Uncertainty evaluation in normalization of isotope delta measurement results against international reference materials

Meija, Juris; Chartrand, Michelle

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/s00216-017-0659-1
Analytical and Bioanalytical Chemistry, 410, 3, pp. 1061-1069, 2017-12-07

Uncertainty evaluation in normalization of isotope delta measurement results against international reference materials

Juris Meija and Michelle Chartrand

Measurement Science and Standards, National Research Council Canada, 1200 Montreal Road, Ottawa ON K1A 0R6, Canada

E-mail: juris.meija@nrc-cnrc.gc.ca

Abstract. Isotope delta measurements are normalized against international reference standards. Although multi-point normalization is becoming a standard practice, the existing uncertainty evaluation practices are either undocumented or are incomplete. For multi-point normalization we present errors-in-variables regression models for explicit accounting of the measurement uncertainty of the international standards along with the uncertainty that is attributed to their assigned values. We also present framework to account for the uncertainty that arises due to small number of replicates in measurements. This manuscript also discusses multi-laboratory data reduction while accounting for inevitable correlations between the laboratories due to the use of same reference materials for calibration between the laboratories. Both frequentist and Bayesian methods of uncertainty analysis are discussed.

Keywords: Isotope delta; normalization; uncertainty evaluation; random effects model

Publication record
Analytical and Bioanalytical Chemistry (2017) to be submitted.

1. Introduction

In 2015, the International Committee for Weights and measures (CIPM) noted that assigned values for isotope delta replacement materials should be done through a formal internationally vetted procedure that assures the continued comparability of delta value measurements. Recent large-scale international isotope delta intercomparison exercise [1] has shown that the statistical analysis of isotope delta materials can be indeed difficult and that more guidance on this matter seems necessary.

The current viewpoint is that normalization based on two or more international standards is the best method for isotope delta analysis [2]. Despite this, the existing uncertainty evaluation practices are largely undocumented or incomplete [3]. In this vein, we present mathematical framework for linear normalization of isotope delta measurement results against secondary international reference materials with exclusive emphasis to the uncertainty evaluation. This article does not address the various prerequisite quality control measures that analysts must undertake during the data generation itself such as the uncertainties due to sample preparation, drift and blank correction[4, 5].

2. General aspects

2.1. Definitions and nomenclature

Isotope delta is the relative difference of the isotope amount ratios of an element between two materials:

$$\delta_{\mathrm{A,B}}(^{13}\mathrm{C}) = R_{\mathrm{A}}/R_{\mathrm{B}} - 1 \qquad (1)$$

The material A is the sample and the material B is the reference standard. The above notation is variably shortened to $\delta_{\mathrm{B}}(^{13}\mathrm{C})$, $\delta(^{13}\mathrm{C})_{\mathrm{B}}$, $\delta^{13}\mathrm{C}_{\mathrm{B}}$, or simply $\delta^{13}\mathrm{C}$. The numerical values of isotope deltas are almost universally reported in parts per thousand (‰) which often leads to appearance of extraneous multiplying factors (0.001 or 1000) in the quantity equations involving isotope deltas [6].

At the Beijing IUPAC General Assembly in 2005, CIAAW recommended that $\delta(^{13}\mathrm{C})$ values of all carbon-bearing materials be measured and expressed relative to the VPDB on a scale normalized by assigning consensus values of −46.6 ‰ to LSVEC lithium carbonate and +1.95 ‰ to NBS 19 calcium carbonate, and authors should clearly state so in their reports[7]. Thus, the measurement results for a given sample are expressed against agreed-upon scale which sets the numerical value of carbon isotope deltas in two international reference materials.

In practice, however, not everyone employs LSVEC or NBS 19 in their daily measurements. One reason for this is that other secondary reference materials are sometimes desirable in order to better match the nature of the calibrators and samples identical treatment. For example, hair reference materials might be more desirable than carbonates when analyzing human hair. Another reason for using secondary isotopic reference materials is out of practical necessity since NBS 19 (also known as the 'toilet seat' limestone) is no longer available.

Since isotope delta measurements cannot always be performed directly against the scale-defining materials, one has to rely on a variety of secondary international reference materials.

2.2. Calibration against secondary reference materials

In addition to scale-defining primary isotope delta reference materials, there are a number of natural or synthetic compounds which have been carefully calibrated versus the primary calibration materials. The list of such materials is published and maintained by IUPAC-CIAAW along with the agreed-upon isotope delta values[8].

The characterization of the secondary isotope delta reference materials is not owned by the institutions that disseminate the physical materials. For this reason, several property values might be available for the same material. As an example, IUPAC lists published delta values from $\delta_{\mathrm{VPDB}}(^{13}\mathrm{C}) = -30.03$ to −29.40 ‰ for NBS 22 oil[8] whereas providers of this reference material give markedly different uncertainty estimates: $-30.03(6)_{k=1}$ ‰ (USGS), $-30.03(9)_{k=2}$ ‰ (NIST) and $-30.031(43)_{k=1}$ ‰ (IAEA). Consequently, when reporting isotope delta measurement results, it is essential that the values assigned to primary and secondary materials are given along the measurement results.

2.3. Calibration algorithms

There are several algorithms which convert measured isotope delta values of a sample to the calibrated international scale: single-, two- and multi-point linear normalization [9]. A two-point calibration, for example, invokes two reference standards whose isotope deltas are measured against an in-house internal standard. The observed results (d) are then plotted against the assigned known isotope delta values (δ) for the primary standards and the linear relationship between these data serves as the calibration function:

$$\delta_{\mathrm{X}} = \delta_1 + \frac{\delta_2 - \delta_1}{d_2 - d_1}(d_{\mathrm{X}} - d_1) \qquad (2)$$

Some authors prefer the model equation of the two-point calibration to be expressed in a different, yet

equivalent, mathematical form [10]:

$$\delta_X = \frac{(nd_X + 1)(\delta_1 + 1)}{(nd_1 + 1)} - 1 \quad (3)$$

where

$$n = \frac{\delta_1 - \delta_2}{d_1 - d_2 - \delta_1 d_2 + \delta_2 d_1} \quad (4)$$

Uncertainty evaluation of isotope delta measurement results with two-point calibration can be done by applying the law of uncertainty propagation to the above measurement models.

Two-point calibration can be naturally extended to multi-point calibration. In this case, however, one no longer can write down a physical measurement model as in Eq. 2. Rather, the measurement model equation becomes a statistical model relating the observed isotope delta values from N standards, d_i ($i = 1 \ldots N$), and the known isotope delta values of these standards, δ_i, in a linear relationship:

$$d_i = a + b \cdot \delta_i + e_i \quad (5)$$

The residual error is typically modeled as normal random variable, $e_i \sim \mathrm{N}(0, \sigma^2)$. Eq. 5 is a statistical model, hence it must be solved for the unknown parameters a and b (and the hyper-parameter σ) using statistical methods which we discuss later. The isotope delta for a given sample, δ_X, is then obtained from the uncalibrated measured value, d_X, by solving the above calibration equation:

$$\delta_X = \frac{d_X - a}{b} \quad (6)$$

2.4. Sources of uncertainty

Multi-point calibration of isotope delta measurement results against international reference materials is summarized visually in Fig. 1.

There are three sources of uncertainty that should be considered in the multi-point calibration of isotope delta measurement results: (1) the uncertainty due to measurement of isotope delta of the sample material, (2) uncertainty due to isotope delta measurement of the primary reference materials, and (3) the uncertainty associated with the isotope delta values of the reference materials used.

Ordinary linear regression is frequently employed to obtain calibration coefficients and their uncertainties. However, such approach neglects the uncertainty that is inherent to the isotope delta values of the reference materials used. To account for this, one can employ errors-in-variables statistical model

$$d_i = a + b(\delta_i + u_i) + e_i \quad (7)$$

which explicitly reflects the fact that isotope delta values of the standards, δ_i, themselves are accompanied with uncertainties, $u_i \sim \mathrm{N}(0, u^2(\delta_i))$.

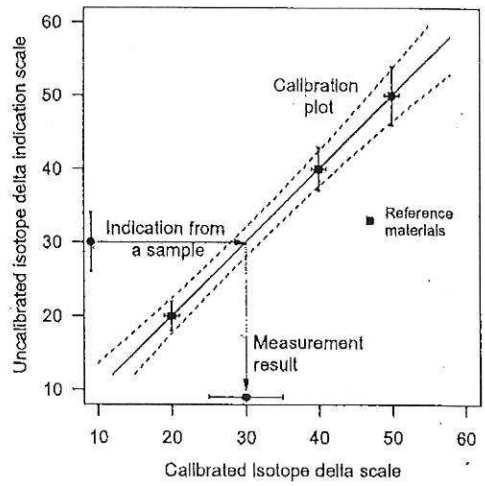

Figure 1. Sources of uncertainty in the isotope delta measurements using multi-point calibration

This statistical model can be solved using generalized distance regression in conjunction with the Monte Carlo uncertainty evaluation. In addition to the above sources of uncertainty, we also consider the possible biases in the results obtained by the various laboratories using random effects statistical models.

3. Single-laboratory results

3.1. Calibration function

Typically, the best multi-point calibration line is obtained using least squares fitting where one assumes that the values of the independent variable are known without error. In the situations encountered in isotope delta measurements, however, both variables d_i and δ_i have non-negligible uncertainties and calibration line must therefore be estimated using statistical methods that accommodate this observation. Such cases can be modeled with the errors-in-variables regression method as described, for example, in ISO 6143 [11]. The weighted least squares procedure minimizes the sum of S_0 for all values of i:

$$S_0(a, b)_i = \frac{(d_i - a - b\delta_i)^2}{u^2(d_i)} \quad (8)$$

In contrast, the errors-in-variables regression involves minimization the sum of S_1 for all values of i:

$$S_1(a, b, \hat{\delta}_i)_i = \frac{(d_i - a - b\hat{\delta}_i)^2}{u^2(d_i)} + \frac{(\delta_i - \hat{\delta}_i)^2}{u^2(\delta_i)} \quad (9)$$

with respect to regression parameters a, b and the unknown 'true' values of δ_i, $\hat{\delta}_i$. The uncertainties $u(d_i)$ and $u(\delta_i)$ are assumed to be known. Regression

based on minimizing the sum $S_{1,i}$ is also known as the generalized Deming regression (available in R using function deming from the package of the same name).

Guenther and Possolo have put forward a modified equations for errors-in-variables regression which recognize the fact that the uncertainties of d_i are typically based on a small number of degrees-of-freedom [12]. This additional source of uncertainty can be accounted for in the errors-of-variables regression by replacing the elements of normal probability density distribution to Student's t-distribution. Subsequently, one minimizes the sum of S_2 for all values of i:

$$S_2(a,b,\hat{\delta}_i)_i = (v_{d,i}+1)\log\left(1+\frac{(d_i-a-b\hat{\delta}_i)^2}{v_{d,i}u^2(d_i)}\right) + (v_{\delta,i}+1)\log\left(1+\frac{(\delta_i-\hat{\delta}_i)^2}{v_{\delta,i}u^2(\delta_i)}\right) \quad (10)$$

where $v_{d,i}$ and $v_{\delta,i}$ are the degrees-of-freedom associated to the estimates $u^2(d_i)$ and $u^2(\delta_i)$, respectively. If unavailable, the degrees-of-freedom that are associated with the uncertainties of the international reference materials can be taken as a sufficiently large number, say $v_{\delta,i} = 100$.

3.2. Uncertainty assessment

Multi-point calibration of isotope deltas typically involves measurements made of three to six international reference materials. While the parameters of linear regression (intercept and slope) can be obtained using the various methods described earlier, the uncertainty of the regression parameter estimates is best estimated using the Monte Carlo method [13]. This consists of applying perturbations to the values of all quantities consistent with the available data. Below is a general description of the Monte Carlo method for evaluating the uncertainty of the isotope delta measurement results with errors-in-variable regression using criterion S_2:

for $k = 1 \ldots K$ (typically $K = 1000$ or higher)

(i) Simulate values for $u(\delta_i)$ as

$$u_k(\delta_i) = u(\delta_i)\sqrt{\frac{v_{\delta,i}}{\chi^2(v_{\delta,i})}}$$

where $\chi^2(v)$ is a sample drawn from chi-squared distribution with v degrees-of-freedom.

(ii) Simulate values for δ_i as random draws from normal distribution with mean δ_i and uncertainty $u_k(\delta_i)$, $\delta_{k,i} = \mathrm{N}(\delta_i, u_k^2(\delta_i))$.

(iii) Simulate values for $u(d_i)$ as

$$u_k(d_i) = u(d_i)\sqrt{\frac{v_{d,i}}{\chi^2(v_{d,i})}}$$

where $\chi^2(v)$ is a sample drawn from chi-squared distribution with v degrees-of-freedom.

(iv) Simulate values for d_i as random draws from normal distribution with mean d_i and uncertainty $u_k(d_i)$, $d_{k,i} = \mathrm{N}(d_i, u_k^2(d_i))$.

(v) Minimize the sum of S_2 in Eq. 10 for all values of i to obtain estimates a_k and b_k. In R this can be done using the general-purpose optimization function optim.

(vi) Simulate the value for the measurement of the sample X, $d_{k,\mathrm{X}}$, similar to how it was done in the preceding steps. As above, this involves simulating the uncertainty $u_k(d_\mathrm{X})$.

(vii) Obtain value $\delta_{k,\mathrm{X}}$ from Eq. 6.

The uncertainty assessments for δ_X are produced from the obtained set of simulated measurement results δ_X which can be conveniently visualized as histograms.

4. Combining multi-laboratory results

Isotope delta results obtained by numerous laboratories for the same material are often compared in order to assign a consensus value. The data reduction problem can be summarized as follows: given a number of measured values x_j ($j = 1 \ldots M$), along with quoted standard uncertainties u_j, the goal is to combine these values into a single best estimate. In order to achieve this, a statistical model is needed to codify the relationship between the data. For this purpose, we adopt a random laboratory effects model for the reported data which represents each one of them as an additive superposition of three effects [14, 15]:

$$x_j = \mu + \lambda_j + \varepsilon_j \quad (11)$$

Here x_j is the reported isotope delta value by jth laboratory, μ is the consensus value of all laboratory results, λ_j denotes the effect of the jth laboratory, which is modeled as a normal random variable with mean 0 and variance τ^2, $\lambda_j \sim \mathrm{N}(0,\tau^2)$, and ε_j denotes measurement error, $\varepsilon_j \sim \mathrm{N}(0, u^2(x_j))$.

The random effects model captures any dark uncertainty, τ^2, expressed in the reported isotope deltas being more scattered than their values of $u(x_j)$ would intimate. Consequently, we accept that the stated uncertainties do not represent complete knowledge. The within-study variances $u^2(x_j)$ are known beforehand whereas the between-study variance τ^2 and the consensus value μ must both be estimated from the data.

In practice, consensus isotope delta values can be obtained from the reported laboratory results and their uncertainties using a variety of methods. Albeit not without its shortcomings, DerSimonian and

Laird method is a popular choice among simple and non-iterative consensus estimators [16]. In addition, Bayesian approaches using non-informative reference priors are gaining widespread acceptance [16,17].

4.1. Correlations between the laboratory results

Often, the laboratory results are not independent because some laboratories might involve same international reference standards for calibration. As an example, during the recent large-scale certification campaign of several carbon isotope delta materials [1], USGS61 caffeine was measured by nine laboratories most of which relied on four common isotope delta reference materials: LSVEC, NBS19, USGS40, and USGS41. Since the uncertainty of these reference materials is a significant component in the overall uncertainty budget, the results from these laboratories cannot be treated as independent.

The between-laboratory correlations can be accounted for in the random effects model (Eq. 11) as follows:

$$\boldsymbol{x} \sim \mathrm{N}(\mu \boldsymbol{1}, \boldsymbol{V} + \tau^2 \boldsymbol{I}) \tag{12}$$

where $\boldsymbol{V}$ is the covariance matrix of x_j which contains the quoted uncertainties and covariances (the diagonal elements of this matrix equal the squared standard uncertainties and the off-diagonal elements are determined as $V_{ij} = \rho_{ij} u(x_i) u(x_j)$ where ρ_{ij} is the correlation between the results x_i and x_j. The laboratory results are represented in Eq. 12 in a form of a vector $\boldsymbol{x}$, and $\boldsymbol{I}$ is the identity matrix (of size $M \times M$). In programming language R, the model in Eq. 12 can be fitted to the data using several approaches. As an example, with the restricted maximum likelihood method one can use the command `rma.mv(X, V, method="reml")`, where the function `rma.mv` is defined in the R package metafor, X is a vector of isotope delta measurement results, $\boldsymbol{x} = \{x_1, \ldots, x_M\}$, and V is their covariance matrix.

4.2. Determination of correlations between the laboratory results

Correlations between the reported laboratory results can be evaluated using Monte Carlo method. With few modifications, this involves the same approach as detailed in Section 3.2. That procedure provides a Monte Carlo simulation of a result from a single laboratory and all the steps that lead to a value $\delta_{k,\mathrm{X}}$ have to be parallelized across all laboratories with the exception of those that perturb the values of $u(\delta_i)$ and δ_i.

for $k = 1 \ldots K$ (typically $K = 1000$ or higher)

(i) Simulate values for $u(\delta_i)$

(ii) Simulate values for δ_i

Table 1. Example dataset for the carbon-13 delta measurements in NIST SRM 350b benzoic acid at NRC.[a]

Material, X	$d_{\mathrm{X,S}}(^{13}\mathrm{C})$	N	$\delta_{\mathrm{X,VPDB}}(^{13}\mathrm{C})$
SRM 350b	+12.235(22) ‰	10	(*measurand*)
IAEA-CH-6	+30.458(27) ‰	3	−10.449(33) ‰
IAEA-CH-7	+8.141(21) ‰	3	−32.151(50) ‰
IAEA-600	+12.729(12) ‰	3	−27.771(43) ‰
USGS40	+14.128(30) ‰	3	−26.39(4) ‰
USGS62	+26.040(46) ‰	3	−14.79(4) ‰
USGS65	+20.355(17) ‰	3	−20.29(4) ‰

[a] Standard deviation from N replicate measurements is given in the parenthesis. $d_{\mathrm{X,S}}(^{13}\mathrm{C})$ is the measurement result of carbon-13 isotope delta in sample X against in-house reference gas S.

(iii) For each laboratory $j = 1 \ldots M$, obtain the value of $\delta_{k,\mathrm{X}}$ which we denote here as $\delta_{k,j}$. This follows the steps (iii)–(vii) of Section 3.2.

The outcome of the above Monte Carlo simulation is a $K \times M$ matrix D containing simulated values of δ_{X} which provides the estimate of the covariance between the results of any two laboratory result necessary to construct matrix $\boldsymbol{V}$.

5. Examples

In order to illustrate the methods described above and the impact of various assumptions that can be taken during the data reduction and interpretation, we provide two detailed examples. Section 5.1 describes a single laboratory result whereas Section 5.2 details a data reduction process from several laboratories providing partially correlated results.

5.1. Single-laboratory results for NIST SRM 350b

As a part of a larger study, measurements of carbon-13 isotope delta in NIST SRM 350b benzoic acid primary (acidimetric) standard were performed at NRC in 2017. For this purpose, four internationally recognized stable isotope reference materials were selected for calibration: sucrose (IAEA-CH-6), polyethylene (IAEA-CH-7), caffeine (IAEA-600), and glutamic acid (USGS40)[8] along with two additional reference materials available from the US Geological Survey: caffeine (USGS62) and glycine (USGS65) [1]. The relevant measurement results and the assigned isotope delta values for the six reference materials are listed in Table 1.

A simple approach to establish the isotope delta value for SRM 350b is to perform ordinary least squares fit from the six standards with complete disregard to their respective uncertainties. The resulting intercept and the slope are then used in Eq. 5, together with the measured value $d_{\mathrm{X,S}}$ for SRM 350b to obtain a value of $\delta_{\mathrm{VPDB}}(^{13}\mathrm{C}) = -28.210$ ‰ for SRM 350b.

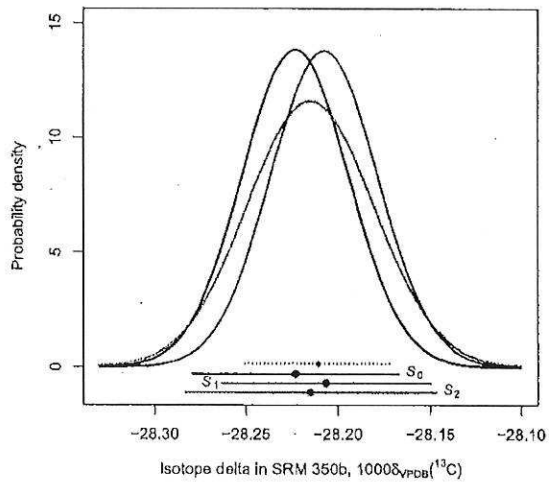

Figure 2. Probability densities corresponding to a Monte Carlo sample of size $K = 1 \times 10^5$ using various statistical approaches to determine the calibration function: weighted ordinary regression (best fit criterion S_0, Eq. 8), errors-in-variables regression which accounts for uncertainties in both the predictor and outcome variables (S_1, Eq. 9), and modified errors-in-variables regression while also accounting for the fact that all measurement uncertainty estimates are based on small degrees-of-freedom (S_2, Eq. 10). Also shown with dotted horizontal line is the naive approach that builds calibration plot by disregarding all uncertainties. Mean values and 95 % coverage intervals (horizontal lines) for the carbon-13 isotope delta in NIST SRM 350b benzoic acid are shown below.

The uncertainty propagation calculations are done from Eq. 5 by taking into account the correlation between the intercept and the slope, and by using $0.022/\sqrt{10}$ ‰ as the uncertainty for the mean value of $d_{X,S}$ of SRM 350b. The uncertainty calculations can be greatly facilitated using the NIST uncertainty machine. We obtain standard uncertainty of 0.020 ‰ for $\delta_{VPDB}(^{13}C)$.

The naive procedure described above is presented only for comparative purposes to demonstrate the need for explicit accounting of all measurement uncertainties; statistical methods that fully incorporate all the available knowledge (as the criterion S_2 does), gives nearly twice larger uncertainty. Six-point calibration line using weighted linear regression (best fit criterion S_0, Eq. 8) provides $\delta_{VPDB}(^{13}C) = -28.223$ ‰ for SRM 350b with standard uncertainty of 0.029 ‰. If, however, the calibration is established using errors-in-variables regression (criterion S_1, Eq. 9), we obtain $\delta_{VPDB}(^{13}C) = -28.206(29)$ ‰. Criterion S_2 gives a more conservative uncertainty estimate of $\delta_{VPDB}(^{13}C) = -28.215(34)$ ‰ as shown in Fig. 2. While the average point estimates resulting from all methods considered herein give nearly identical results, same cannot be said about the uncertainty estimates.

Table 2. Example dataset for the carbon-13 delta measurements relevant to the NRC sugar isotope delta reference material BEET-1.[a,b]

Material, X	Lab.	$d_{X,S}(^{13}C)$	N
BEET-1	A	−26.028(85) ‰	12
BEET-1	B	−26.054(59) ‰	12
BEET-1	C	−26.037(37) ‰	12
BEET-1	D	−25.987(73) ‰	12
IAEA-CH-6	A	−10.418(79) ‰	4
IAEA-CH-6	B	−10.450(14) ‰	4
IAEA-CH-6	C	−10.455(66) ‰	4
IAEA-CH-6	D	−10.465(71) ‰	4
USGS40	A	−26.385(73) ‰	4
USGS40	B	−26.430(71) ‰	4
USGS40	C	−26.455(66) ‰	4
USGS40	D	−26.402(61) ‰	4
USGS62	A	−14.840(92) ‰	4
USGS62	B	−14.797(13) ‰	4
USGS62	C	−14.846(30) ‰	4
USGS62	D	−14.675(182) ‰	4

[a] Standard deviation from N replicate measurements is given in the parenthesis.
[b] Isotope delta values for the three reference materials (IAEA-CH-6, USGS40, USGS62) are given in Table 1.

Note that ordinary least squares (weighted or not), does not provide unbiased parameter estimates when the predictor variable has measurement uncertainty [20]. Generally speaking, uncertainty in the predictor variable leads to negatively biased estimates of the calibration slope and, in turn, biased value for the isotope deltas.

5.2. Multi-laboratory results for NRC BEET-1

A variability in carbon isotope ratio of approx. 180 ‰ is observed in naturally occurring terrestrial materials. Sugar beet has $\delta_{VPDB}(^{13}C) \approx -14$ ‰ whereas sugar cane has $\delta_{VPDB}(^{13}C) \approx -28$ ‰. Given that sugar cane is one of the major sources of added sugar in processed food, isotope delta standards can enable tracking added sugars[21]. For these reasons, NRC has recently produced a suite of isotope delta sugar reference materials aided by interlaboratory comparison study. We use a small portion of this dataset to illustrate the data handling from multi-laboratory results as summarized in Table 2. Although Table 2 lists the reported calibrated isotope delta measurement results from four laboratories, for the purposes of this example we treat these results as non-calibrated indications, $d_{X,S}(^{13}C)$ instead of $\delta_{X,VPDB}(^{13}C)$.

Following the general procedure outlined in Section 4.2, we obtain estimates of $\delta_{VPDB}(^{13}C)$ for BEET-1 by each of the four laboratories. This involves a three-point calibration using errors-in-variables regression with criterion S_2. The following

results were obtained:

$$\begin{array}{llll} \text{A} & -26.022(78)\ ‰ & \text{B} & -26.013(52)\ ‰ \\ \text{C} & -26.009(70)\ ‰ & \text{D} & -26.035(76)\ ‰ \end{array} \quad (13)$$

In addition, the following correlations are obtained between the results of the four laboratories:

$$\mathrm{cor}(\delta_i, \delta_j) = \begin{bmatrix} 1 & 0.39 & 0.27 & 0.25 \\ 0.39 & 1 & 0.38 & 0.36 \\ 0.27 & 0.38 & 1 & 0.22 \\ 0.25 & 0.36 & 0.22 & 1 \end{bmatrix} \quad (14)$$

Fitting these data to the multivariate meta-analysis model of random laboratory effects (Eq. 12) provides with a consensus value $\delta_{\mathrm{VPDB}}(^{13}\mathrm{C}) = -26.017(46)$ ‰ for BEET-1 from the example dataset. In R this can be achieved by using code metafor::rma.mv(x,V) where x is the vector of mean laboratory results (Eq. 14) and V is their covariance matrix.

Dismissing positive correlations between the laboratories has the effect of significantly lowering the uncertainty of the consensus value to $\delta_{\mathrm{VPDB}}(^{13}\mathrm{C}) = -26.018(33)$ ‰. Hence, it is important to evaluate the possible correlations between the laboratories in order to avoid unrealistic (incorrect) uncertainty assessments.

Establishing consensus estimate from partially correlated datasets is not unique to this study. Bodnar *et al.* describe similar problem to determine fundamental constants from various studies using Bayesian methods [17]. Analysis of the NRC BEET-1 data can also be performed using Bayesian paradigm. For this purpose, we have adopted the following Bayesian model:

(i) Data

- results from each laboratory, δ_j $(j = 1 \ldots 4)$
- covariances of the laboratory results, V

(ii) Priors

- non-informative prior for the consensus isotope delta value, $\alpha \sim \mathrm{N}(0, 10^5)$
- non-informative prior for the between-laboratory variance term, $\tau^{-2} \sim \mathrm{G}(10^{-4}, 10^{-4})$, using gamma distribution

(iii) Likelihood (data generation model)

laboratory results are like outcomes of multivariate normal distribution, $\delta_j \sim \mathrm{N}(\mu_j, V)$ with hierarchical model for the laboratory effects, $\mu_j \sim \mathrm{N}(\alpha, \tau^2)$

The model parameters (consensus isotope delta value and the between-laboratory variance) can be obtained using Markov chain Monte Carlo method [22]. Alternatively, the above model (with slightly different specification of the non-informative priors according to the currently favored principles of Berger and Bernardo) can be solved analytically[17]. Using Markov Chain Monte Carlo method, we obtain consensus value $\delta_{\mathrm{VPDB}}(^{13}\mathrm{C}) = -26.018(53)$ ‰ for BEET-1 from the example dataset. For this purpose, we use openBUGS software from R as shown in the Listing 1. The posterior probability densities for the consensus isotope delta and the between-laboratory uncertainty are both shown in Fig. 3.

```
library(R2OpenBUGS)

cor2cov = function(cor, u) diag(u)%*%cor%*% diag(u)

HB <- function(df) {
# DATA ##
delta = t(matrix(df$mu))
sigma = solve(cor2cov(df$cor, df$u))
N = length(delta)

# MODEL ##
model = function() {
  # priors
    tau ~ dgamma(1.0E-4,1.0E-4)
    alpha ~ dnorm(0,1.0E-5)

  # likelihood
    for(i in 1:N) mu[i] ~ dnorm(alpha, tau)
    delta[1,1:N] ~ dmnorm(mu[], sigma[,])
  }

file = file.path(tempdir(), "model.txt")
write.model(model, file)

data = list("delta", "sigma", "N")
par = c("alpha","tau")
init = function() list(alpha=median(delta), tau=1/mad
     (df$u)^2)
out = bugs(data, init, par, file, n.iter=50000)
out$summary
}

# Input data
df = list(
       mu = c(-26.022, -26.013, -26.009, -26.035),
        u = c(0.078, 0.052, 0.070, 0.076),
      cor = matrix(c(1.000,0.391,0.271,0.251,
            0.391,1.000,0.376,0.358,
            0.271,0.376,1.000,0.224,
            0.251,0.358,0.224,1.000),ncol=4)
              )

HB(df) # fit data to the Bayesian random effects model
##              mean    sd    2.5%   97.5%
## alpha     -26.018 0.053 -26.120 -25.920
```

Listing 1. Implementation of Markov Chain Monte Carlo method for Bayesian inference of the interlaboratory random effects model with regard to possible correlations between the laboratory results in R programming language with openBUGS software.

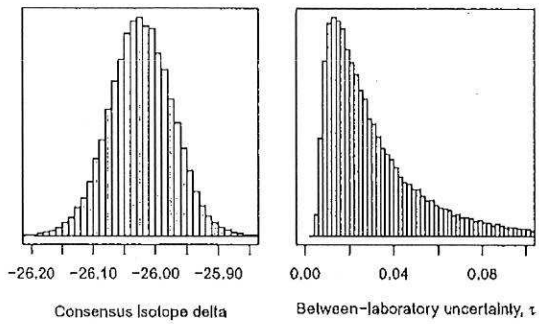

Figure 3. Posterior distribution of isotope delta consensus value and the between-laboratory uncertainty.

6. Conclusions

We have outlined data reduction framework for normalizing isotope delta measurement results which are applicable for two- and multi-point normalization algorithms. We prefer the adoption of random effects models in multi-laboratory data reduction routines as they allow to interpret mutually inconsistent observed interlaboratory results to be consistent with a common consensus value. In addition, Bayesian paradigm might be preferred over the classical statistical methods as it obviates the need for asymptotic arguments when dealing with small sample sizes.

Acknowledgments

The authors have benefited from discussions with Blaza Toman, National Institute of Standards and Technology, USA.

References

[1] Schimmelmann A, Qi H, Coplen T B, Brand W A, Fong J, Meier-Augenstein W, Kemp H F, Toman B, Ackermann A, Assonov S, Aerts-Bijma A T, Brejcha R, Chikaraishi Y, Darwish T, Elsner M, Gehre M, Geilmann H, Gröning M, Hélie J F, Herrero-Martín S, Meijer H A J, Sauer P E, Sessions A L and Werner R A 2016 *Analytical Chemistry* **88** 4294–4302 URL https://doi.org/10.1021%2Facs.analchem.5b04392

[2] Skrzypek G 2012 *Analytical and Bioanalytical Chemistry* **405** 2815–2823 URL https://doi.org/10.1007/s00216-012-6517-2

[3] Szpak P, Metcalfe J Z and Macdonald R A 2017 *Journal of Archaeological Science: Reports* **13** 609–616 URL https://doi.org/10.1016/j.jasrep.2017.05.007

[4] Werner R A and Brand W A 2001 *Rapid Communications in Mass Spectrometry* **15** 501–519 URL https://doi.org/10.1002/rcm.258

[5] Dunn P J H, Hai L, Malinovsky D and Goenaga-Infante H 2015 *Rapid Communications in Mass Spectrometry* **29** 2184–2186 URL https://doi.org/10.1002/rcm.7376

[6] Coplen T B 2011 *Rapid Communications in Mass Spectrometry* **25** 2538–2560 URL https://doi.org/10.1002/rcm.5129

[7] Wieser M E 2006 *Pure and Applied Chemistry* **78** 2051–2066 URL https://doi.org/10.1351/pac200678112051

[8] Brand W A, Coplen T B, Vogl J, Rosner M and Prohaska T 2014 *Pure and Applied Chemistry* **86** 425–467 URL https://doi.org/10.1515/pac-2013-1023

[9] Paul D, Skrzypek G and Fórizs I 2007 *Rapid Communications in Mass Spectrometry* **21** 3006–3014 URL https://doi.org/10.1002/rcm.3185

[10] Carter J F and Barwick V J 2011 *Good Practice Guide for Isotope Ratio Mass Spectrometry* (FIRMS) ISBN 978-0-948926-31-0

[11] ISO 6143:2001 2001 *Gas analysis – comparison methods for determining and checking the composition of calibration gas mixtures* International Organization for Standardization Geneva, Switzerland

[12] Guenther F R and Possolo A 2011 *Analytical and Bioanalytical Chemistry* **399** 489–500 URL https://doi.org/10.1007/s00216-010-4379-z

[13] Possolo A M 2015 *Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Technical Note 1900* (NIST)

[14] Kacker R N 2004 *Metrologia* **41** 132–136 URL https://doi.org/10.1088%2F0026-1394%2F41%2F3%2F004

[15] Toman B and Possolo A 2009 *Accred Qual Assur* **14** 553–563 URL http://dx.doi.org/10.1007/s00769-009-0547-2

[16] Koepke A, Lafarge T, Possolo A and Toman B 2017 *Metrologia* **54** S34–S62 URL https://doi.org/10.1088/1681-7575/aa6c0e

[17] Bodnar O, Elster C, Fischer J, Possolo A and Toman B 2016 *Metrologia* **53** S46–S54 URL https://doi.org/10.1088/0026-1394/53/1/s46

[18] Linsinger T P J, Pauwels J, van der Veen A, Schimmel H and Lamberty A 2001 *Accreditation and Quality Assurance* **6** 20–25 URL https://doi.org/10.1007/s007690000261

[19] van der Veen A, Linsinger T P and Pauwels J 2001 *Accreditation and Quality Assurance* **6** 26–30 URL https://doi.org/10.1007/s007690000238

[20] Carroll R J, Ruppert D, Stefanski L A and Crainiceanu C M 2006 *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition* (Chapman and Hall/CRC) ISBN 1584886331

[21] Callier V 2015 *Science* **348** 488–488 URL https://doi.org/10.1126%2Fscience.348.6234.488

[22] Possolo A and Toman B 2007 *Metrologia* **44** 464–475 URL https://doi.org/10.1088/0026-1394/44/6/005