error. In the ideal case of a Gaussian distribution and assuming that your scatter error represents the standard deviation of the distribution anything within 2 even 3σ should be considered as acceptable. If your scatter error is say 0.1 then an accuracy of 0.2 even 0.3 can be expected. This is another indication for arguing that adding the errors the way you do with eq. 4 is problematic.

- It makes very little sense to say that calibrating the water standards with the CFA system would result in a better calibration result because you simulate closer the operating conditions of the measurement. A well calibrated standard is independent of the measurement system. It is the quality of your system, your calibration protocol as well as the storage and handling of the standards that will allow for a good result. This statement also appears in the conclusions and can be very misleading to the reader.

- "...so one explanation for the relatively large error is that our drift correction is poorly adapted to this upper range. Ideally, we would use a quality-control standard that falls between the values of our two calibration standards, RICE and

C10

ITASE". All the drift correction does is weigh the influence of two neighbouring calibrations thus if you have a relatively large calibration error then it is your calibration protocol that has an issue. In this sentence you are also speculating again. What makes you think that your drift correction is poorly adapted to the upper range. Show some evidence. What I am wondering about when I am reading this is that you do have such a standard. If you use WP1 and ITASE for your slope/interscept the RICE will be the middle check standard also falling very close to the measurements' level. So the data is there. Adjust you calibration scheme to use WP1 and ITASE then use RICE as check.

- "The overall system performance became more variable in 2014". Looking in fig 7 it is clear that this statement is false. Your total error has fallen considerably for 2014.

- There is no single piece of evidence presented in the manuscript supporting that ice breaks, drill fluid contamination and leaks or valve degradation are reasons for poor quality data. The drill fluid would likely cause considerable effects in the form of outliers if there was a spectroscopic intereference. Have you looked into the instruments spectra and the quality of the spectral fits (fit residuals are usually saved in laser spectrometers)? All of your error metrics (Allan, scatter and calibration error) are based on some assesement of the RICE and WS1 standards measurement or mq water. Thus how do breaks in the ice affect you total error? Leaks and poor pressure control can be checked by looking into the pressure log of the instrument. Do you have any indication based on these data?

3 Proposed improvements

Here I outline some proposed improvements and changes in methods used that the authors are welcome to consider in a new version of the manuscript.

3.1 Spectral estimate of the noise level

You can get a precision/noise estimate **directly from your data**. Calculate the power spectral density of a section of your data (say 10 m long). It should slowly decay reaching some sort of plateau for the very high frequencies. Take the level of this "tail" and integrade over the full range of frequencies (the surface under a straight line from $-f_{Nyquist}$ to $+f_{Nyquist}$ will be good enough.) The square root of your integral gives you a direct estimate of the noise level from you data. Do this for the full record. Be aware about the type of spectral density you calculated (single/double sided)

3.2 Measurement accuracy on WS1

The values you ontain for your accuracy based on the WS1 measurements should be presented for all your calibrations. Make a plot with all of them.

3.3 The calibration stability

Give statistics and plots on the values of the calibration line slope and intercept. This is a very interesting statistic and it is great you have all these data. Consider switching to a scheme where WS1 and ITASE are you calibration standards and RICE is your check. Compare to the current scheme where WS1 is your check.

3.4 A section on instrument/system differences

Consider a well written section on the instrumental differences between 2013 and 2014. You have a plethora of data on the instruments. Look into spectral fit quality, cavity pressure and temperature for indications of issues with these parameters. Explain better the differences between the two instruments and make sure you understand

well the impact of the higher acquisition rate of the 2014 instrument.

References

B. D. Emanuelsson, W. T. Baisden, N. A. N. Bertler, E. D. Keller, and V. Gkinis. High-resolution continuous-flow analysis setup for water isotopic measurement from ice cores using laser spectroscopy. *Atmos. Meas. Tech.*, 8(7):2869–2883, July 2015.

AMTD

Interactive comment

Printer-friendly version

Discussion paper