Improving memory effect correction to achieve high precision analysis of $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess and d-excess in water using cavity ring-down laser spectroscopy

Christine Vallet-Coulomb, Martine Couapel, Corinne Sonzogni

► To cite this version:

Christine Vallet-Coulomb, Martine Couapel, Corinne Sonzogni. Improving memory effect correction to achieve high precision analysis of $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess and d-excess in water using cavity ring-down laser spectroscopy. Rapid Communications in Mass Spectrometry, 2021, 10.1002/rcm.9108 . hal-03209687

HAL Id: hal-03209687

<https://hal.science/hal-03209687>

Submitted on 20 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Improving memory effect correction to achieve high precision analysis of $\delta^{17}\text{O}$, $\delta^{18}\text{O}$ $\delta^2\text{H}$, ^{17}O -excess and d-excess in water using cavity ring-down laser spectroscopy

Christine Vallet-Coulomb¹, Martine Couapel¹, Corinne Sonzogni¹

¹ Aix Marseille Univ, CNRS, IRD, INRA, Coll France, CEREGE, Aix-en-Provence, France

Abstract

Rationale: The precision obtained in routine isotope analysis of water ($\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess and d-excess) by Cavity Ring-Down Spectroscopy is usually below the instrument specifications provided by the manufacturer. This study aimed at reducing this discrepancy, with particular attention paid to mitigating the memory effect (ME).

Methods: We used a Picarro L2140i analyzer coupled with a high-precision A0211 vaporizer and an A0325 autosampler. The magnitude and duration of the ME were estimated using 24 series of 50 successive injections of samples with contrasting compositions. Four memory correction methods were compared, and the instrument performance was evaluated over a 17-month period of routine analysis, using two different run architectures.

Results: The ME remains detectable after the 30th injection, implying that common corrections procedures only based on the last preceding sample need to be revised. We developed a new ME correction based on the composition of several successive samples, and designed a run architecture to minimize the magnitude of the ME. The standard deviation obtained from routine measurement of a QA water over a 7 months-period was 0.015‰ for $\delta^{17}\text{O}$, 0.023‰ for $\delta^{18}\text{O}$, 0.078‰ for $\delta^2\text{H}$, 0.006‰ for ^{17}O -excess and 0.173‰ for d-excess. In addition, we provided the first $\delta^{17}\text{O}$ and ^{17}O -excess values for the GRESP certified reference material.

Conclusions: This study demonstrates the long-term persistence of the ME, which is often overlooked in routine analysis of natural samples. As already evidenced when measuring labelled water, it calls for consideration of the compositions of several previous samples to obtain an appropriate correction, a prerequisite to achieve high precision data.

1- Introduction

Laser absorption spectrometric techniques are more accessible and easier to operate than Isotope Ratio Mass Spectrometers (IRMS), which has led to a growing interest in producing water isotope data, as revealed by the increasing number of laboratories involved in the international proficiency tests organized by the International Atomic Energy Agency (IAEA) between 2011 and 2016.^{1,2} The most recent instruments offer the opportunity to measure the $^{17}\text{O}/^{16}\text{O}$ ratio of water in addition to the traditionally analyzed ratios of $^{18}\text{O}/^{16}\text{O}$ and $^2\text{H}/^1\text{H}$. The analysis of all three oxygen isotopes allows one to calculate the ^{17}O -excess value, defined as $^{17}\text{O}\text{-excess} = \delta'^{17}\text{O} - 0.528 \times \delta'^{18}\text{O}$ (with $\delta' = \ln(\delta + 1)$; all values in the decimal notation), similarly to the classic d-excess value ($\text{d-excess} = \delta^2\text{H} - 8 \times \delta^{18}\text{O}$). Regarding the natural variability of the ^{17}O -excess in meteoric waters,³ a measurement precision better than 10^{-5} (10 permeg) is required to allow the application of this new tracer in water cycle studies. This remains challenging in routine analysis, and appropriate data processing and measurement protocols are necessary to resolve the small variations of ^{17}O -excess observed in nature. Post measurement data processing involves compensation for instrument drift and memory effect, and VSMOW-SLAP normalization.⁴⁻⁷ Additionally, regular measurements of a quality control standard are required to ensure long-term reproducibility with high precision.

The so-called “memory effect” (henceforth ME), affecting isotopic measurements due to the carry-over of residual amounts of water during sample analysis, is an important pitfall that limits the precision obtained when measuring disparate samples under routine conditions.^{5,8-10} For the Picarro wavelength-scanned cavity ring-down spectrometers (WS-CRDS, Picarro, Santa Clara, CA, USA), the manufacturer specifies the magnitude of the ME, as a percentage of the final value, being better than 98% (for $\delta^2\text{H}$) and 99% (for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$) after the 4th injection, and recommends to simply discard the first few injections. Nevertheless, the ME remaining after the 4th injection may strongly affect measurement precision, especially when measuring samples with significant differences in isotopic composition, as the reference waters used for calibration, or even more when analyzing strongly enriched (labelled) waters. As an illustration, for a difference in $\delta^2\text{H}$ composition of ± 60 ‰ between consecutive samples, which is a typical value encountered when measuring batches of precipitation samples, the ME would induce an offset of ± 1.2 ‰ from the true value. Therefore, except in cases when innovative modifications of the internal measurement settings could be developed to mitigate the ME,⁸ it remains highly

recommended to systematically apply a numerical correction of the ME during post-measurement data processing.

In usual practices of laser data processing, the magnitude of the ME is generally evaluated during each run using reference waters included in the measurement batches, and corrected in consideration of the difference in isotopic composition between 2 successive samples.^{4-7,9,11} In most cases, satisfying precisions can be obtained. Nevertheless, the succession of injections is similar when the same sample is injected, and when moving to the following sample, except for the syringe rinses performed between vials. Therefore, assuming that the measurement of a sample is only affected by the last preceding one would imply that the persistence of the ME is equivalent to the time (or injection number) between two samples. The number of successive injections of the same sample hardly overtake 6-8 during routine analysis (or even less) to maintain reasonable data throughput.^{4-6,9,11,12} An overlooked longer persistence of memory would challenge the current practices, since more than one preceding sample could be involved in the ME.

It is generally assumed that 10-15 injections of the same sample are sufficient to obtain a composition free of memory.^{4-6,9,10} While the exact duration of the ME has never been fully assessed, some previous studies already suggested that substantial ME could remain after 15 injections.⁷ In addition, no evaluation has been performed so far of the impact of the ME on ¹⁷O-excess data, despite the high level of precision required to obtain reliable values. As the ME is expected to be the main factor reducing the precision and accuracy compared to the basic specifications of the instrument,⁸ an in-depth exploration of its impact and of the appropriate correction strategies is required.

The objectives of the present study were therefore to: 1) estimate the magnitude, duration, and variability of the ME and its impact on data precision and accuracy ; 2) propose a new correction method for the ME during post-run data processing and compare it with current practices, including the multi-reservoir correction method which has been developed for measuring labelled (*i.e.* strongly enriched) water¹³; 3) propose an optimal architecture for routine runs, and 4) evaluate the resulting improvement in measurement precision for the stable oxygen and hydrogen isotopes of the water molecule ($\delta^{17}\text{O}$, $\delta^{18}\text{O}$ and $\delta^2\text{H}$) and their derived tracers (¹⁷O-excess and d-excess).

2- Method

2-1 Basic protocol and routine measurement runs

We used a commercially available Picarro L2140i isotopic analyzer (Picarro, Santa Clara, CA, USA) to simultaneously measure $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and derive the values of ^{17}O -excess and d-excess. Water samples were stored in 2-ml glass vials and injected with a liquid autosampler (Picarro A0325) into a vaporizer module (Picarro A0211), using pure nitrogen as the carrier gas. A 10- μL syringe was used to collect the samples. The syringe was rinsed between each vial, and purged twice with the sample water, prior to injection into the vaporizer. The injection volume was 1.6 μL , the injection speed was 1 $\mu\text{L} \times \text{s}^{-1}$, and throughout the study, only the ^{17}O -mode of the analyzer was used. The initial instrument specifications for liquid water measurement provided by the manufacturer, estimated as the standard deviation (SD) of the mean of groups of 6 injections over 8 hours, were 0.025‰, 0.025‰, 0.1‰ and 0.015‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and ^{17}O -excess, respectively.

Two different run architectures were used for routine analysis (Table 1). The run architecture #1, used until June 2019, was designed to have an average duration of 48h, *i.e.* a total of about 320 injections, divided into 8 injections of 40 water samples. The run included 2 to 3 replicates of all samples in order to obtain the best precision possible for all isotopes and their derived secondary parameters ^{17}O -excess and d-excess. Normalization to the VSMOW-SLAP scale was performed by using a set of 3 in-house Working Standards (WS in the following) measured at the beginning, middle, and end of each run in order to monitor instrument drift.⁵ A linear calibration was applied, and the linearity of the calibration was evaluated by comparing the scaling factors corresponding for the different intervals between WS. These in-house WS were named ICE (low δ value), SEA (high δ value) and TAP (intermediate composition). A fourth in-house standard (ROB) was dedicated to Quality Assurance (QA), and treated as an unknown sample in all runs (the same number of replicates as for the samples). To normalize the $\delta^{17}\text{O}$ value in the VSMOW-SLAP scale, we used the SLAP2 composition proposed by Schoenemann et al. (2013).¹⁴

From September 2019, a different run architecture was adopted (run architecture #2). The modification was based on the evidence of a ME higher than expected, and after having tested different numbers of injections (12 to 15 injections, applied either to all vials or to WS only). In the new scheme, two sets of WS were measured, at the beginning and at the end of a run. A

“symmetrical ordering” was adopted for the two sets of WS (ascending *versus* descending delta-values), and for sample replicates. The number of injections was increased to 15 for calibration WS, while keeping 8 injections per sample. This organization was expected to mitigate most of the ME when averaging replicates, especially for intermediate composition ranges. However, to tackle the problem of very low and very high isotopic compositions, each group of calibration WS was preceded by a conditioning vial filled with the same water (Table 1). The data throughput was almost the same for both versions of the routine run architecture, with a total of 30 unknown water samples measured during approximately 48 hours, *i.e.* 15 replicated samples including the QA samples.

The post measurement data processing was carried out in the Python programming language¹⁵ (script available in supplementary information). All sample compositions are provided on the VSMOW-SLAP scale, after ME correction. In addition, the following information was systematically gathered in a “run summary” file to verify the overall quality of each measurement run: raw composition of the WS, mean water concentration and associated standard deviation for all injections of the run, calibrated values of the QA standard. Plots are also drawn to visualize all raw values and water concentration during the run, and a focus on all injections of the WS.

2-2 Measurement run dedicated to assessing the memory effect

The magnitude of the ME can be estimated from a memory factor representing the proportion of residual water vapor in the instrument during a measurement:

$$x_i = \frac{\delta_{i,meas} - \delta^n}{\delta^r - \delta^n} \quad \text{Eq. (1)}$$

where, during the injection i of a sample n (composition δ^n), x_i is the proportion of residual vapor (composition δ^r), and $\delta_{i,meas}$ is the measured composition. Assuming that the composition of the residual vapor corresponds to that of the previous sample (*i.e.* $\delta^r = \delta^{n-1}$), x_i is equivalent to previous formulations proposed to quantify the deviation of a measurement from its expected value (*e.g.* the term “ m ” in Guidotti et al.(2013),¹³ Van Geldern et al. (2011),⁹ or “ Φ ” in Olsen et al. (2006)¹⁶). Here it should be mentioned that the term “residual vapor” will be used in the following since all measurements are in the vapor phase, but strictly

speaking, the ME could also result from water adsorbed in its liquid form on the different parts of the measurement line, and thereafter desorbed.

The assumption that x_i represents the entire magnitude of the ME is only verified if the compositions of the two successive samples are measured completely free of ME. A run dedicated to evaluate the ME factors was designed with a total of 50 injections per sample vials. This number was chosen to keep a reasonable duration of the run (3 days), and represents a compromise between obtaining memory free values, and the effect of a potential drift. To encompass all the differences in δ values between the WS, the sample vials were sorted in both ascending and descending isotopic composition. In addition, three successions of vials filled with identical water were included to evaluate the robustness of the number of injections. Two similar dedicated runs were performed in October and November 2019. In addition, 3 other runs including 50 injections of successive samples were included in the ME analysis.

3- Results

3-1- Calibration of working standards and QA water

The compositions of the three in-house WS used for calibration were determined in two runs that included the VSMOW2-VSLAP2 international standards for calibration and normalization,^{14,17} and the GRESP certified reference water¹⁸ for quality control (Table 2). Sample vials were sorted, first, in a decreasing order of δ values, and then replicated in the reverse order, such that the symmetric succession of water compositions compensated for the possible ME. In addition, conditioning vials were added for VSMOW and SLAP. For each vial 50 successive injections were performed, and the corresponding selected raw value was estimated as the average of the last 15 injections. The total run lasted 4 days (715 injections) with a vaporizer septum change in the middle. A second run with the same architecture was performed one week apart. The normalized isotopic compositions of the 3 WS and of the GRESP standard were calculated from the average calibrated values of each of the 4 half-sequences. The composition of the 3 WS used for calibration covers approximately the VSMOW-GRESP range of isotopic composition. The measured $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values for the GRESP water fell in the range of the total combined uncertainties of IAEA composition.¹⁸ In addition, we determined the first values of $\delta^{17}\text{O}$ and ^{17}O -excess for the GRESP certified reference material:¹⁸ $\delta^{17}\text{O} = -17.7784 \pm 0.02 \text{ ‰}$; ^{17}O -excess = $0.025 \pm 0.005 \text{ ‰}$ (n=4).

Regarding the significantly greater precision obtained for ^{17}O -excess compared to $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$, and following the recommendations of Schoenemann et al.,¹⁴ the $\delta^{17}\text{O}$ value and $\delta^{18}\text{O}$ are reported here and in Table 2 with a number of decimals higher than allowed by their absolute precisions, in order to be consistent with the ^{17}O -excess calculation.

The assigned composition of the in-house standard used for QA (ROB) was determined from four dedicated runs, including the SEA and TAP in-house standards : two runs with 25 injections of all vials, taking the last 5 injections, and two runs with 50 injections, taking the last 15 injections (Table 2). The results of both runs were in good agreement, comprising differences less than 0.001, 0.002, 0.022 ‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, and $\delta^2\text{H}$, respectively.

3-2- Memory effect

3-2-1 Evaluation of memory effect factors

Regarding the expected asymptotic shape of the ME, obtaining “memory-free” values needs to be considered with regards to the measurement precision. The two runs dedicated to assess the ME (cf. section 2-2) indicated that the ME was no longer detectable after injection #45 of the same sample: the SD calculated for sets of similar water samples were of the order of the instrument precision (Table 3). Nevertheless, a closer look at the succession of data pointed that for $\delta^2\text{H}$ values SEA was lowest after having followed ICE, while ICE was highest after having just followed SEA, suggesting that a slight ME remained after 45 injections. As stated above, the total of 50 injections represents a compromise between memory removing and the effect of a potential drift, regarding the total duration of the dedicated runs (>3days), which is consistent with the measurement precision.

The last 5 injections of each sample were then used as the reference compositions for the current sample ($\delta^n = \delta_{i,meas}$ in Eq. 1) and for the residual vapor ($\delta^r = \delta^{n-1}$). In the run architecture, the succession of water sample compositions covered 3 different isotopic differences, in both directions, from which 6 series of ME factors (x_i) were calculated following Equation 1. In addition to these two runs, we also calculated ME factors from 3 other runs that involved 50 injections of successive samples with significant isotopic differences. Two of these runs (February 2020) included the VSMOW2 and SLAP2 standards, and therefore covered the

largest differences in δ -values. This provided a total of 24 series of ME factors. The averages and corresponding standard deviations are shown in Figure 1A and 1B. The highest variability of ME factors was found for $\delta^{17}\text{O}$, and the lowest for $\delta^2\text{H}$ (Figure 1G). This is consistent with the relative differences in the signal/noise ratio for each of the isotopic compositions, and roughly corresponds to their respective precisions compared to their natural variation ranges. For each isotopic species, the variability was especially high for the first 2 injections. Potential drivers of variations in ME factors were tested but did not evidence any clear trend. No relation was found for ME factors with the magnitude or the direction of isotopic differences between samples. In addition, the time variation of the ME was evaluated by plotting the average x_i values of each run against time for the injections #3 to #8 (Figure 2). This plot shows the decreasing ME with more injections, as already illustrated in Figure 1, but no variation with time appeared with regards to the SD of each ME factor average value. No change in the measurement procedure that might explain a variation in the ME (*e.g.* a change in the injection method, in the injection volume, the analyze of saline water, or a vaporizer cleaning) was performed. In the absence of a clear trend, the average series of ME factors were considered to be stable.

A semi-logarithmic scale allowed to evidence the exponential behavior of the ME decay with time, which can be fitted with a 2-term exponential function (Figure 1C and 1D). Similar average ME factors were found for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$, showing a similar behavior for the two oxygen isotopic species with regards to memory (Figure 1A and 1C). For $\delta^2\text{H}$ (Figure 1B and 1D), the ME was stronger and persisted longer than for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ (Figure 1A and 1C). The absolute impact of the ME with respect to the instrument precision is illustrated for the transition between the high- and the low-value WS (Figure 1E and 1F), showing the need for a correction until 36 injections for $\delta^{17}\text{O}/\delta^{18}\text{O}$, and 41 injections for $\delta^2\text{H}$. As the injections are fully automated, the number of injections can be translated into a time scale, providing that the same injection protocol is maintained: with our measurement conditions the ME needs approximately 5 hours to disappear. It should be specified that some exploratory experiments suggested that a lower injection volume, and thus a lower vapor concentration during pulse analysis, would increase the magnitude and the duration of the ME. Further data would be necessary to quantify this trend.

Reference series of ME factors were established using a smoothing of the average ME factors calculated from the 24 individual ME factor series, assuming a progressive decrease in ME from one injection to the next until its complete disappearance. Smoothing was performed with the Python programming language¹⁵ using Sequential Least Squares Programming (SLSQP in the SciPY module of PYTHON). The minimization of the difference between smoothed and raw ME factors was subjected to two constraints: 1) a decreasing trend, and 2) null values for the last 5 injections, used to estimate the reference composition. Smoothing was performed to avoid adding some noise to the data when applying an ME correction. The smoothing is based on the method proposed by Van Geldern and Barth⁹ which aimed at reducing the variance of all corrected injections for a vial, after the application of an increasing correction coefficient. Note that an alternative would be to use the exponential fits (Figure 1C and 1D), but the smoothing better fitted the measured ME factors. Two distinct reference series were established for $\delta^{18}\text{O}$ and $\delta^2\text{H}$, respectively. Based on the similarity between ME factors obtained for the two oxygen isotopic species (Figure 1A), the reference series built for $\delta^{18}\text{O}$ was also used to correct $\delta^{17}\text{O}$ measurements. Therefore, the application of the ME correction did not affect the precision of ^{17}O -excess.

3-2-2 Memory effect correction

A mathematical description of the ME is needed to apply a correction. Mixing equations are commonly used, but they may correspond to two different conceptual frameworks. In the first approach, the composition of the residual vapor is taken as that of the preceding sample. The contribution of the residual vapor decreases with successive injections, as residual vapor is flushed, but the end-member compositions are assumed to be constant. Most of the classical methods used to correct for the ME are based on this conceptual scheme, implying that $\delta^r = \delta^{n-1}$ in Equation 1.^{4-7,9,11,19} In the second approach, the residual vapor exchanges constantly with the current sample, and its composition progressively evolves towards that of the current sample, while the exchange rate is kept constant. This is the basis of the sophisticated correction method proposed by Guidotti et al.¹³ in the context of the very large variations in δ -values encountered when measuring doubly labelled waters (*e.g.* differences of several thousands of ‰ in $\delta^2\text{H}$). Based on the exponential decay of their ME factors with time, which can be fitted with a 3-term exponential function, they concluded that 3 water pools are involved in the ME. The correction equation describes exchange processes between these water pools, constrained

by several parameters (e.g. 6 parameters, when assuming 3 water pools), which are adjusted for each individual run, by minimizing the value for the SD calculated on all injections of the same sample. This approach better describes the physical processes, and implicitly accounts for the influences of several successive samples, as long as compositions of the water pools are not fully renewed. It has been developed in the context of measuring labelled water, and the parameter adjustment is based on selected samples whose δ -values strongly differ from that of the preceding samples.

Here, we propose a simpler correction procedure that is able to account for the long-lasting memory without the need to adjust any parameters. It assumes the variation with time (*i.e.* from one injection to the next, since the injections are automated) of 1) the mixing proportions between the current sample and the residual vapor, following the reference series of ME factors, and 2) the composition of the residual vapor δ^r . The evolution of the δ^r term was calculated assuming that, after the last injection of a given sample, its contribution to δ^r follows the ME factor reference series, whatever the further changes in water samples, until $x_i=0$. Each of the following samples contribute in turn to δ^r in decreasing proportions as the injections continue. A mixing equation was thus used to calculate the composition of the residual vapor during the injection i of the current sample n (δ_i^r) as a function of the compositions of four previous samples (the number of samples expected to contribute to the residual vapor depends on the number of injections performed for each sample and of the decay of the ME, and can be adjusted) :

$$\delta_i^r = \frac{x_i^{n-1} \delta^{n-1} + x_i^{n-2} \delta^{n-2} + x_i^{n-3} \delta^{n-3} + x_i^{n-4} \delta^{n-4}}{x_i^{n-1} + x_i^{n-2} + x_i^{n-3} + x_i^{n-4}} \quad \text{Eq. (2)}$$

where δ^{n-1} to δ^{n-4} are the compositions of the four samples that precede the sample n and x_i^{n-1} to x_i^{n-4} are the ME factor series counted from just after the last injection of the sample $n-1$ to $n-4$, as described in Figure 3. Practically, this composition is calculated in the data processing script by using 4 series of ME factors associated to 4 successive samples. The initial δ^r value is set to the composition of the first sample vial, therefore the number of injections of this conditioner sample need to be properly chosen.

Then, the measurement of the current sample n is corrected for the ME using the classical mixing equation:

$$\delta_{i,corr}^n = \frac{\delta_{i,meas} - x_i^{n-1} \delta_i^r}{1 - x_i^{n-1}} \quad \text{Eq. (3)}$$

Nevertheless, the application of this equation differs from its classical use, since both the terms δ_i^r (calculated from Eq. 2) and x_i vary from the first to the last injection of a sample. Similarly to the method proposed by Guidotti et al.,¹³ this simplified correction scheme is able to account for several preceding samples. Nevertheless, it doesn't need to make any assumptions on the number of water reservoirs involved in the ME, and it doesn't require to use parameters experimentally determined from the data being considered for ME correction.

3-2-3 Evaluation of the ME correction method

To evaluate the magnitude of the ME and the efficiency of the new correction procedure, we used three runs, which were initially designed to evaluate the short-term performances of the instrument (Table 4). They included blocks of 6 vials filled with the same water, treated as unknown samples, each block separated from the others by a group of 3 WS dedicated to calibration. The sequence was run three times, first with ICE as unknown samples, secondly with TAP and the last run with SEA. Ten injections were performed per vial. The vials placed immediately after the calibration WS group (*i.e.* positions #5, #14 and #23) were the most strongly affected by the compositions of previous samples, while the last vials in each block of similar samples (*i.e.* positions #10, #19 and #28) were not affected by ME.

For all these runs, four different methods were applied to correct the ME, and the calibrated results were compared (Figure 4). In the first method (M1), no ME correction was applied but the first 6 injections of all vials were discarded. In the second method (M2), the data were corrected by considering only the composition of the preceding sample (*i.e.* $\delta_i^r = \delta^{n-1}$). The third data processing (M3) followed the ME correction method proposed by Guidotti et al.¹³ Two water pools were considered, as suggested by the 2-component exponential fit of our ME factors (Figure 1C and 1D). The corresponding 4 parameters were adjusted for each run using the measured data corresponding to the WS used for calibration, which covered the highest difference in δ -values between successive samples, and therefore the highest level of information on the ME. The fourth data processing (M4) was performed with the new correction method described in Figure 3, therefore by calculating δ_i^r with Eq. 2, and correcting injections data using Eq.3. In this case, the first group of WS was discarded from the calibration, and kept

to initialize the ME correction. For the last three methods, the first 2 injections after the application of the ME correction were discarded.

For the first samples in each sub-block (Figure 4A to 4D), a slight offset was found when no numerical correction was applied (M1): on average 0.03 ‰, 0.06‰, 0.79‰ and 0.28‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess, respectively. The simple ME correction (M2) over-corrected the values, leading to a reverse offset of -0.03 ‰, -0.06 ‰, -1.06‰ and -0.57‰ on average for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess, respectively. The third method (M3) led to results almost similar to what was obtained by simply discarding the first injections (M1). The new correction (M4) led to intermediate values, still slightly over-corrected in $\delta^2\text{H}$ but very close to the reference compositions with offsets of 0.00 ‰, -0.01 ‰, -0.40‰ and -0.31‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess. For each method, the highest offsets were found for the lowest δ -value (the ICE WS). The ^{17}O -excess composition was not affected by the ME correction because similar ME coefficients were used for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ (not shown). However, slight differences were found when discarding the first 6 injections, since a lower number of injections was used for ^{17}O -excess determination.

The last vials of each block of unknown samples (#10, #19 and #28) were expected to be deprived of the ME, but their calibrated compositions depend on the correction applied to the WS data used for the calibration. When simply discarding the first 6 injections (M1), and with the 2-pool assumption (M3), significant offsets were observed, whose magnitude and direction depend on the isotopic composition of the unknown samples (Figure 4E to 4H). The offset was particularly marked for values close to the VSMOW composition (the SEA WS).

The poor improvement obtained with the Guidotti et al. approach compared to the method M1 came from the criterion used to adjust the correction parameters. The method was developed for measuring spiked water samples which carry a huge ME. In these cases, the ME correction greatly improve the SD calculated for successive injections of a sample. However, our results showed that the raw and ME-corrected groups of injections of a sample did not clearly differ in their SD, although they were different in their absolute values, with regards to the expected accuracy, which is almost one order of magnitude higher than what is expected for labelled water samples.¹³

3-3 Long-term statistics on routine runs

The new ME correction was applied retroactively to all sequences from October 2018 to March 2020. The data processing followed several successive steps. First, all the injections of a run were ME-corrected. The first two injections of each vial were systematically discarded because of the high variability in the corresponding ME factors (Figure 1G). Then, ME-corrected raw values of $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ were estimated for individual WS vials by averaging the remaining injections. The calibration equations were established using the average of all replicates of the 3 WS and applied to all the injections of a run. Finally, the calibrated data were averaged for each vial to obtain $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, and thereafter ^{17}O -excess and d-excess.

We observed slight differences between the compositions of the two QA replicates. The composition of the QA sample placed after the first group of WS was systematically lower than that of the second replicate in the run architecture #2. The difference between the 2 replicates (0.03 ‰ for $\delta^{18}\text{O}$ and 0.29‰ for $\delta^2\text{H}$) remained in the range of the final long-term reproducibility, but this imperfect ME correction showed that careful attention should also be paid to the run architecture to minimize the magnitude of the ME.

Despite very good linear regression coefficients (>0.999995), the linearity of the calibration was not fully reached, since the scaling factors based on the ICE-TAP WS (0.969, 0.978 and 0.971 for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$ and $\delta^2\text{H}$), were found slightly different from those based on the TAP-SEA intervals (0.975, 0.984 and 0.978 for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$ and $\delta^2\text{H}$). This slight non-linearity remained lower than what could result from the bias due the ME. In addition, the 3-point calibration was compared to a normalization performed using only the two most contrasting WS (ICE and SEA). The differences in calibrated data remained lower than 0.01‰ for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ and lower than 0.2‰ for $\delta^2\text{H}$.

Long-term statistics were calculated for the two run architectures used for routine analysis, using the average composition of QA replicates per run (Table 5). In addition, individual compositions of all QA are plotted for the entire measurement period (Figure 5). The resulting precision (estimated as the SD of QA compositions) was slightly improved with the run architecture #2 (Table 5: 0.015‰, 0.023‰, 0.078‰ and 0.006‰ and 0.173 ‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess, and d-excess respectively), and fell within the range of the instrument

specifications, corresponding to the continuous measurement of the same water. In addition, the accuracy estimated as the deviation (bias) from the expected composition (Table 5) was also improved for the run architecture #2: 0.005‰, 0.015‰, -0.124‰, -0.002‰ and -0.252‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess, and d-excess respectively. These offsets were lower than the SD for each tracer, except for d-excess, as a result of opposite, though low, offsets for $\delta^{18}\text{O}$ and $\delta^2\text{H}$.

4- Discussion and conclusion

Our results demonstrate the significant level and persistence of the ME when analyzing water samples with WS-CRDS. For example, the contribution of residual vapor is still between 0.5% and 1% in the 10th injection of the same water for oxygen and hydrogen. It is thus essential to apply a robust correction during post measurement data processing to obtain high precision isotopic data. A similar impact of the ME was evidenced for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$, allowing to use the same correction for the two isotopic species. Consequently, the application of a ME correction is transparent for the ^{17}O -excess determination, and affects neither the value nor the reproducibility of ^{17}O -excess data.

We showed that the ME required up to 45 successive injections of the same sample to be removed, with respect to the instrument precision, and to the δ -value intervals corresponding to our WS. This corresponds to about 5 hours of measurement with the “high precision mode” coordinator and an injection volume of 1.6 μL . This duration is longer than usually expected^{4-6,9} and has two main consequences. First, it makes the systematic evaluation of the ME in routine runs questionable, as it would be too time-consuming for a full assessment. With regards to the relative stability of the ME with time, it is preferable to perform a robust determination of ME factors based on several dedicated runs to obtain a robust average series. Second, it implies that several successive samples contribute to the composition of residual vapor affecting the measurement of a given sample in routine runs. Such a long-term persistence of the memory was already pointed out and accounted for in the 3-pool-exchange model proposed by Guidotti et al. (2013) for memory correction of doubly labelled waters characterized by huge differences in δ -values.¹³ We proposed here an alternative ME correction method based on a mixing equation to calculate the composition of the residual water vapor as a function of the compositions of a number of previous samples. Compared to the method proposed by Guidotti

et al., this simplified correction scheme was found more suitable for natural samples, for which the magnitude of the ME is lower, and the expected precision higher than for labelled water.

The insufficient number of injections used to assess the ME in most current practices^{4-7,9} leads to the underestimation of the full magnitude of the memory. Nevertheless, the influence of an inappropriate ME correction may remain hidden in a number of situations, *e.g.* in cases where the under-estimated ME is compensated for by the alternation of variable sample compositions mitigated long-lasting ME, *e.g.* when δ^{n-2} is closer to δ^n than δ^{n-1} . However, in a monotonic succession of sample compositions (either increasing or decreasing δ values) the under-estimated ME results in inaccurate data. The new correction formula led to better precision whatever the distribution of sample compositions in the run architecture. A systematic offset in calibrated data may also be hidden when a constant order is adopted in the routine runs for WS, since the calibrated data may be indirectly affected by a bias in the measurement of the WS used for calibration. This effect, shown in Figure 4E to 4H, may remain negligible for sample compositions falling in between WS values. However, significant bias were found for samples close to WS compositions.

The determination of robust ME factors depends on the assumption of a memory-free value for the composition of the current vial. However, the asymptotic decay of the ME makes difficult to assert a complete disappearance of memory. We found that 50 injections were a good compromise between memory dampening and the possible instrumental drift, with regards to the measurement precision. Nevertheless, this major pitfall combined with the natural variability of ME factors (Figure 1) implies that a perfect correction can never be accomplished, and that it is necessary to minimize the magnitude of the ME by adopting an appropriate run architecture. The magnitudes of the ME (estimated as non-corrected *minus* corrected calibrated data) were compared for two runs of precipitation samples, conducted with the two architectures described previously. Precipitation samples usually encompass a large range of isotopic compositions; in these cases, values varied from $\delta^{18}\text{O} = -8\text{‰}$ to 0‰ . In the run architecture #1, the magnitude of the ME reached 0.07‰ , 0.15‰ , 1.7‰ and 0.7‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess respectively. In the run architecture #2, the differences were reduced by half, remaining below 0.04‰ , 0.06‰ , 0.9‰ and 0.7‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$ and d-excess respectively. The dampening of the ME was obtained by performing 15 injections of each WS used for calibration, including conditioning samples before each WS group, and

systematically analyzing replicates of WS and samples, the replicates being sorted in reverse order in a run sequence. The symmetrical organization also eliminated the need for a drift correction, as this effect is implicitly corrected by averaging replicates. In recent instruments the drift generally remains low (e.g. Schauer et al.⁸ did not evidence any drift during periods of 10 consecutive days), but when a drift correction is required, a robust memory correction is essential to avoid the misattribution of the ME to a drift effect.

The use of a QA WS included in all runs is essential. To avoid the hidden effect of a constant shift, the QA should be placed randomly among unknown samples. In addition, a Quality Control (QC) water would be required to fully assess the data accuracy (trueness). Then, the composition of the QA/QC water has to be determined independently, using primary international standards, with a run architecture different from that of routine measurements, and with a number of injections sufficient to eliminate the ME. Laboratories using IRMS techniques in parallel with laser technology have the opportunity to perform independent measurements. However, the precision obtained for $\delta^2\text{H}$ using IRMS techniques hardly reaches 0.1‰,^{16,20} which would be necessary with respect to what is expected in laser data. Ultimately, dedicated inter-laboratory comparisons are necessary to fully evaluate the data accuracy.

The difference between the 3-point calibration used in this study and a normalization performed using the two most contrasting WS (ICE and SEA), evaluated on calibrated data, remained lower than 0.01‰ for $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ and lower than 0.2‰ for $\delta^2\text{H}$. This would plead for the 2-point calibration classically recommended (REF), with a large range of compositions for the WS, since it has the advantage of minimizing error propagation during calibration. However, the unexpectedly long duration of the ME underlines the need for an appropriate correction to compensate for the persistence of a highly different residual vapor during the measurement of the samples that follow the WS. An alternative would be to choose WS compositions closer to the expected sample compositions (e.g. no strongly depleted compositions, when working on temperate climate regions), and use up to 3 WS for the calibration. But the choice of 3 WS for the calibration is time-consuming. Therefore, the use of 2 WS covering a large range of composition would remain the best choice, provided that the ME is properly corrected.

Finally, the new ME correction method proposed in this study (Python script in the supplementary information) combined with an appropriate run architecture, allowed us to achieve a long-term precision similar to the instrumental specifications provided by the

manufacturer. The standard deviation estimated for the QA after modification of the run architecture (7-month period) showed a precision of 0.015‰, 0.023‰, 0.078‰, 0.006‰ and 0.173‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess., and d-excess., respectively. The offset from the expected composition of the QA was 0.005‰, 0.015‰, -0.124‰, -0.002‰ and -0.252‰ for $\delta^{17}\text{O}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$, ^{17}O -excess, and d-excess respectively (Table 5). In addition, the precision we obtained for ^{17}O -excess, allowed us to demonstrate the first determination of $\delta^{17}\text{O}$ and ^{17}O -excess compositions for the GRESP certified reference material¹⁸ (Table 2).

Acknowledgements

This study benefited from the ANR HUMI-17 (ANR-17-CE01-0002), ECCOREV and LABEX OT-Med grants. Many thanks to Clément Piel for fruitful discussions, especially during the instrument set-up.

References

1. Wassenaar LI, Ahmad M, Aggarwal P, et al. Worldwide proficiency test for routine analysis of d2H and d18O in water by isotope-ratio mass spectrometry and laser absorption spectroscopy. *Rapid Commun Mass Spectrom.* 2012;26(15):1641-1648. doi:10.1002/rcm.6270
2. Wassenaar LI, Terzer-Wassmuth S, Douence C, Araguas-Araguas L, Aggarwal PK, Coplen TB. Seeking excellence: An evaluation of 235 international laboratories conducting water isotope analyses by isotope-ratio and laser-absorption spectrometry. *Rapid Communications in Mass Spectrometry.* 2018;32(5):393-406. doi:10.1002/rcm.8052
3. Aron PG, Levin NE, Beverly EJ, et al. Triple oxygen isotopes in the water cycle. *Chemical Geology.* Published online December 31, 2020:120026. doi:10.1016/j.chemgeo.2020.120026
4. Pierchala A, Rozanski K, Dulinski M, Gorczyca Z, Marzec M, Czub R. High-precision measurements of delta H-2, delta O-18 and delta O-17 in water with the aid of cavity ring-down laser spectroscopy. *Isotopes in Environmental and Health Studies.* 2019;55(3):290-307. doi:10.1080/10256016.2019.1609959
5. Wassenaar LI, Coplen TB, Aggarwal PK. Approaches for Achieving Long-Term Accuracy and Precision of delta O-18 and delta H-2 for Waters Analyzed using Laser Absorption Spectrometers. *Environmental Science & Technology.* 2014;48(2):1123-1131. doi:10.1021/es403354n
6. Coplen TB, Wassenaar LI. LIMS for Lasers 2015 for achieving long-term accuracy and precision of delta H-2, delta O-17, and delta O-18 of waters using laser absorption spectrometry. *Rapid Communications in Mass Spectrometry.* 2015;29(22):2122-2130. doi:10.1002/rcm.7372
7. Groening M. Improved water delta H-2 and delta O-18 calibration and calculation of measurement uncertainty using a simple software tool. *Rapid Communications in Mass Spectrometry.* 2011;25(19):2711-2720. doi:10.1002/rcm.5074
8. Schauer AJ, Schoenemann SW, Steig EJ. Routine high-precision analysis of triple

- water-isotope ratios using cavity ring-down spectroscopy. *Rapid Communications in Mass Spectrometry*. 2016;30(18):2059-2069. doi:10.1002/rcm.7682
9. van Geldern R, Barth JAC. Optimization of instrument setup and post-run corrections for oxygen and hydrogen stable isotope measurements of water by isotope ratio infrared spectroscopy (IRIS). *Limnology and Oceanography-Methods*. 2012;10:1024-1036. doi:10.4319/lom.2012.10.1024
10. Penna D, Stenni B, Sanda M, et al. Technical Note: Evaluation of between-sample memory effects in the analysis of delta H-2 and delta O-18 of water samples measured by laser spectrometers. *Hydrology and Earth System Sciences*. 2012;16(10):3925-3933. doi:10.5194/hess-16-3925-2012
11. Holko L. Syringe life and memory effects in isotopic analyses performed by liquid water isotopic analysers - a case study for natural waters from central Europe. *Isotopes in Environmental and Health Studies*. 2016;52(4-5):553-559. doi:10.1080/10256016.2015.1090987
12. Qu D, Tian L, Zhao H, Yao P, Xu B, Cui J. Demonstration of a memory calibration method in water isotope measurement by laser spectroscopy. *Rapid Communications in Mass Spectrometry*. 2020;34(8):UNSP e8689. doi:10.1002/rcm.8689
13. Guidotti S, Jansen HG, Aerts-Bijma AT, Verstappen-Dumoulin BM a. A, van Dijk G, Meijer H a. J. Doubly Labelled Water analysis: Preparation, memory correction, calibration and quality assurance for delta H-2 and delta O-18 measurements over four orders of magnitudes. *Rapid Communications in Mass Spectrometry*. 2013;27(9):1055-1066. doi:10.1002/rcm.6540
14. Schoenemann SW, Schauer AJ, Steig EJ. Measurement of SLAP2 and GISP 17O and proposed VSMOW-SLAP normalization for 17O and 17Oexcess. *Rapid Commun Mass Spectrom*. 2013;27(5):582-590. doi:10.1002/rcm.6486
15. Python Software Foundation. Python Language Reference, version 3.7. Available at <http://www.python.org>.
16. Olsen J, Seierstad I, Vinther B, Johnsen S, Heinemeier J. Memory effect in deuterium analysis by continuous flow isotope ratio measurement. *International Journal of Mass Spectrometry*. 2006;254(1-2):44-52. doi:10.1016/j.ijms.2006.05.004
17. IAEA. Reference Sheet for VSMOW2 and SLAP2 international measurement standards. Published online 2009.
18. IAEA. Reference sheet for certified reference material - GRESP, Greenland Summit Precipitation water. Published online March 2021.
19. Gupta P, Noone D, Galewsky J, Sweeney C, Vaughn BH. Demonstration of high-precision continuous measurements of water vapor isotopologues in laboratory and remote field deployments using wavelength-scanned cavity ring-down spectroscopy (WS-CRDS) technology. *Rapid Communications in Mass Spectrometry*. 2009;23(16):2534-2542. doi:10.1002/rcm.4100
20. Donnelly T, Waldron S, Tait A, Dougans J, Bearhop S. Hydrogen isotope analysis of natural abundance and deuterium-enriched waters by reduction over chromium on-line to a dynamic dual inlet isotope-ratio mass spectrometer. *Rapid Communications in Mass Spectrometry*. 2001;15(15):1297-1303. doi:10.1002/rcm.361

Table 1: Architectures of the runs used for routine measurements. In the run architecture #1 (used over October 2018 – May 2019) unknown and QA samples are replicated (2 or 3 replicates) in the same order. In the run architecture #2 (used over September 2019 – March 2020) unknown samples are replicated in reverse order.

run architecture #1		purpose	
vial position	identifier		number of injections
1	DI water	conditioner	8
2	ICE	Low δ -value WS for calibration	8
3	TAP	Intermediate δ -value WS for calibration	8
4	SEA	high δ -value WS for calibration	8
5-19	samples	unknown samples, including ROB as QA WS	8
20	ICE	Low δ -value WS for calibration	8
21	TAP	Intermediate δ -value WS for calibration	8
22	SEA	high δ -value WS for calibration	8
23-37	samples	Replicated unknown samples, including ROB as QA WS	8
38	ICE	Low δ -value WS for calibration	8
39	TAP	Intermediate δ -value WS for calibration	8
40	SEA	high δ -value WS for calibration	8
run architecture #2			
vial position	identifier		number of injections
1	ICE	conditioner	8
2	ICE	Low δ -value WS for calibration	15
3	TAP	Intermediate δ -value WS for calibration	15
4	SEA	high δ -value WS for calibration	15
5	ROB	QA WS	8
6-20	samples	unknown samples	8
21	ROB	QA WS	8
22-36	samples	replicated unknown samples, in reverse order	8
37	SEA	conditioner	8
38	SEA	high δ -value WS for calibration	15
39	TAP	Intermediate δ -value WS for calibration	15
40	ICE	Low δ -value WS for calibration	15

Table 2: Mean compositions of in-house WS normalized to the VSMOW-SLAP scale (all values in ‰). ICE, TAP and SEA were measured in the same runs as VSMOW2, SLAP2 and GRESP, including 50 injections per vial, GRESP being treated as an unknown. ROB was measured from 4 runs including SEA and TAP for calibration. The total number of injections used to calculate the δ -values are 60 for ICE, SEA, TAP and GRESP (4 vials, 15 injections per vial), and 40 for ROB (2 vials with the last 15 injections of 50, and 2 vials with the last 4 injections of 15). The sample size is 4 for all mean values and corresponding SDs. See text for a more detailed description of the runs. Regarding the significantly greater precision obtained for ^{17}O -excess compared to $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$, and following the recommendations of Schoenemann et al.,¹⁴ the $\delta^{17}\text{O}$ value and $\delta^{18}\text{O}$ are reported here with a number of decimals higher than allowed by their absolute precisions, in order to be consistent with the ^{17}O -excess calculation ($^{17}\text{O}\text{-excess} = 1000 \times (\delta'^{17}\text{O} - 0.528 \times \delta'^{18}\text{O})$, with $\delta' = \ln(\delta/1000 + 1)$, when all values are expressed in ‰).

purpose	identifier		$\delta^{17}\text{O}$	$\delta^{18}\text{O}$	$\delta^2\text{H}$	$^{17}\text{O}\text{-excess}$	d-excess
calibration WS	ICE	mean	-14.1836	-26.7411	-203.05	0.026	10.88
		SD	0.026	0.041	0.53	0.007	0.28
calibration WS	TAP	mean	-4.5285	-8.5926	-59.06	0.018	9.68
		SD	0.017	0.021	0.49	0.006	0.41
calibration WS	SEA	mean	-0.1395	-0.2457	0.32	-0.010	2.28
		SD	0.006	0.018	0.09	0.004	0.19
IAEA reference material	GRES	<i>assigned composition</i>		-33.40	-258.0		9.32
		<i>cum. Uncertainties</i>		0.04	0.4		0.72
		mean	-17.7784	-33.4487	-257.87	0.025	9.72
		SD	0.020	0.034	0.52	0.005	0.30
QA WS	ROB	mean	-4.0588	-7.7084	-52.75	0.018	8.92
		SD	0.025	0.035	0.04	0.010	0.29

Table 3: Architecture of the run dedicated to estimating the ME factors and corresponding raw (uncalibrated) delta values obtained from the average of the last 5 out of the 50 injections of each vial (all values in ‰).

vial position	identifier	$\delta^{17}\text{O}$	$\delta^{18}\text{O}$	$\delta^2\text{H}$	^{17}O -excess
1	ICE	-14.98	-28.18	-205.61	0.004
2	ICE	-15.02	-28.25	-205.92	-0.001
3	SEA	-0.55	-1.20	3.02	0.085
4	SEA	-0.55	-1.18	3.41	0.069
5	TAP	-5.01	-9.64	-57.44	0.088
6	TAP	-5.03	-9.65	-57.39	0.079
7	SEA	-0.56	-1.18	3.50	0.066
8	ICE	-14.99	-28.18	-205.44	-0.007
9	TAP	-5.04	-9.66	-57.69	0.074
10	ICE	-15.02	-28.24	-205.68	-0.008
Corresponding SD					
	ICE	0.02	0.03	0.20	0.006
	TAP	0.01	0.01	0.16	0.007
	SEA	0.00	0.01	0.26	0.010

Table 4: Architecture of the run used to compare ME corrections (10 injections per vial). The sequence was run 3 times, each of them with one of the WS used to fill the 18 sample vials and treated as unknown samples.

vial position	identifier 1	identifier 2
1	water	conditioner
2	ICE	WS
3	TAP	WS
4	SEA	WS
5		Sample
6		Sample
7		Sample
8		Sample
9		Sample
10		Sample
11	ICE	WS
12	TAP	WS
13	SEA	WS
14		Sample
15		Sample
16		Sample
17		Sample
18		Sample
19		Sample
20	ICE	WS
21	TAP	WS
22	SEA	WS
23		Sample
24		Sample
25		Sample
26		Sample
27		Sample
28		Sample
29	ICE	WS
30	TAP	WS
31	SEA	WS

Table 5: Long term statistics on the QA WS (ROB) calculated separately for the two run architectures described in Table 1. Bias is calculated as measured *minus* accepted values (all values in ‰).

	run architecture #1	run architecture #2
n	46	20
$\delta^{17}\text{O}$ bias	0.008	0.005
$\delta^{17}\text{O}$ SD	0.018	0.015
$\delta^{18}\text{O}$ bias	0.021	0.015
$\delta^{18}\text{O}$ SD	0.025	0.023
$\delta^2\text{H}$ bias	-0.214	-0.124
$\delta^2\text{H}$ SD	0.146	0.078
^{17}O -exc bias	-0.002	-0.002
^{17}O -exc SD	0.009	0.006
d-exc bias	-0.381	-0.252
d-exc SD	0.147	0.173

Figure captions

Figure 1: Average ME factors (solid lines) with their associated standard deviations (dotted lines: $\pm 1SD$, $n = 24$) calculated for $\delta^{17}O$ and $\delta^{18}O$ (1A), and δ^2H (1B). Same average ME factors plotted on a semi-logarithmic scale, with a 2-component exponential fit (1C, 1D). Example of raw data obtained for the low δ -value WS (ICE) following 50 injections of the high δ -value WS (SEA) (1E, 1F). The continuous lines represent the upper and lower limits of their uncalibrated composition, assumed deprived of ME (average of the last 5 injections), considering the instrument precision provided by the manufacturer ($\pm 0.025\text{‰}$ for $\delta^{17}O$ and $\delta^{18}O$, and $\pm 0.1\text{‰}$ for δ^2H). Comparison of the standard deviations of each series of average ME factors (1G).

Figure 2: ME factors corresponding to the injections #3 to #8. Average values and corresponding SD (error bars) are calculated for each of the 4 runs performed between September 2019 and February 2020, while the first run performed in 2018 included only one series.

Figure 3: Schematic description of the build-up of residual vapor (δ_i^r), following Eq. 2, during the injections of a sample of composition δ^n

Figure 4: Comparison of offsets remaining after different ME corrections (Methods M1 to M4; see text for details). Offsets were estimated from the 3 runs described in Table 4, as the deviation from expected values calculated for vials #5, #14 and #23 (4A to 4D) and for vials #10, #19 and #28 (4E to 4H)

Figure 5: Calibrated composition of all replicates of QA samples (ROB) included in routine runs between October 2018 and March 2020. Also shown are the reference compositions (black lines) and the average composition of the QA d samples corresponding to the two run architectures (red dashed lines)

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5