Journal of Archaeological Science 186 (2026) 106457

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: www.elsevier.com/locate/jas

An automated spreadsheet for determining analytical uncertainty of stable isotope measurements

Paul Szpak

Department of Anthropology, Trent University, Peterborough, Ontario, K9L 0G2, Canada

ARTICLE INFO

Keywords:
Stable isotopes
Analytical uncertainty
Quality control
Quality assurance

ABSTRACT

The quantification and reporting of the analytical uncertainty that accompanies all stable isotope measurements is critical, providing a mechanism for assessing their comparability against other datasets. The ability to compare isotopic datasets generated across many labs is especially important as there has been a proliferation in the number of stable isotope databases across a range of disciplines, facilitating large scale metanalyses of a rapidly expanding library of isotopic measurements. This paper presents a simple and automated spreadsheet-based approach for quantifying and reporting analytical uncertainty of stable isotope measurements.

1. Introduction

There are many ways to quantify analytical uncertainty, varying in terms of their complexity, additional costs (in terms of analysis time, labor, and materials), and importantly, scale (Dunn et al., 2025; Jardine and Cunjak, 2005; Meija and Chartrand, 2018; Skrzypek et al., 2010). For example, the ISO/EIC Guide to Expression of Measurement Uncertainty (GUM) presents a widely recognized framework for determining measurement uncertainty (Joint Committee for Guides in Metrology, 2008), but relies on partial differential equations and is not especially user friendly. One approach that presents a non-calculus method of determining uncertainty is the Kragten spreadsheet (Kragten, 1994), which has been used in a number of instances for stable isotope data. While a good approach, the Kragten method quantifies uncertainty at the level of individual samples, necessitating the analysis of a quantity of sample replicates that may be impossible (because of sample size limitations), not cost-effective (e.g., for very large projects), or not especially useful. This approach may be less useful when there is no reason to suspect that the variability in replicate sample measurements would vary, for example, with especially well-homogenized materials. Bone collagen, for example, is typically dissolved in solution, frozen, and lyophilized, producing a highly homogenous mixture. For sample types that are inherently more heterogeneous or are challenging to homogenize during sample preparation, an uncertainty approach that is sample-specific, such as the Kragten spreadsheet, may be more appropriate.

This paper presents a simple and automated spreadsheet-based approach for quantifying and reporting analytical uncertainty in stable isotope measurements. Unlike the sample-specific Kragten approach, this sheet is designed to determine the analytical uncertainty associated with a set of samples within an analytical session (run) or across multiple analytical sessions, as might be associated with a discrete project. This paper does not attempt to provide an exhaustive set of best practices or guidelines for stable isotope research. For those interested in an excellent resource on these topics, the freely available IRMS Good Practice Guide produced by members of the Forensic Isotope Ratio Mass Spectrometry (FIRMS) Network (Dunn et al., 2025) provides a very accessible and comprehensive assessment of these matters in addition to other papers that deal with related issues in a great level of detail (Carter and Fry, 2013; Coleman and Meier-Augenstein, 2014; Coplen, 2011; Dunn and Skrzypek, 2023; Paul et al., 2007). Nonetheless, a brief overview of some of the relevant terms used in the spreadsheet and throughout this text are necessary.

Analytical session. Sometimes colloquially called a ‘run’, this refers to the total samples and reference materials analyzed within a defined period of time, most typically without long periods of interruption (down time), altered instrument tuning, or changes to the instrument itself (e.g., replacing traps, columns, or other materials).

Reference material (RM). In some cases, the shorthand ‘standard’ may be used in place of reference material. Reference materials are used for a variety of purposes. Isotope ratio mass spectrometers initially compare the isotope ratio of a gas sample to a laboratory reference gas, or less commonly a solid reference material, producing a raw δ value relative to the laboratory working gas that requires further calibration or normalization to a recognized isotope delta scale such as VPDB or AIR. This calibration is accomplished via reference materials with known δ values

---

E-mail address: paulszpak@trentu.ca.

https://doi.org/10.1016/j.jas.2025.106457
Received 3 August 2025; Received in revised form 23 November 2025; Accepted 14 December 2025

that are interspersed between unknown samples. These materials (hereafter calibration RMs) often have assigned δ values defined by interlaboratory comparisons (ILCs) and are available for purchase through organizations such as the IAEA or USGS (Dunn et al., 2020). In some cases these δ values may change over time. Since calibration RMs are used to define the relationship between the measured and known values, they cannot be used to monitor systematic error. RMs that are interspersed amongst unknown samples but are not used in the process of calibration serve as independent checks on the analytical accuracy and precision of the analyses (hereafter quality control or 'QC RMs'). These materials may similarly have internationally agreed-upon δ values based on ILCs, but they are frequently in-house RMs that laboratories have selected because they have similar physical and chemical properties to the samples that a given lab most commonly analyzes (Dunn et al., 2021). These QC RMs should also have similar δ values to those that are expected for the samples being analyzed. There is a robust literature documenting important characteristics of reference materials in the context of stable isotope analysis (Dunn et al., 2020, 2021, 2025; Meier-Augenstein and Schimmelmann, 2018; Skrzypek, 2013). Readers are encouraged to engage with this literature and think critically about the appropriateness of various reference materials for their applications to provide them with the most complete picture of analytical uncertainty.

Matrix matching. Reference materials that are analyzed alongside unknown samples should have a similar chemical composition or 'matrix' to those samples. Matching the reference materials and the unknown samples in this way helps to ensure that both behave similarly during combustion, gas purification, gas separation, and ionization (Balint et al., 2024). For samples that are proteinaceous in nature (e.g., bone collagen, hair), reference materials that are also proteinaceous represent the best match, but at the very least, organic reference materials are superior to inorganic reference materials when analyzing organic samples (Gentile et al., 2013). Precise matrix matching will not always be feasible for calibration and QC RMs as a sufficiently broad range of materials with the appropriate δ values have not been characterized by ILCs. For example, USGS88 and USGS89 are the most well matched RMs available relative to bone collagen, but the relative difference between the δ^{13}C, δ^{15}N, and δ^{34}S values are all far too small (~2 ‰ for δ^{13}C, ~8 ‰ for δ^{15}N, and ~13 ‰ for δ^{34}S) to be effectively used for calibration (Balint et al., 2024), meaning that other organic, but not collagenous, RMs must be used.

Calibration. Also referred to as normalization, calibration involves analyzing reference materials with known δ values alongside samples. By determining the mathematical relationship between the measured (expressed relative to a laboratory working gas, for example) and known (expressed relative to an isotope delta scale such as VPDB or VSMOW) δ values of these RMs, it is possible to create an equation (frequently called a calibration curve, although this may often be a simple least squares regression) that can be applied to all RMs and samples analyzed within that session. This equation may be based on as few as two reference materials that bracket the range of δ values expected in unknown samples (Skrzypek, 2013) and span at least 20 ‰ (Balint et al., 2024). Additional RMs can be used to construct a multi-point calibration that may offer some benefits with respect to reducing error (Paul et al., 2007), especially when the RMs and standards are matrix-matched (Balint et al., 2024). One limitation of the simple least squares regression approach is that it fails to account for uncertainty in the assigned δ values of the reference materials, which often varies among RMs. For example, the reported uncertainties for δ^{13}C and δ^{15}N of two glutamic acid RMs commonly used to anchor calibration curves (USGS41a and USGS40) are quite different, with USGS41a possessing more than twice the reported uncertainty for δ^{13}C and δ^{15}N than USGS40. While not employed with any regularity to date, it is also possible to develop alternative models for calibration that incorporate the uncertainty attributed to the assigned δ values of the calibration RMs (Meija and Chartrand, 2018). These uncertainties assigned to the RMs used for calibration are not accounted for by the uncertainty sheet presented in this paper, which works only with the widely used least squares regression method of calibration.

Replicate. Any unknown sample that is analyzed more than once (e.g., duplicate, triplicate), thus allowing for an assessment of random error that is specific to the samples being analyzed.

Drift. A change in the measured δ values over time within an analytical session. Interspersing RMs throughout the session provides an opportunity to observe drift (Fig. 1A). If the relationship between time and the δ value of an RM is consistent, a correction can be applied (Dunn et al., 2025). Drift is not something that can be corrected for with the uncertainty sheet, nor is there a specific area to differentiate any error specifically associated with drift. If, however, reference materials not used for drift correction are dispersed evenly throughout an analytical session, any error caused by instrumental drift should be reflected in higher variability in the δ values of these materials. This higher variability in these RMs would in turn be reflected in the uncertainty calculation used in the attached spreadsheet.

Linearity. A linearity effect occurs when the measured δ values varies as a function of the size of the sample being analyzed, more specifically the amount of gas (expressed as peak area or peak height) reaching the mass spectrometer (Fig. 1B). Linearity effects are relevant when they are observed across the range of peak areas observed for the samples being analyzed and are most likely to be an issue where elemental compositions vary widely among samples. For instance, there is 100 % more sulfur in fish collagen than in mammalian collagen and analyzing these two materials at the same mass could produce linearity effects. As with drift, the uncertainty sheet cannot correct for linearity effects and best practices should be followed to first minimize the probability of appreciable linearity effects, and secondarily by including relevant reference materials to apply corrections to unknown samples (Dunn et al., 2025) as tests of reference gases may not always reflect linearity effects with actual combusted samples (Balint et al., 2024). Including reference materials that are matrix-matched to the unknown samples throughout the session at a range of peak areas approximating those of the unknown samples should effectively integrate any error caused by linearity effects into the uncertainty calculation used in this spreadsheet.

Systematic error. Systematic errors are consistent and repeatable errors that bias measurements in the same direction (i.e., too high or too low) (Fig. 2). They are frequently the product of issues in the process of calibration. They can only be monitored through the use of RMs that did not factor into the production of the calibration curve (i.e., QC RMs).

Random error. Random error is associated with noise, often caused by small variations in the instrument or laboratory's conditions or the inherent heterogeneity of the material being analyzed (Fig. 2). Random error can be monitored via RMs that were included in the calibration curve, those that were not included in the calibration curve (QC RMs), and sample replicates.

2. Overview of the uncertainty calculation

The underlying math of this spreadsheet and the resulting calculations are the same as those that were presented in the supplementary material of Szpak et al. (2017), but the spreadsheet presents a number of important improvements, specifically requiring minimal copying and pasting and increasing the speed with which one can use the sheet by orders of magnitude, especially for large projects or datasets. Anecdotally, uncertainty calculations previously performed by the author for large datasets that took more than a full working day (Szpak et al., 2019), could be achieved in a few minutes with this new sheet. What follows here is a brief recap of the various elements of uncertainty that are captured in the sheet and an overview of the various calculations. The standard uncertainty (u_c) is determined by taking the root-sum-square of two terms that are associated with random error ($u(R_w)$) and systematic error ($u(bias)$) as in Equation (1):

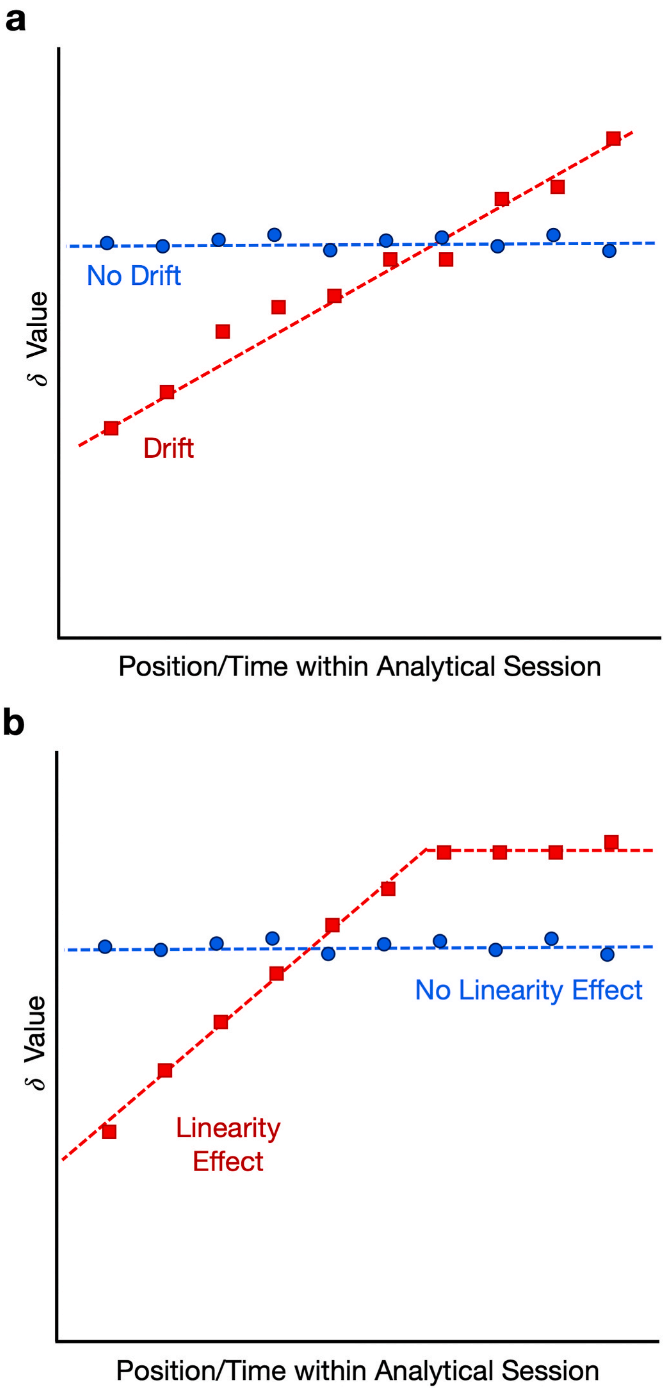

Fig. 1. (A) *Drift.* Drift is not observed and the δ values of a reference material stay stable throughout the session (circles). Drift is observed and the δ values of a reference material change in a consistent way throughout the session (squares). (B) *Linearity*. There is no relationship between the amount of gas reaching the mass spectrometer and the δ values of a reference material (circles). Below a threshold, there is a relationship (linearity effect) between the amount of gas reaching the mass spectrometer and the δ values of a reference material (squares).

$$u_c = \sqrt{u(R_w)^2 + (u(bias))^2} \qquad \text{Equation 1}$$

The term associated with random error ($u(R_w)$) can be captured by repeated measurements of materials of both known and unknown δ values as it reflects the 'noise' of the measurements. This random error is itself the root-sum-square of the pooled standard deviations of any repeated measurements across all analytical sessions, with knowns or RMs (s_{RM}) and unknowns or samples (s_{rep}) being differentiated as in Equation (2):

$$u(R_w) = \sqrt{(s_{RM})^2 + (s_{rep})^2} \qquad \text{Equation 2}$$

The s_{RM} term includes reference materials that were part of the calibration curve as well as those that were not, since all of these materials will reflect the random error of the relevant stable isotope measurements through the covered sessions. Presumably these reference materials are all reasonably well matrix-matched to the samples being analyzed. This condition may not always be met, for example when δ^{34}S measurements are made on bone collagen (an organic matrix with <1 % sulfur by weight) and the δ^{34}S scale is calibrated using silver sulfide or barium sulfate RMs (inorganic matrices with >10 % sulfur by weight). In a situation such as this, care should be taken not to underestimate uncertainty by biasing calculations towards materials that are not representative of the samples being measured (see FAQ13 for more discussion). The s_{rep} term includes any unknown materials that were measured repeatedly. Both s_{RM} and s_{rep} are calculated as in Equation (3):

$$s_x = \sqrt{\frac{(n_1-1){s_1}^2 + (n_2-1){s_2}^2 + \ldots + (n_k-1){s_k}^2}{n_1+n_2+\ldots+n_k-k}} \qquad \text{Equation 3}$$

In Equation (3), k represents the total number of RMs (for s_{RM}) or unknown samples (s_{rep}) that were measured across all sessions. For s_{RM}, n_1 and n_2 represent the number of a given reference material that was analyzed in a given analytical session, rather across all sessions. For example if 4 USGS40s were included in all three sessions, the numbers of USGS40s in each of the three sessions would be reflected as n_1, n_2, n_3 each being 4 rather than n_1 representing the 12 total USGS40s. The reason for treating the standards on a per session basis is expanded upon in FAQ1. For both s_{RM} and s_{rep}, s_k represents the standard deviation of a given reference material within a session or a given sample replicate across all sessions.

Systematic errors are reflected in the *u(bias)* term of Equation (1), which itself is defined by Equation (4):

$$u(bias) = \sqrt{(RMS_{bias})^2 + u(CRM)^2} \qquad \text{Equation 4}$$

The *u(CRM)* term represents the root-mean-square of the known standard deviation of the reference materials that were not included in the calibration curve (Equation (5)). This metric may be determined from long-term measurements of in-house reference materials or the supplied uncertainty from an external provider may be used. This term is important because it reflects the uncertainty associated with the known value of the reference material, which may be more or less homogenous than an absolutely pure compound.

$$u(CRM) = \sqrt{\frac{\sum u(CRM_i)^2}{n}} \qquad \text{Equation 5}$$

RMS_{bias} reflects the relative difference between measured and expected values for reference materials that were not used as part of the calibration curve. It is the root-mean-square of the difference ($bias_i$) between the observed ($Clab_i$) and known (CRM_i) δ values of the QC RMs within each analytical session (Equations (6) and (7)). It is important to clarify that determining the mean observed δ value for a particular RM across *all sessions* and comparing to the known value will not accurately reflect the uncertainty in the measurements as is discussed in more detail in FAQ1.

$$RMS_{bias} = \sqrt{\frac{\sum (bias_i)^2}{n}} \qquad \text{Equation 6}$$

$$bias_i = Clab_i - CRM_i \qquad \text{Equation 7}$$

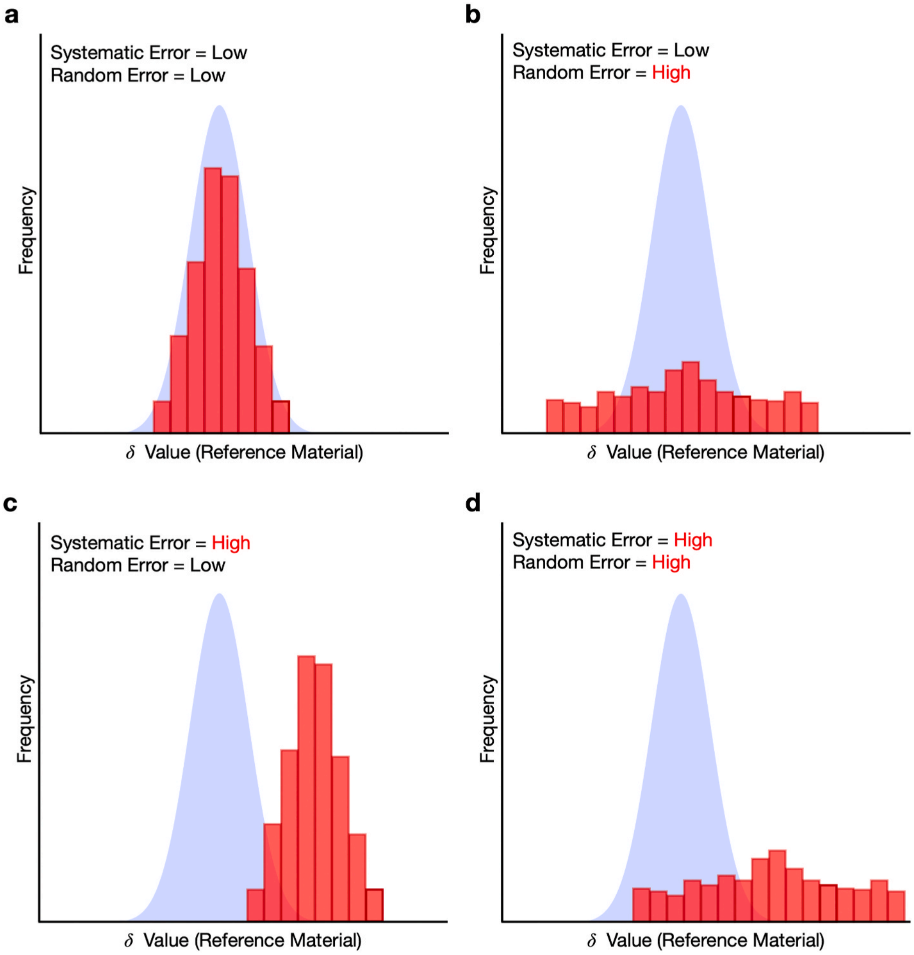

Fig. 2. Four examples showing low and high amounts of random and systematic error for a given reference material (RM). In each panel, the shaded Gaussian distribution represents the known δ value of the RM and the bars represent measured δ values within a given session.

For this sheet to most accurately reflect measurement uncertainty, it is especially important that the QC RMs are similar to the unknown samples being measured in terms of: (1) δ values, (2) elemental composition, (3) amount of gas produced.

3. Description of the sheets

The uncertainty calculator consists of an Excel workbook with six sheets (Online Supplementary Material). The current formatting works for δ^{13}C, δ^{15}N, and δ^{34}S measurements, although it could easily be modified for any other measurements. Below is a description of each of the sheets included, which are currently filled with a mix of simulated δ^{13}C, δ^{15}N, and δ^{34}S data as an example.

3.1. Data sheet

This sheet provides the primary area where users will paste in their data. It is recommended that users make use of the Paste Values function, rather than the simple Paste. This sheet contains conditional formatting that will highlight any replicate entries in the Sample column; this formatting would apply to both samples and RMs. Failing to use the Paste Values function will remove the conditional formatting. It is important to enter any replicate samples with exactly the same identifier (i.e., 29335 and 29335 rather than 29335A and 29335B or 29335 and 29335 DUP). Similarly, RMs that have been analyzed multiple times must have exactly the same identifier (i.e., USGS40, USGS40, USGS40 not USGS40-1, USGS40-2, USGS40-3). In addition to the conditional formatting, which is only a visual cue, the spreadsheet requires that these entries be identical to facilitate the calculations of the mean and variance within each session. The analytical session column contains dates in the example included here, but any identifier can be used and a date is not strictly necessary. Times of analysis of samples and RMs should not be used as each will be different and it will prevent the sheet from identifying the analytical session(s) and grouping the RMs accordingly. The Sample; Session column contains formulas that are used for assigning groups of RMs or samples to particular sessions and does not need to be modified. There are approximately 1000 rows in the data file, but this can be expanded as needed, provided the formulas in the Sample; Session column are dragged down as well.

3.2. RMs sheet

The second sheet lists all the reference materials, both international and internal, that were used in the analysis. For convenience, most of the materials available with certified δ^{13}C, δ^{15}N, and δ^{34}S values from USGS and IAEA are provided with relevant links. Note that these values will not update automatically if there are changes made to the accepted δ values, as has happened in the past (Qi et al., 2003). Users should always ensure that the known δ values reported on the RMs sheet match what they used in their own analysis. For example, USGS40 has a δ^{13}C value that is reported to three decimal places −26.389 ‰, but some may choose to use this RM to anchor their calibration curve with a known value that is rounded to −26.39 ‰. Users can add their own reference materials to this table, entering the long-term observed δ values ± one standard deviation. The Type column allows the user to select from a dropdown menu and indicate whether the RM was used in the calibration/normalization (Calibration) or if the material was not used in the calibration/normalization and was instead used as a QC RM (QC). If the reference material information is entered, but the Type column is left blank, this/these RM(s) will not factor into the uncertainty calculations and will be treated as sample replicates.

3.3. 13C, 15N, 34S sheets

These sheets require no input from the user and will populate automatically. The first four columns (A-D) provide the summarized aspects of the uncertainty calculation from Szpak et al. (2017). Columns E and F provide a summarized list of the unique RMs that were used during the sessions; this list includes all RMs rather than only those that contained the relevant element (e.g., in the example USGS40 is included in the δ^{34}S sheet but does not factor into any of those calculations). Users should avoid including δ values for elements that are not present in a given RM; for example, do not include δ^{15}N values for IAEA-CH-7 on the Data Sheet. Columns H through K summarize all the samples that were analyzed more than once during the sessions, with Column I specifying whether the sample was a duplicate, triplicate, and so on. Column J calculates the standard deviation of each replicate and Column K calculates the sum of squares for each replicate. Column M list the unique analytical sessions based on the information in the Data sheet. Columns O-AH summarize the mean δ values observed for each RM within each session. Columns AI-BB summarize the differences between the known and measured δ values for each of the QC RMs, excluding those that are identified as calibration RMs. The standard deviations of the calibration and QC RMs are summarized in Columns BC-BV and BW-CP, respectively. CQ-DJ summarizes the number of each RM that was analyzed in each session. DK-ED presents the total variance (sum of squares) for each RM and EE-EX presents the known standard deviation of those RMs that were used as QCs.

3.4. Summary tables

This sheet conveniently summarizes important information about the RMs and sample replicates. The information presented here is not used in any of the calculations but may be useful for compiling summary tables to report in a supplementary file (see FAQ15).

4. Frequently asked questions (FAQ)

FAQ1. *Why do I need to differentiate different analytical sessions (runs)?*

The differentiation of analytical sessions is most important with respect to the measured δ values of QC RMs and the calculation of systematic error (u_{bias}). To illustrate this point, consider a simple example focusing only on δ^{15}N wherein the three reference materials were used as QC RMs (long-term δ^{15}N values in parentheses): IRM-1 (+1.87 ± 0.15 ‰), IRM-14 (+21.51 ± 0.17 ‰), IRM-26 (+14.71 ± 0.19 ‰). The δ^{15}N values observed across three sessions are presented in Table 1 (note that these values do not correspond to those that are used in the example uncertainty calculation).

The δ^{15}N values of all three RMs are systematically off relative to the long-term values by approximately +0.2 ‰ in the 2024-06-18 session, right on the mark in the 2024-08-08 session, and off by approximately −0.2 ‰ in the 2024-11-24 session. Across all sessions, the mean δ^{15}N values for these RMs that are within 0.01 ‰ of the long-term δ^{15}N values. Since the δ^{15}N values of all the RMs were systematically off by +0.2 ‰ in the first session, and the long-term δ^{15}N values of these RMs span a wide range (approximately 20 ‰), it is safe to assume that the δ^{15}N values of any samples of similar matrix analyzed on this day would also be systematically off by +0.2 ‰. Similarly, all the samples analyzed in the 2024-11-24 session would likely be systematically off by −0.2 ‰. Not differentiating the three analytical sessions would create a situation where the RMS_{bias} would be near 0, and therefore the systematic error (u_{bias}) would be equivalent to *u(Cref)*, simply reflecting the long-term variability in the RMs' δ^{15}N values. Specifically, when the sessions are differentiated for the example here, u_{bias} is 0.27 ‰ and when they are not u_{bias} is 0.17 ‰. Therefore, not differentiating RMs by analytical session will tend to underrepresent systematic error.

Table 1
Example δ^{15}N data for three RMs not included in the calibration curve across three analytical sessions (for FAQ1).

Sample	$\delta^{15}N_{AIR}$/‰	Analytical Session	Mean $\delta^{15}N_{AIR}$/‰	StDev $\delta^{15}N_{AIR}$/‰
IRM-1	2.35	2024-06-18		
IRM-1	2.31	2024-06-18		
IRM-1	2.18	2024-06-18		
IRM-1	2.03	2024-06-18		
IRM-1	2.14	2024-06-18	2.20	0.13
IRM-1	1.86	2024-08-08		
IRM-1	1.87	2024-08-08		
IRM-1	1.87	2024-08-08		
IRM-1	1.84	2024-08-08		
IRM-1	1.94	2024-08-08	1.88	0.04
IRM-1	1.35	2024-11-24		
IRM-1	1.53	2024-11-24		
IRM-1	1.57	2024-11-24		
IRM-1	1.75	2024-11-24		
IRM-1	1.64	2024-11-24	1.57	0.15
IRM-1		**All Sessions**	**1.88**	**0.29**
IRM-14	21.69	2024-06-18		
IRM-14	21.71	2024-06-18		
IRM-14	21.87	2024-06-18		
IRM-14	21.58	2024-06-18	21.72	0.12
IRM-14	21.43	2024-08-08		
IRM-14	21.64	2024-08-08		
IRM-14	21.48	2024-08-08		
IRM-14	21.53	2024-08-08	21.52	0.09
IRM-14	21.35	2024-11-24		
IRM-14	21.37	2024-11-24		
IRM-14	21.28	2024-11-24		
IRM-14	21.15	2024-11-24	21.29	0.10
IRM-14		**All Sessions**	**21.51**	**0.21**
IRM-26	14.87	2024-06-18		
IRM-26	15.03	2024-06-18		
IRM-26	14.66	2024-06-18		
IRM-26	14.85	2024-06-18		
IRM-26	14.96	2024-06-18		
IRM-26	14.93	2024-06-18		
IRM-26	14.96	2024-06-18		
IRM-26	15.13	2024-06-18		
IRM-26	15.00	2024-06-18		
IRM-26	15.07	2024-06-18		
IRM-26	14.94	2024-06-18	14.95	0.12
IRM-26	14.66	2024-08-08		
IRM-26	14.95	2024-08-08		
IRM-26	14.76	2024-08-08		
IRM-26	14.52	2024-08-08		
IRM-26	14.81	2024-08-08		
IRM-26	14.88	2024-08-08		
IRM-26	14.65	2024-08-08		
IRM-26	14.71	2024-08-08		
IRM-26	14.45	2024-08-08		
IRM-26	14.66	2024-08-08	14.70	0.15
IRM-26	14.58	2024-11-24		
IRM-26	14.47	2024-11-24		
IRM-26	14.36	2024-11-24		
IRM-26	14.60	2024-11-24		
IRM-26	14.33	2024-11-24		
IRM-26	14.23	2024-11-24		
IRM-26	14.68	2024-11-24		
IRM-26	14.41	2024-11-24		
IRM-26	14.46	2024-11-24		
IRM-26	14.46	2024-11-24	14.46	0.13
IRM-26		**All Sessions**	**14.71**	**0.24**

FAQ2. *What happens if I use a reference material as a QC in one session and for calibration in another?*

This scenario should not cause any problems, however, it is important to appropriately differentiate the RMs that were used to calibrate vs. those that were not. For example, assume that USGS62 is used as part of the calibration curve in two sessions and is used as a QC in another. In the RMs sheet, there should be two lines for USGS62, with the Name being differentiated as USGS62-CA and USGS62-CH. In the Type column, USGS62-CA could be specified with Calibration from the dropdown and USGS62-CH could be specified with QC. In the Data sheet, these USGS62s would also need to be differentiated accordingly, taking care to ensure that the text matches exactly.

FAQ3. *What happens if my in-house reference material is not listed in the RMs sheet?*

Adding reference materials is as easy as typing them in on the next available row, no other changes are required to any of the sheets. At least 100 rows of RMs can be accommodated with the existing formulas. If a reference material is entered on the Data sheet but is not also listed on the RMs sheet or is listed on the RMs sheet but is not identified as either a QC or Calibration, it will be treated like a sample replicate in the calculations.

FAQ4. *What if I am using the δ^{15}N values of one element to calibrate and the δ^{34}S values only as a QC RM because there is no certified value, such as for IAEA-N-1 or IAEA-N-2?*

This case is slightly more complicated and will also require two rows in the RMs sheet, for example IAEA-N-1N (with Calibration in the Type column) and IAEA-N-1S (with QC in the Type column). In the Data sheet, there will need to be two rows for each instance of IAEA-N-1, one specified as IAEA-N-1N containing only the δ^{15}N measurements and one specified as IAEA-N-1S containing only the δ^{34}S measurements.

FAQ5. *How should my data be rounded and to how many decimal places should I report the uncertainty?*

The spreadsheet presented here does not round values, but it can be modified relatively easily by the user with the ROUND function in Excel according to their preferences. As a default, u_c is rounded to two decimal places and all the other non-counting metrics (e.g., s_{rep}, RMS_{bias}) are reported to two decimal places but are not rounded. This spreadsheet offers no mechanism to improve the resolution of isotopic measurements, it simply provides a means with which to quantify the uncertainty of these measurements. At the very least, the number of decimal places reported by the user should never exceed those of the reference materials that they have used to calibrate and assess the accuracy of their measurements. In experimenting with the sheet, using δ values that have not been rounded does not produce meaningfully different uncertainties than using δ values that have been rounded.

FAQ6. *The calculated uncertainty is much higher than I expected or is several times greater than what the lab that produced the data stated in the report that I was provided with. Why?*

It is relatively common for commercial labs to report metrics such as the long-term uncertainty of RMs that they regularly analyze. As discussed in FAQ1, this is not a good representation of the uncertainty of the individual user's data because it fails to take into account variation among analytical sessions, especially as it relates to systematic error. Furthermore, the uncertainty commonly reported by commercial labs is typically for extremely homogenous reference materials that may not be perfectly analogous to the user's samples. Since this uncertainty sheet considers the random error in both samples and RMs, it is more appropriate to report these values than it is to report the long-term variance on a reference material.

FAQ7. *I have not analyzed any samples in duplicate and the sheet generates an error, what do I do?*

Sample replicates are important because they provide information about the isotopic homogeneity of the materials being analyzed, which may differ from the RMs that were analyzed alongside those samples. There may be some instances in which sample replication is impossible. For instance, some of the methods for producing incremental samples of tooth dentine produce so little material that sample replication is difficult, if not impossible, particularly for smaller teeth. If replicates were not analyzed, at a minimum, the s_{rep} should be set as equal to s_{RM}. This can be achieved by modifying the sheet and setting C6 as being equal to C3 in the 13C, 15N, 34S sheets. In practical terms, this assumes that the variance of the samples would be the same as the variance in the RMs that were analyzed. In reporting the uncertainty in an instance such as this, it would be important to clarify that sample replicates were not analyzed.

FAQ8. *I used a process standard or a method duplicate, what do I do with this?*

Performing pretreatments on a homogenous material that is perfectly matrix-matched to the samples is an excellent way to ensure that there are no systematic or random errors introduced by sample processing. It is critical to ensure that this material is indeed isotopically homogenous, which may be particularly challenging for some types of materials, such as bones (Hall et al., 2025). These process standards are similar to QC RMs in that they have a known δ value that has been characterized over time, but they are likely to have a greater uncertainty because of their large size. Accordingly, they should be treated like QC RMs when using the uncertainty sheet with long-term δ values being added to the RMs sheet.

FAQ9. *How do I differentiate RMs used to correct for linearity or drift?*

It is not necessary to differentiate RMs that are used as simple QCs from those that are used to correct for linearity or drift as part of the uncertainty calculation. Any shifts in δ value caused by non-linear effects or instrumental drift should be reflected in the δ valuess of the RMs, with residual drift or linearity effects producing larger systematic and/or random errors. It would, however, be important to report which RMs may have been used to correct for linearity and drift, being explicit about the intervals at which they were analyzed (e.g., every 10 samples for drift) and the relative mass of each element analyzed (e.g., 20–200 μg N for linearity).

FAQ10. *The lab that I contacted will not provide me with the information about the RMs that they analyzed alongside my samples.*

Not providing this information upon request is unusual and it may be advisable to look for an alternative provider.

FAQ11. Can I create different uncertainties for each analytical session?

It is possible to do so and would require a different file to be used for each session. If all of the samples included across the different sessions are composed of the same material and are part of the same project, it may make more sense to have a single uncertainty value to present for all of the samples. Creating session-specific uncertainties would necessitate reporting the different uncertainties at the level of individual samples in a supplementary file.

FAQ12. *I have conducted a methodological study of different sample pretreatments. One of the pretreatments produced much more variable δ values in the sample material, what should I do?*

In this situation it would make sense to determine uncertainties for the different sample pretreatments using separate spreadsheets. Doing so would be informative, providing insight into the material being produced under different conditions, aiding the objectives of the study.

FAQ13. *What are the implications for the uncertainty calculation if my reference materials are not matrix-matched to my samples, as is the case with sulfur?*

The extent to which this is an issue is unclear, but whenever possible, matrix-matching the reference materials and unknown samples should be prioritized (Balint et al., 2024). Prior to the introduction of the glutamic acid RMs USGS40 and USGS41 in 2003, it was commonplace to calibrate δ^{15}N values of organic samples using two inorganic (ammonium sulfate) RMs (IAEA-N-1 and IAEA-N-2), which has since been demonstrated to be potentially problematic (Gentile et al., 2013). Relative to inorganic substances, glutamic acid is obviously a better match for organic samples, such as proteins, but the δ values reported for UGS40 and USGS41 were themselves determined in reference to inorganic materials. A separate issue may exist for sulfur because most available reference materials well-suited for calibration (e.g., silver sulfide and barium sulfate) have two orders of magnitude more sulfur than the collagenous samples that are commonly analyzed from archaeological contexts. In this case, there is little to be done until a more robust set of isotopically disparate organic reference materials (analogous to USGS40 and USGS41) become available for sulfur. Reporting which RMs were used and how they performed is the best that can be done under these circumstances.

FAQ14. *If I am using a reference material that contains carbon but not nitrogen (*e.g.*, IAEA-CH-6), will the sheet still treat this as a reference material for C and N?*

In this instance, it is necessary to delete any δ^{15}N values associated with the IAEA-CH-6 in the Data Sheet since these are not meaningful results. The sheet will only recognize RMs based on their listing in the RMs sheet, and does not differentiate those RMs that contain some elements but not others.

FAQ15. *How should I report the information from this uncertainty calculator?*

Reporting more information is always better than reporting less. The entire workbook could be included as a supplementary file, which would provide a convenient means of presenting the data that was produced in the study alongside all the RMs that accompanied the samples. The Summary Tables sheet provides a relatively convenient summary of the relevant information. In the main text of a paper, much of this information would be cumbersome to report, but Szpak et al. (2017) provide a useful example of what could be presented as a summary. In addition to the uncertainty calculation method, important information about how the samples were prepared and analyzed, the δ values assigned to the reference materials used for normalization (these assigned δ values may change over time), the δ values assigned to the reference materials used as quality controls (checks) and their traceability would all be important to report as outlined with examples in Skrzypek et al. (2022).

FAQ16. *How many sample replicates do I need for a robust uncertainty calculation?*

The robusticity of the uncertainty calculation increases with more sample replicates and this is indicated by the degrees of freedom associated with sample replicates df_{rep}. Reporting this value, along with the degrees of freedom associated with the reference materials analyzed alongside the samples (df_{RM}), would help to ensure that readers are aware of the relative reliability of the uncertainty estimate. The number of replicates will vary depending on the nature of the samples being analyzed (i.e., homogeneity), the budget of the project, the amount of material available, and the δ values. For example, our lab analyzes a higher percentage of sample replicates for bulk leaf material than for bone collagen because leaves are a less pure material that are not homogenized as effectively as collagen during sample pretreatment. Furthermore, we analyze a higher percentage of sample replicates when determining δ^{34}S relative to δ^{13}C and δ^{15}N because of greater concerns around memory effects.

Declaration of competing interest

None.

Acknowledgements

Members of Paul Szpak's research group (Trent Environmental Archaeology Lab) and Lesley Chesson's isotope team at the Defense POW/MIA Accounting Agency provided useful feedback on an earlier version of this sheet that improved its functionality. Angela Pettipas provided assistance with checking the sheet for errors. This work was supported by the Canada Research Chairs program and an NSERC Discovery Grant (2020-04740).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jas.2025.106457.

References

Balint, S., Schwartz, M., Fowler, D.N., Linnekogel, S., Clemons, S.C., Burkemper, L.K., 2024. Experimental assessment of elemental analyzer isotope ratio mass spectrometry normalization methodologies for environmental stable isotopes. Rapid Commun. Mass Spectrom. 38, e9837.

Carter, J.F., Fry, B., 2013. Ensuring the reliability of stable isotope ratio Data—Beyond the principle of identical treatment. Anal. Bioanal. Chem. 405, 2799–2814.

Coleman, M., Meier-Augenstein, W., 2014. Ignoring IUPAC guidelines for measurement and reporting of stable isotope abundance values affects us all. Rapid Commun. Mass Spectrom. 28, 1953–1955.

Coplen, T.B., 2011. Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. Rapid Commun. Mass Spectrom. 25, 2538–2560.

Dunn, P.J.H., Brodie, C., Strak, E., Carter, J.F., 2025. Good practice guide for isotope ratio mass spectrometry. Forensic IsotopeRatio Mass Spectrometry, third ed.

Dunn, P.J.H., Malinovsky, D., Goenaga-Infante, H., 2020. Calibration hierarchies for light element isotope delta reference materials. Rapid Commun. Mass Spectrom. 34, e8711.

Dunn, P.J.H., Malinovsky, D., Holcombe, G., Cowen, S., Goenaga-Infante, H., 2021. Guidance for characterization of in-house reference materials for light element stable isotope analysis. Rapid Commun. Mass Spectrom. 35, e9177.

Dunn, P.J.H., Skrzypek, G., 2023. Perspective: hidden biases in isotope delta results and the need for comprehensive reporting. Rapid Commun. Mass Spectrom. 37, e9623.

Gentile, N., Rossi, M.J., Delémont, O., Siegwolf, R.T.W., 2013. δ15N measurement of organic and inorganic substances by EA-IRMS: a speciation-dependent procedure. Anal. Bioanal. Chem. 405, 159–176.

Hall, O.R., Derian, A.A.Y., Rausch, A., McCuaig, J., Szpak, P., 2025. Documenting the extent of intra-bone isotopic variation. Available at SSRN 5025228 https://doi.org/10.2139/ssrn.5025228.

Jardine, T., Cunjak, R., 2005. Analytical error in stable isotope ecology. Oecologia 144, 528–533.

Joint Committee for Guides in Metrology, 2008. Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement. International Organization for Standardization, Geneva.

Kragten, J., 1994. Tutorial review. Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique. Analyst 119, 2161–2165.

Meier-Augenstein, W., Schimmelmann, A., 2018. A guide for proper utilisation of stable isotope reference materials. Isot. Environ. Health Stud. 1–16.

Meija, J., Chartrand, M.M.G., 2018. Uncertainty evaluation in normalization of isotope delta measurement results against international reference materials. Anal. Bioanal. Chem. 410, 1061–1069.

Paul, D., Skrzypek, G., Fórizs, I., 2007. Normalization of measured stable isotopic compositions to isotope reference scales – a review. Rapid Commun. Mass Spectrom. 21, 3006–3014.

Qi, H., Coplen, T.B., Geilmann, H., Brand, W.A., Böhlke, J.K., 2003. Two new organic reference materials for δ^{13}C and δ^{15}N measurements and a new value for the δ^{13}C of NBS 22 oil. Rapid Commun. Mass Spectrom. 17, 2483–2487.

Skrzypek, G., 2013. Normalization procedures and reference material selection in stable HCNOS isotope analyses: an overview. Anal. Bioanal. Chem. 405, 2815–2823.

Skrzypek, G., Allison, C.E., Böhlke, J.K., Bontempo, L., Brewer, P., Camin, F., Carter, J.F., Chartrand, M.M.G., Coplen, T.B., Gröning, M., Hélie, J.-F., Esquivel-Hernández, G., Kraft, R.A., Magdas, D.A., Mann, J.L., Meija, J., Meijer, H.A.J., Moossen, H., Ogrinc, N., Perini, M., Possolo, A., Rogers, K.M., Schimmelmann, A., Shemesh, A., Soto, D.X., Thomas, F., Wielgosz, R., Winchester, M.R., Yan, Z., Dunn, P.J.H., 2022. Minimum requirements for publishing hydrogen, carbon, nitrogen, oxygen and sulfur stable-isotope delta results (IUPAC technical report). Pure Appl. Chem. 94, 1249–1255.

Skrzypek, G., Sadler, R., Paul, D., 2010. Error propagation in normalization of stable isotope data: a Monte Carlo analysis. Rapid Commun. Mass Spectrom. 24, 2697–2705.

Szpak, P., Metcalfe, J.Z., Macdonald, R.A., 2017. Best practices for calibrating and reporting stable isotope measurements in archaeology. J. Archaeol. Sci. Rep 13, 609–616.

Szpak, P., Savelle, J.M., Conolly, J., Richards, M.P., 2019. Variation in late holocene marine environments in the Canadian arctic archipelago: evidence from ringed seal bone collagen stable isotope compositions. Quat. Sci. Rev. 211, 136–155.