



HAL
open science

Full propagation of analytical uncertainties in $\Delta 47$ measurements

Mathieu Daëron

► **To cite this version:**

Mathieu Daëron. Full propagation of analytical uncertainties in $\Delta 47$ measurements. *Geochemistry, Geophysics, Geosystems*, 2021, 22 (5), pp.e2020GC009592. 10.1029/2020GC009588. hal-03094085v1

HAL Id: hal-03094085

<https://hal.science/hal-03094085v1>

Submitted on 4 Jan 2021 (v1), last revised 26 Nov 2021 (v2)

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.

Full propagation of analytical uncertainties in Δ_{47} measurements

M. Daëron⁽¹⁾

(1) *Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France. daeron@lsce.ipsl.fr*

Abstract

Clumped-isotope measurements in CO₂ and carbonates (Δ_{47}) present a number of technical challenges and require correcting for various sources of analytical non-linearity. For now we lack a formal description of the analytical errors associated with these correction steps, which are not accounted for in most data processing methods currently in use. Here we formulate a quantitative description of Δ_{47} error propagation, fully taking into account standardization errors and their properties. We describe various standardization strategies, along with the assumptions they rely on, in the context of this model, and propose a new, “pooled” standardization approach designed to yield more robust/accurate corrections. User-friendly online resources and an open-source Python library are also provided to facilitate the use of these error models. Among other uses, the mathematical framework described here may be helpful to improve standardization protocols (e.g., anchor/unknown ratios) and inform future efforts to define community reference materials.

1 Introduction

Clumped-isotope geochemistry is the study of statistical anomalies in the abundance of multiply substituted isotopologues in natural materials [Eiler & Schauble, 2004; Eiler, 2011]. Mass spectrometric measurements of Δ_{47} , quantifying the excess abundance of ¹³C – ¹⁸O bonds in CO₂ and, by extension, in carbonate minerals, constitute the most widely used branch of this relatively young but rapidly evolving field of research. The appeal of carbonate Δ_{47} measurements is largely based on the fact that the clumped-isotope compositions of natural carbonates directly or indirectly constrain their crystallization temperatures and/or thermal histories, with a broad range of Earth science applications. Establishing a robust calibration of the carbonate Δ_{47} thermometer, however, has long remained a vexing challenge, with inter-laboratory discrepancies equivalent to large uncertainties in reconstructed temperatures, sometimes exceeding 10 °C [e.g., Bonifacie *et al.*, 2017; Petersen *et al.*, 2019].

25 Despite keeping in mind that “true” calibration differences between certain types of carbonates
 26 are not to be excluded *a priori*, various potential causes for these discrepancies have been put for-
 27 ward, such as (a) inconsistent or inaccurate ^{17}O correction parameters [Daëron *et al.*, 2016; Schauer
 28 *et al.*, 2016; Olack & Colman, 2019], (b) systematic effects arising from different data processing
 29 methods, and (c) poorly-corrected analytical biases resulting from instrumental and/or methodolog-
 30 ical differences between laboratories. Petersen *et al.* [2019] tested the first two of these hypotheses
 31 and found that using unified methods for ^{17}O correction and subsequent data processing reduced
 32 inter-laboratory discrepancies without eliminating them. Testing the third hypothesis is one of the
 33 goals of the recently completed inter-laboratory comparison exercise “InterCarb”, whose results are
 34 reported in a companion study [Bernasconi *et al.*, 2021].

35 How to accurately estimate the analytical uncertainties affecting Δ_{47} measurements constitutes
 36 a distinct but related issue. Although Fernandez *et al.* [2017] pointed out that non-robust statis-
 37 tics based on small numbers of observations frequently yield underestimated uncertainties, there is
 38 no consensus today as to whether this is the primary cause of inter-laboratory discrepancies. Both
 39 Daëron & Blamart [2016] and Kocken *et al.* [2019] called attention to the uncertainties associated
 40 with Δ_{47} standardization (i.e. conversion of “raw Δ_{47} ” measurements to “absolute” values), but for
 41 now we lack an explicit, formal description of this source of analytical error, which most data pro-
 42 cessing methods do not currently account for. This issue is critical in the context of the InterCarb
 43 exercise, which aims to test whether different laboratories, when analyzing a common set of four
 44 unknown and three reference carbonate samples, obtain analytically consistent results, i.e. results
 45 displaying no more inter-laboratory scatter than expected based on intra-laboratory analytical errors.
 46 The present work aims to formulate a comprehensive model of analytical errors in Δ_{47} measure-
 47 ments, including those arising from standardization using carbonate and/or carbon dioxide stan-
 48 dards; to describe various standardization strategies along with the assumptions they rely on; and to
 49 provide user-friendly data processing tools implementing these error models.

50 2 Methods

51 2.1 A brief summary of mass spectrometric measurements of Δ_{47} in carbonates

52 Clumped-isotope analyses of carbonates are typically performed using dual-inlet gas-source isotope-
 53 ratio mass spectrometry. In each analysis, a certain amount of carbonate mineral reacts with pure
 54 phosphoric acid at a fixed temperature (usually 25, 70 or 90 °C). Each such reaction typically pro-
 55 duces between 1 and 100 μmol of CO_2 , which is collected in a series of cryogenic traps and carefully
 56 purified to eliminate isobaric contaminants (i.e. species with a molecular mass of 47 amu, or com-
 57 pounds liable to produce such species through fragmentation/recombination reactions). Although
 58 our primary focus is on measurements of Δ_{47} in carbonate minerals, most aspects of the present study
 59 apply just as well to Δ_{47} in CO_2 samples which were not produced by acid digestion of carbonates.

60 The purified CO_2 is then introduced to the “sample” side of a dual-inlet system and, from there,
 61 into a Nier-type ion source. In most systems currently in use, analyte CO_2 is compared with a “work-
 62 ing gas” reference CO_2 through the frequent, regular toggling of a change-over valve. The bulk
 63 isotopic composition ($\delta^{13}\text{C}$, $\delta^{18}\text{O}$) and mass-47 to mass-44 abundance ratio of each analyte are deter-

64 mined by comparing ion currents for the analyte and the working gas, averaged over long integration
 65 times, typically tens of minutes or longer. These integration times are necessary because counting
 66 statistics are one of the primary factors limiting precision when observing rare isotopologues such
 67 as $^{16}\text{O}^{13}\text{C}^{18}\text{O}$, which makes up only 45 ppm of natural CO_2 [Huntington *et al.*, 2009].

68 For the past decade Δ_{47} measurements have been standardized by comparison with specially
 69 prepared CO_2 standards with known clumped-isotope compositions and variable bulk isotope com-
 70 positions [Dennis *et al.*, 2011]. Carbonate reference materials have increasingly been also used for
 71 standardization, either in addition to or as a replacement for CO_2 standards [Meckler *et al.*, 2014;
 72 Bernasconi *et al.*, 2018]. Although here we primarily consider standardization using carbonate ref-
 73 erence materials, the mathematical framework presented below generally applies as well to CO_2
 74 standards.

75 2.2 Terminology

76 We define below, in the context of this work, a number of terms. A **sample** is an amount of pre-
 77 sumably homogeneous carbonate material subjected to one or more **analyses** (otherwise known
 78 as replicate measurements/observations). Each analysis corresponds to a single acid reaction fol-
 79 lowed by purification of the evolved CO_2 and by a series of dual-inlet IRMS measurements, yielding
 80 **working-gas delta values** (δ_{45} to δ_{49}). These working-gas deltas are then converted to “raw” (non-
 81 standardized) values of $\delta^{13}\text{C}$, $\delta^{18}\text{O}$, and Δ_{47}^{raw} . The specifics of this conversion have been extensively
 82 covered elsewhere [e.g., Huntington *et al.*, 2009; Daëron *et al.*, 2016], and are not directly relevant to
 83 the topics discussed here. Analyses are generally grouped into **sessions**, each of them usually cor-
 84 responding to a given time span over which analytical conditions are presumed to have remained
 85 stable. One key assumption is that the various analytical/instrumental non-linearities which affect
 86 Δ_{47}^{raw} observations remain constant over the duration of each session. These non-linearities include
 87 a **scrambling effect** likely reflecting recombination of isotopologues in the IRMS gas source or else-
 88 where in the sample preparation apparatus [Dennis *et al.*, 2011]; a **compositional slope** reflecting
 89 small biases in the electrical background of the ion beam measurements [He *et al.*, 2012; Bernasconi
 90 *et al.*, 2013]; and a **working gas offset** resulting from the (knowingly inaccurate) assumption that
 91 the dual-inlet working gas is stochastic. Within each session, the samples/analyses are divided into
 92 two groups: **anchors**, whose Δ_{47} values are assigned *a priori*, and **unknowns**, whose Δ_{47} values are
 93 to be determined. Here we define a **standardization model** as any mathematical procedure aiming
 94 to estimate these unknown Δ_{47} values by comparing the anchor and unknown analyses, explicitly
 95 or implicitly constraining analytical non-linearities within each session.

96 2.3 Objectives and strategy

97 Our aim is to model how random, zero-centered, presumably Gaussian measurement errors propa-
 98 gate into final, “absolute” Δ_{47} values averaged over a number of analyses/sessions. We do not attempt
 99 to account for non-random biases such as those potentially arising, for instance, from errors in the
 100 isotopic composition of the working gas, or from assigning inaccurate Δ_{47} values to one or more
 101 anchors. The models described here will hopefully provide a framework to report more accurate es-
 102 timates of the uncertainty associated with clumped-isotope measurements, and inform the choices
 103 we make in the laboratory.

104 We start by describing a general formulation of the standardization function used to compute
 105 the “absolute” Δ_{47} value of each analysis. Quantifying the parameters defining this function within
 106 a given session is equivalent to constraining the analytical/instrumental non-linearities mentioned
 107 above, and may be treated as a classical least-squares minimization problem.

108 We follow up by estimating the analytical precision of “raw” measurements (before standard-
 109 ization) based, following oft-repeated recommendations, on the pooled external repeatability of a
 110 group of standards and/or unknown samples. The general formulation used here then makes it
 111 straightforward to propagate the raw measurement errors into the “autogenic” uncertainty of each
 112 analysis (that directly arising from the raw errors of this particular analysis) and an independent
 113 component of “allogenic” uncertainties arising from the least-squares model errors, i.e. from the
 114 standardization itself.

115 We finish by describing the general properties of these two components of error, and briefly
 116 discuss several practical standardization approaches applicable to real-world data sets.

117 2.4 Standardization to an “absolute” Δ_{47} reference frame within a single session

118 Computing “absolute” Δ_{47} values traditionally involves two chained affine transformations designed
 119 to correct for known instrumental non-linearities (eqs. 5-6 of *Dennis et al.* [2011], using the original
 120 notation):

$$\Delta_{47\text{-[SGvsWG]0}} = \Delta_{47\text{-[SGvsWG]}} - \delta_{\text{[SGvsWG]}}^{47} \times \text{Slope}_{\text{EGL}} \quad (1)$$

$$\Delta_{47\text{-RF}} = \Delta_{47\text{-[SGvsWG]0}} \times \text{Slope}_{\text{ETF}} + \text{Intercept}_{\text{ETF}} \quad (2)$$

121 This is mathematically equivalent to the following formulation:

$$\Delta_{47}^{\text{raw}} = a \Delta_{47} + b \delta_{47} + c \quad (3)$$

122 In this equation, the parameters (a , b , c) respectively account for scrambling effects, the compo-
 123 sitional slope, and the working gas offset. To estimate these parameters, a natural approach is to use
 124 classical least-squares minimization methods, treating Δ_{47}^{raw} as the response/dependent variable and
 125 (Δ_{47} , δ_{47}) as explanatory variables. Despite uncertainties on δ_{47} usually being as large as those on
 126 Δ_{47}^{raw} , the former may safely be treated as an explanatory variable because b is typically small enough
 127 (10^{-2} or less) for errors on δ_{47} to have a negligible impact.

128 Without compelling reasons to do otherwise, we assign equal weights to all measurements be-
 129 longing to the same session. The best-fit standardization parameters (a, b, c) for any given session
 130 are thus those minimizing the following χ^2 statistic, summed over all anchor analyses within that
 131 session (unknown analyses are not considered here because their Δ_{47} values are not known *a priori*):

$$\chi^2 = \sum (\Delta_{47}^{\text{raw}} - a \Delta_{47} - b \delta_{47} - c)^2 \quad (4)$$

132 This computation, whose mathematical details are summarized in appendix A, yields a triplet
 133 of best-fit values for (a, b, c), thus defining the standardization function of eq. (3) for this session.
 134 It also yields a covariance matrix V_0 for the best-fit values of (a, b, c). At this stage, the covariance
 135 matrix is unscaled, meaning that it only constrains the *relative* scaling between model standard er-
 136 rors and covariances in (a, b, c). The additional piece of information needed to scale these model
 137 errors is the uncertainty assigned to each observation, i.e. the analytical precision of individual Δ_{47}^{raw}
 138 measurements.

139 2.5 Estimating the analytical precision of raw measurements

140 The uncertainty assigned to individual Δ_{47}^{raw} measurements, noted σ_{47}^{raw} , may be quantified in various
 141 ways, but always keeping in mind that over-reliance on the statistics of small numbers is problematic
 142 [Fernandez *et al.*, 2017]. We propose that in most cases a robust estimate of σ_{47}^{raw} can be obtained
 143 by considering carbonate samples deemed free of contaminants and isotopically homogeneous, be
 144 them anchors, unknown samples, or carbonate standards treated as unknowns, and computing the
 145 pooled Δ_{47} repeatability of analyses within this group:

$$\sigma_{47}^2 = \frac{1}{N_a - N_s} \sum (\Delta_{47} - \overline{\Delta_{47}})^2 \quad (5)$$

146 where $\overline{\Delta_{47}}$ is the average Δ_{47} value for the sample considered, N_a the total number of analyses
 147 considered, and N_s the number of different samples considered. The Δ_{47}^{raw} repeatability of analyses
 148 is then:

$$(\sigma_{47}^{\text{raw}})^2 = (a \sigma_{47})^2 = \frac{1}{N_a - N_s} \sum (\Delta_{47}^{\text{raw}} - a \overline{\Delta_{47}} - b \delta_{47} - c)^2 \quad (6)$$

149 It bears noting that if the group of samples used to estimate σ_{47}^{raw} within a single session only
 150 comprises three anchors, then $(\sigma_{47}^{\text{raw}})^2$ is equal to the reduced chi-squared statistic $\chi^2/(N_a - 3)$ for
 151 that session. In such case, scaling the standardization errors by σ_{47}^{raw} is equivalent to the common
 152 practice of estimating least-square model errors based on the scatter/variance of residuals. Taking
 153 additional samples into account increases confidence in our estimate by virtue of increasing the sta-
 154 tistical degrees of freedom ($N_f = N_a - N_s$), on the condition that the replicability of these additional
 155 samples is equal to (or indistinguishable from) that of carbonate standards. In our experience this
 156 condition is frequently met when samples are well-mixed, finely ground, relatively pure carbonate
 157 powders.

158 **2.6 Propagation of standardization errors within a single session**

159 Regardless of its estimation method, σ_{47}^{raw} may now be used to quantify the standard model errors
 160 ($\sigma_a, \sigma_b, \sigma_c$) on the best-fit standardization parameters and their covariances (c_{ab}, c_{bc}, c_{bc}):

$$\begin{bmatrix} \sigma_a^2 & c_{ab} & c_{ac} \\ c_{ab} & \sigma_b^2 & c_{bc} \\ c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} = (\sigma_{47}^{\text{raw}})^2 V_0 \quad (7)$$

161 These model errors and covariances fully describe the standardization uncertainty associated
 162 with anchor measurement errors, and can now be propagated explicitly to the session average Δ_{47}
 163 value of a given unknown sample, noted $\overline{\Delta_{47}}$:

$$\overline{\Delta_{47}} = (\overline{\Delta_{47}^{\text{raw}}} - b \overline{\delta_{47}} - c)/a \quad \Rightarrow \quad \sigma(\overline{\Delta_{47}})^2 = J \times C \times J^T \quad (8)$$

164 with $\overline{\Delta_{47}^{\text{raw}}}$ and $\overline{\delta_{47}}$ being the session average values of Δ_{47}^{raw} and δ_{47} , respectively; J the Jacobian
 165 matrix of $\overline{\Delta_{47}}$; and C the covariance matrix of $(\overline{\Delta_{47}^{\text{raw}}}, a, b, c)$:

$$J = \left[\frac{\partial \overline{\Delta_{47}}}{\partial \overline{\Delta_{47}^{\text{raw}}}}, \frac{\partial \overline{\Delta_{47}}}{\partial a}, \frac{\partial \overline{\Delta_{47}}}{\partial b}, \frac{\partial \overline{\Delta_{47}}}{\partial c} \right] = \frac{1}{a} \left[1, -\overline{\Delta_{47}}, -\overline{\delta_{47}}, -1 \right] \quad (9)$$

$$C = \begin{bmatrix} \sigma(\overline{\Delta_{47}^{\text{raw}}})^2 & 0 & 0 & 0 \\ 0 & \sigma_a^2 & c_{ab} & c_{ac} \\ 0 & c_{ab} & \sigma_b^2 & c_{bc} \\ 0 & c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} = \begin{bmatrix} (\sigma_{47}^{\text{raw}})^2/N_a & 0 & 0 & 0 \\ 0 & \sigma_a^2 & c_{ab} & c_{ac} \\ 0 & c_{ab} & \sigma_b^2 & c_{bc} \\ 0 & c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} \quad (10)$$

166 The structure of the above covariance matrix makes it clear that $\sigma(\overline{\Delta_{47}})^2$ for an unknown sample
 167 is the sum of two statistically independent sources of error: an “autogenic” component σ_u reflecting
 168 uncertainties in Δ_{47}^{raw} measurements for that sample, and an “allogenic” component σ_s reflecting
 169 uncertainties in the standardization model:

$$\sigma(\overline{\Delta_{47}})^2 = \sigma_u^2 + \sigma_s^2 \quad (11)$$

$$\sigma_u^2 = \sigma_{47}^2/N_a \quad (12)$$

$$\sigma_s^2 = \frac{1}{a^2} \left(\overline{\Delta_{47}}^2 \sigma_a^2 + \overline{\delta_{47}}^2 \sigma_b^2 + \sigma_c^2 + 2(\overline{\Delta_{47}} \overline{\delta_{47}} c_{ab} + \overline{\Delta_{47}} c_{ac} + \overline{\delta_{47}} c_{bc}) \right) \quad (13)$$

170 **2.7 Combining data from several independent sessions**

171 As long as the standardization of each session only takes into accounts analyses from that session, the
 172 values of $\overline{\Delta_{47}}$ computed as above within each session are statistically independent from each other.
 173 The final $\overline{\Delta_{47}}$ value for a given unknown sample may thus be simply computed as the weighted
 174 average of $\overline{\Delta_{47}}$ from different sessions. Using a weighted average for this last step is necessary to
 175 account for inter-session differences in the number of analyses of that sample, and also potentially in
 176 raw analytical repeatability (unless, for instance, a conscious choice is made to use a single estimate
 177 of σ_{47}^{raw} constrained by all sessions):

$$\Delta_{47}^{\text{final}} = \sum_i \omega_i (\overline{\Delta_{47}})_i \quad \text{with sessions noted as } i \quad (14)$$

$$\omega_i = \sigma(\overline{\Delta_{47}})_i^{-2} / \sum_i \sigma(\overline{\Delta_{47}})_i^{-2} \quad (15)$$

$$\sigma(\Delta_{47}^{\text{final}})^2 = \sum_i \omega_i^2 \sigma(\overline{\Delta_{47}})_i^2 = 1 / \sum_i \sigma(\overline{\Delta_{47}})_i^{-2} \quad (16)$$

178 **3 Discussion**

179 **3.1 Properties of standardization errors**

180 The standardization model of section 2.4 is mathematically equivalent to the least-squares fitting of
 181 a two-dimensional plane described by eq. (3) in a three-dimensional space (δ_{47} , Δ_{47}^{raw} , Δ_{47}). Most
 182 properties described below arise naturally from this geometry.

183 Standardization uncertainties depend greatly on the bulk (δ_{47}) and clumped-isotope (Δ_{47}) com-
 184 position of the unknown samples to be standardized. It is thus useful to describe this uncertainty in
 185 terms of an “error field” which can be mapped in (δ_{47} , Δ_{47}) space, as shown in fig. 1. The minimum
 186 standardization error coincides, in (δ_{47} , Δ_{47}) space, with the barycenter of the anchor analyses, and
 187 its value is equal to $\sigma_{47}/N^{1/2}$, with N being the total number of anchor analyses.

188 Outside of a polygon defined by the anchor samples, standardization errors increase steeply. As
 189 illustrated in fig. 1, this increase is comparatively slower if analyses are evenly distributed between
 190 anchor samples, which tightens constraints on parameters *a* and *b*.

191 Fig. 1 also illustrates the benefits of using anchors with extreme isotopic compositions, which
 192 increases the area of the anchor polygon. One potential drawback of relying on isotopically extreme
 193 anchors, however, is that our “planar” model approximation might then break down. For instance,
 194 a small quadratic component to the compositional nonlinearity (term *b* δ_{47} in eq. 3), whose effect
 195 would be negligible over a δ_{47} range of 30 ‰, might introduce a significant bias over a range of 60 or
 196 100 ‰ (e.g., fig. 7 from *He et al.* [2012]).

197 The properties outlined above are fully consistent with the Monte Carlo simulations of *Kocken*
 198 *et al.* [2019]. In particular, they explain all of the main patterns displayed in their fig. 5. The primary
 199 difference between our approach and that of *Kocken et al.*, beyond the difference in mathematical
 200 methods, is that their simulations focus on the empirical transfer function (eq. 2), which corresponds

201 to parameters a and c . Here we show that the uncertainty from compositional non-linearities (eq. 1)
 202 behaves in a similar way, and that all of these corrections can be propagated explicitly in a unified
 203 manner, side-stepping the need for Monte Carlo simulations.

204 3.2 Impact of standardization errors

205 Because the relative contributions of the autogenic and allogenic error components defined above
 206 (eqs. 11–13) are sensitive to the distribution of analyses among anchor and unknown samples and
 207 on the isotopic composition of unknowns relative to the anchor polygon, they are expected to vary
 208 greatly between laboratories and/or sessions. The InterCarb dataset, comprising over five thousand
 209 analyses from 22 different laboratories [Bernasconi *et al.*, 2021], offers an excellent opportunity to
 210 quantify these two components in a wide range of realistic settings.

211 A compilation of σ_s versus σ_u for the average Δ_{47} value of unknown samples obtained in each
 212 of the 77 InterCarb sessions is shown in figure 2. As expected, standardization errors for IAEA-
 213 C1, a marble sample which plots within the anchor polygon defined in $(\delta_{47}, \Delta_{47})$ space by ETH-
 214 1/2/3, are generally slightly smaller than autogenic errors, resulting in a modest increase of the total
 215 Δ_{47} error (σ_{47}) relative to the autogenic error. Samples IAEA-C2 (natural travertine) and ETH-4
 216 (synthetic calcite), both of them located outside of the anchor polygon, display larger standardization
 217 errors, thereby increasing σ_{47} by an average factor of 1.5 and up to a factor of 2. Finally, in the
 218 case of the MERCK sample, a synthetic carbonate with extremely depleted $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values,
 219 standardization errors generally dominate. As a result, propagating them into the total Δ_{47} error
 220 increases σ_{47} by an average factor of 2.5 and up to a factor of 4.

221 3.3 Correlations between samples

222 Compared to most other isotopic or elemental tracers, carbonate clumped isotopes stand out in that
 223 analytical uncertainties remain large relative to the range of Δ_{47} values typical of natural samples.
 224 Standardization errors contribute a sizable portion of these uncertainties, but it is notable that they
 225 do so in a way that is strongly correlated between samples, as illustrated by the joint 95 % confidence
 226 ellipses for the average Δ_{47} values of unknown samples shown in figure 3.

227 Appendix B provides full computational details for the covariance between the session-averaged
 228 Δ_{47} values of two unknowns samples (B.1); the uncertainties characterizing Δ_{47} differences between
 229 samples (B.2); and weighted mean Δ_{47} values averaged over several samples (B.3). The key point to
 230 keep in mind is that full analytical errors are not independent between samples of the same session,
 231 with the following consequences: (1) when averaging many Δ_{47} measurements within a single ses-
 232 sion, analytical errors will not tend to zero but to the standardization error for this sample; (2) the
 233 error on Δ_{47} differences between samples of similar compositions (as is often the case in paleocli-
 234 mate records) is largely unaffected by standardization errors, but only if they are analyzed within
 235 the same session.

236 **3.4 Gaussian approximation of standardization errors**

237 The error propagation formulas presented in section 2.6 are first-order Taylor approximations. Be-
 238 cause Δ_{47} is not a linear combination of $(\Delta_{47}^{\text{raw}}, \delta_{47}, a, b, c)$, propagated errors in Δ_{47} are not strictly
 239 Gaussian. However, after quantifying the non-Gaussian effects of these approximations using Monte
 240 Carlo simulations of the full InterCarb dataset (Appendix C), we find that these deviations from nor-
 241 mality remain entirely negligible in practice.

242 As an extreme example, figure 4 shows the Monte Carlo distribution of standardization errors
 243 for the average Δ_{47} value of IAEA-C2 in one of the InterCarb sessions, chosen because it is the “least
 244 Gaussian” distribution of the whole dataset, i.e. the least likely to be Gaussian ($p = 2 \times 10^{-7}$) based on
 245 a Kolmogorov-Smirnov test. Even in this worst-case example, differences between the Monte Carlo
 246 cumulative distribution function (CDF) and the Gaussian CDF computed from eqs. (11–13) remain
 247 minuscule: the Monte Carlo average of Δ_{47} for this particular sample is 0.6659 ‰ (versus 0.6640 ‰
 248 for the propagated value), and the width of the 95 % spread of values in the Monte Carlo dataset is
 249 ± 0.0494 ‰ (versus ± 0.0491 ‰ for the Gaussian approximation).

250 **3.5 Statistical weighting options**

251 For the sake of simplicity, the error model described above rests on simple assumptions, for example
 252 by assigning equal statistical weights to all analyses. In the following sections we briefly discuss vari-
 253 ous ways in which this error model could be modified to better reflect real-life analytical conditions.

254 **3.5.1 Equal session weights**

255 In the general case where all sessions are considered equal, we recommend that each session should
 256 first be standardized using eqs. (21) and (23). The *overall* Δ_{47}^{raw} repeatability should then be computed
 257 using a slightly modified version of eq. (6), with N_A being the number of anchor samples, N_U the
 258 number of unknown samples, and N the total number of analyses:

$$(\sigma_{47}^{\text{raw}})^2 = \frac{1}{N - N_A - N_U} \sum \left(\Delta_{47}^{\text{raw}} - a \overline{\Delta_{47}} - b \delta_{47} - c \right)^2 \quad (17)$$

259 This overall repeatability should then be used to scale the covariance matrix of each session
 260 according to eq. (7).

261 **3.5.2 Different session weights**

262 It may be justified in some cases to assign different statistical weights to analyses from different
 263 sessions. We would not generally recommend doing so based only on observed differences in σ_{47}
 264 (which will inevitably vary slightly between sessions), unless these differences are statistically sig-
 265 nificant with a high level of confidence. On the other hand, data produced under different analytical
 266 conditions may in some cases reasonably be expected to be more or less precise: for example, mea-
 267 surements obtained using greater ion currents should be more precise due to counting statistics
 268 alone.

269 In such cases, we may first divide sessions into groups expected to share similar analytical preci-
 270 sion levels. Pooled Δ_{47}^{raw} repeatabilities for each group may then be computed according to eq. (17),
 271 and subsequently applied to covariance matrix scaling according to eq. (7).

272 3.6 Pooled standardization model taking unknown samples into account

273 3.6.1 Principle

274 By only considering anchor samples to constrain the standardization parameters (a , b , c) of each
 275 session, the models described so far neglect some useful information. As a matter of fact, even
 276 without prior knowledge of the Δ_{47} values of unknown samples, we expect the relative mapping
 277 of anchors and unknowns in $(\delta_{47}, \Delta_{47})$ space to be preserved between sessions. This approach is
 278 only useful when some of the sessions have unknown samples in common, but in that case it is
 279 likely to substantially increase the number of observations constraining the standardization model,
 280 making it more robust (less sensitive to outliers in the anchor analyses) and more precise (by virtue
 281 of increasing the model’s degrees of freedom).

282 In practice, instead of treating each session as a separate least-squares problem, we now aim
 283 to minimize an “pooled” version of the χ^2 defined in eq. (4), this time summed over all analyses
 284 (including both anchors and unknowns) in all sessions considered:

$$\chi^2 = \sum (\Delta_{47}^{\text{raw}} - a_i \Delta_{47} - b_i \delta_{47} - c_i)^2 \quad (18)$$

285 where Δ_{47}^{raw} and δ_{47} are the observations from each analysis, (a_i, b_i, c_i) are the standardization
 286 parameters for session (i), and Δ_{47} is either a nominal value assigned *a priori* (for anchor analyses)
 287 or an additional, free model parameter equal to the Δ_{47} value of the relevant unknown sample. The
 288 pooled regression model now rests on a number of observations equal to the total number of analyses,
 289 with a number of model parameters equal to the number of unknown samples plus three times the
 290 number of sessions.

291 Because some of the χ^2 terms include the product of two model parameters, this is not a linear
 292 least-squares problem and the direct solution of Appendix A no longer applies. One may, however,
 293 call upon well-established numerical approaches designed to optimize non-linear problems. In our
 294 experience, the classical Levenberg-Marquardt method [Levenberg, 1944; Marquardt, 1963], as im-
 295 plemented by the [LMFIT](#) Python package [Newville *et al.*, 2014], is well suited to this task. Even for
 296 large datasets of several thousand analyses, it is able to quickly and reliably output a vector of best-
 297 fit values for all model parameters (including Δ_{47} values for all unknown samples) along with the
 298 corresponding covariance matrix, thus directly providing standard errors and covariances between
 299 unknown sample Δ_{47} values.

3.6.2 Benefits

The benefits of a pooled standardization model may not be immediately obvious, but this approach should yield systematic improvements in the robustness and accuracy of the standardization procedure. For instance, considering the samples shown in figure 1, it may be clear that forcing the Δ_{47} value of MERCK to remain consistent between sessions should greatly contribute to constrain variations in the compositional slope (b) between sessions, even without knowing MERCK's true composition. The same argument could be made if one were to analyze heated and equilibrated gases along with carbonate standards, treating them as entirely unknown samples: even without any knowledge of CO_2 equilibrium values nor of acid fractionation effects, the large spread of Δ_{47} between heated and equilibrated gases would strongly constrain variations of the scrambling effect (a) between sessions, thereby reducing standardization errors for all samples.

Figure 5 illustrates this reduction in standardization errors by showing (δ_{47} , Δ_{47}) plots for the four sessions from Lab #12 in the InterCarb dataset, comparing the error fields resulting either from the pooled standardization approach (one model with 153 degrees of freedom) or from the earlier approach ignoring unknown samples (four independent models with 20, 16, 24, and 16 degrees of freedom, respectively). Although the locations and values of the error field minima remain largely unaffected by the choice of standardization method, in this case the pooled model strongly reduces standardization errors for analyses plotting outside of the anchor polygon (from 10–11 ppm down to 6 ppm for MERCK), despite the fact that no assumption was made regarding the true Δ_{47} values of unknown samples.

3.6.3 Caveat

The pooled approach depends critically on our earlier assumption that samples are homogeneous, which we acknowledge to be generally but not universally true. It is however simple enough, in the presence of samples suspected to be heterogeneous (i.e. whose Δ_{47} repeatability is demonstrably worse than for carbonate standards with a statistically high level of confidence), to treat each of the corresponding analyses as belonging to separate samples.

3.7 ClumpyCrunch and D47crunch

The calculations discussed above may be tedious to implement from scratch. The simplest way to take advantage of these error models is to use the latest version of the open-source ClumpyCrunch web application (<https://clumpycrunch.pythonanywhere.com>), which implements both the independent-sessions method of section 2.6 and the pooled standardization approach of section 3.6. Those wishing to experiment at a deeper lever may install the underlying, open-source D47crunch library for Python (<https://doi.org/10.5281/zenodo.4314550>), which also supports computing different repeatabilities for different groups of sessions (section 3.5.2); explicitly treating some samples as potentially inhomogeneous (section 3.6); modeling temporal drifts in parameters a , b , c (appendix A.2); computing standard errors for Δ_{47} differences and/or means accounting for analytical covariance between samples (appendix B.2-B.3); and assessing whether the Δ_{47} repeatabilities of two samples differ significantly. Both D47crunch and ClumpyCrunch also output robust 95 % confidence limits for final Δ_{47} values based on the number of degrees of freedom in the standardization models.

4 Conclusion

The framework presented here provides a quantitative/predictive description of Δ_{47} error propagation, fully taking into account standardization errors and their properties. It corroborates and extends earlier investigations based on Monte Carlo simulations [Kocken *et al.*, 2019]. This mathematical formulation is found not to introduce large deviations from normality: in other words, if Δ_{47}^{raw} errors are Gaussian, the fully propagated Δ_{47} errors may also be treated as Gaussian for all practical purposes. What's more, as reported by Bernasconi *et al.* [2021], using this framework yields a very reasonable ($p = 0.19$) prediction for the distribution of inter-laboratory scatter in Δ_{47} values within the InterCarb dataset.

Based on this framework, we describe a new, “pooled” standardization method designed to make full use of the constraints available from both anchor and unknown analyses. This approach is expected to yield substantially improved standardization models, both in terms of precision and robustness. We also provide new online resources and a Python library aiming to make the use of such error models as simple as possible. This library being open-source and fully documented, implementing the methods described here in existing software such as Easotope [John & Bowen, 2016] should be straightforward.

Most published clumped-isotope studies so far have lacked a rigorous propagation of standardization errors. This, of course, is not a problem in itself, but the InterCarb results unambiguously demonstrate that these standardization uncertainties are both necessary and sufficient to explain the inter-laboratory scatter observed in this large dataset ($N = 5329$). Going one step further, it could be argued that the ongoing persistence of inter-laboratory discrepancies in Δ_{47} calibrations [Petersen *et al.*, 2019] is due, at least in part, to largely ignored standardization errors. Whatever the case, it seems likely that future comparisons between results obtained in different laboratories would greatly benefit from more accurate error estimates.

363 Finally, although all statistical models are interpretative approximations, their ultimate value
 364 depends less on their exactness than on their practical usefulness. At the very least, the framework
 365 described here may help improve standardization protocols (e.g., anchor/unknown ratios, composi-
 366 tional distribution of anchors) and inform our collective effort to establish inter-laboratory reference
 367 materials.

368 Data and code

369 The complete raw data and all associated code used in this work are available under a Modified BSD
 370 License at <https://doi.org/10.5281/zenodo.4314593>.

371 Notations

372 **a** : scrambling factor, one of the standardization parameters, quantifying the amount of molecular
 373 recombination during the analytical procedure; its value should lie between 0 and 1.

374 **b** : compositional slope, one of the standardization parameters, quantifying small systematic errors
 375 in the electrical background of the ion collectors; it may be positive or negative and its absolute
 376 value should ideally remain small (10^{-2} or less).

377 **c** : working gas offset, one of the standardization parameters, accounting for the fact that the work-
 378 ing gas is not necessarily stochastic; in settings where the working gas is equilibrated at room
 379 temperature, $c \approx -a$

380 **N_a** : Number of analyses.

381 **N_f** : Degrees of freedom in a regression model.

382 **N_A** : Number of anchor samples.

383 **N_U** : Number of unknown samples.

384 **Δ_{47}** : delta notation (in ‰) for the clumped-isotope anomaly associated with mass-47 CO_2 ; either
 385 denotes the “true” value for a given sample, or the “absolute” value computed from one or
 386 more IRMS measurements after standardization.

387 **Δ_{47}^{raw}** : “raw” Δ_{47} value from an IRMS measurement, before standardization.

388 **δ_{47}** : delta notation (in ‰) for the mass-47 to mass-44 abundance ratio of an analyte CO_2 , generally
 389 defined relative to a working reference gas.

390 **σ_{47}** : analytical error/uncertainty assigned to individual measurements of Δ_{47} (eq. 5).

391 **σ_{47}^{raw}** : analytical error/uncertainty assigned to individual measurements of Δ_{47}^{raw} (eq. 6).

392 **σ_s** : allogenic error, i.e. the analytical error/uncertainty on a Δ_{47} measurement arising from the
 393 standardization function (eq. 13).

394 **σ_u** : autogenic error, i.e. the analytical error/uncertainty on a Δ_{47} measurement arising from the
 395 analyses of the unknown sample itself (eq. 12).

396 **Appendix A: Least squares regression**

397 **A.1 General linear case**

398 Consider a linear model f defined as:

$$y = f(x, a_1, a_2 \dots a_p) = \sum_{i=1}^p a_i f_i(x) \tag{19}$$

399 where x is a scalar or vectorial explanatory variable; y the response variable; $(f_1 \dots f_p)$ a series of functions
400 of x ; and $(a_1 \dots a_p)$ a series of scalar factors which are the model parameters to be estimated.

401 Given n observations $((x_1, y_1) \dots (x_n, y_n))$ to fit, we construct the following matrices:

$$A = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \dots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_p(x_2) \\ \vdots & \vdots & & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_p(x_n) \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{20}$$

402 The best-fit parameters $(a_1 \dots a_n)$ and their unscaled variance-covariance matrix V_0 are then:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = V_0 \times A^T \times Y \quad V_0 = (A^T \times A)^{-1} \tag{21}$$

403 **A.2 Application to the standardization model**

404 The standardization model of eq. (3) is equivalent to the above formulation if:

$$\begin{aligned} y &= \Delta_{47}^{\text{raw}} & f_1(x) &= \Delta_{47} \\ x &= (\delta_{47}, \Delta_{47}) & f_2(x) &= \delta_{47} \\ (a_1, a_2, a_3) &= (a, b, c) & f_3(x) &= 1 \end{aligned} \tag{22}$$

405 In this case:

$$A = \begin{bmatrix} \Delta_{47} & \delta_{47} & 1 \\ \Delta_{47} & \delta_{47} & 1 \\ \vdots & \vdots & \vdots \\ \Delta_{47} & \delta_{47} & 1 \end{bmatrix} \quad \begin{array}{l} \leftarrow \text{analysis \#1} \rightarrow \\ \leftarrow \text{analysis \#2} \rightarrow \\ \leftarrow \text{analysis \#n} \rightarrow \end{array} \quad Y = \begin{bmatrix} \Delta_{47}^{\text{raw}} \\ \Delta_{47}^{\text{raw}} \\ \vdots \\ \Delta_{47}^{\text{raw}} \end{bmatrix} \tag{23}$$

406 To take into account an uncertainty, noted σ , assigned to the observations, A and Y should both be divided
407 by σ , which will leave the best-fit parameters unchanged and scale the variance-covariance matrix V_0 by a
408 factor of σ^2 (as in eq. 7).

409 Alternatively, to assign individual uncertainties, noted $(\sigma_1 \dots \sigma_n)$ to the n analyses, each line of A and each
410 element of Y should be divided by the corresponding σ value:

$$A = \begin{bmatrix} \Delta_{47}/\sigma_1 & \delta_{47}/\sigma_1 & 1/\sigma_1 \\ \Delta_{47}/\sigma_2 & \delta_{47}/\sigma_2 & 1/\sigma_2 \\ \vdots & \vdots & \vdots \\ \Delta_{47}/\sigma_n & \delta_{47}/\sigma_n & 1/\sigma_n \end{bmatrix} \quad \begin{array}{l} \leftarrow \text{analysis \#1} \rightarrow \\ \leftarrow \text{analysis \#2} \rightarrow \\ \leftarrow \text{analysis \#n} \rightarrow \end{array} \quad Y = \begin{bmatrix} \Delta_{47}^{\text{raw}}/\sigma_1 \\ \Delta_{47}^{\text{raw}}/\sigma_2 \\ \vdots \\ \Delta_{47}^{\text{raw}}/\sigma_n \end{bmatrix} \tag{24}$$

411 Extending this model with additional parameters should be rather straightforward. For instance, in order
 412 to account for a temporal drift in the compositional non-linearity, one could reformulate the model, with t
 413 denoting time and an additional standardization parameter d , as:

$$\Delta_{47}^{\text{raw}} = a \Delta_{47} + (b + td) \delta_{47} + c \quad (25)$$

414 Which would correspond to:

$$A = \begin{bmatrix} \Delta_{47} & \delta_{47} & 1 & t \delta_{47} \\ \Delta_{47} & \delta_{47} & 1 & t \delta_{47} \\ \vdots & \vdots & \vdots & \vdots \\ \Delta_{47} & \delta_{47} & 1 & t \delta_{47} \end{bmatrix} \begin{matrix} \leftarrow \text{analysis \#1} \rightarrow \\ \leftarrow \text{analysis \#2} \rightarrow \\ \\ \leftarrow \text{analysis \#n} \rightarrow \end{matrix} \begin{bmatrix} \Delta_{47}^{\text{raw}} \\ \Delta_{47}^{\text{raw}} \\ \vdots \\ \Delta_{47}^{\text{raw}} \end{bmatrix} = Y \quad (26)$$

415 **Appendix B: Δ_{47} covariance**

416 **B.1 Covariance between unknown samples**

417 Consider two unknown samples A and B, whose session-averages compositions ($\overline{\delta_{47}}$, $\overline{\Delta_{47}^{\text{raw}}}$, and $\overline{\Delta_{47}}$) are re-
 418 spectively noted δ_A , δ_B , Δ_A^{raw} , Δ_B^{raw} , Δ_A , and Δ_B . Defining X as the column vector $[\Delta_A, \Delta_B]$, we can express its
 419 Jacobian J_X relative to the system of variables $(\Delta_A^{\text{raw}}, \Delta_B^{\text{raw}}, a, b, c)$ and the covariance C of this quintuplet as:

$$J_X = \frac{1}{a} \begin{bmatrix} 1 & 0 & -\Delta_A & -\delta_A & -1 \\ 0 & 1 & -\Delta_B & -\delta_B & -1 \end{bmatrix} \quad C = \begin{bmatrix} (\sigma_{47}^{\text{raw}})^2/N_A & 0 & 0 & 0 & 0 \\ 0 & (\sigma_{47}^{\text{raw}})^2/N_B & 0 & 0 & 0 \\ 0 & 0 & \sigma_a^2 & c_{ab} & c_{ac} \\ 0 & 0 & c_{ab} & \sigma_b^2 & c_{bc} \\ 0 & 0 & c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} \quad (27)$$

420 The covariance matrix of X is then:

$$C_X = J_X \times C \times J_X^T \quad (28)$$

421 Because of the structure of J and C , the non-zero terms of C_X are equal to:

$$\text{cov}(\Delta_A, \Delta_B) = \frac{1}{a^2} \begin{bmatrix} \Delta_A & \delta_A & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_a^2 & c_{ab} & c_{ac} \\ c_{ab} & \sigma_b^2 & c_{bc} \\ c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} \times \begin{bmatrix} \Delta_B \\ \delta_B \\ 1 \end{bmatrix} \quad (29)$$

$$\text{cov}(\Delta_A, \Delta_B) = \frac{1}{a^2} (\Delta_A \Delta_B \sigma_a^2 + \delta_A \delta_B \sigma_b^2 + \sigma_c^2 + (\Delta_A \delta_B + \delta_A \Delta_B) c_{ab} + (\Delta_A + \Delta_B) c_{ac} + (\delta_A + \delta_B) c_{bc}) \quad (30)$$

422 The covariance between mean Δ_{47} values of two samples averaged over several sessions is zero if the
 423 samples were never analyzed in the same session. Otherwise, with ω_{Ai} and ω_{Bi} weights defined as in (15):

$$\Delta_A^{\text{final}} = \sum_i \omega_{Ai} \Delta_{Ai} \quad \text{with } i \text{ denoting all sessions including A} \quad (31)$$

$$\Delta_B^{\text{final}} = \sum_j \omega_{Bj} \Delta_{Bj} \quad \text{with } j \text{ denoting all sessions including B} \quad (32)$$

$$\text{cov}(\Delta_A^{\text{final}}, \Delta_B^{\text{final}}) = \sum_k \omega_{Ak} \omega_{Bk} \text{cov}(\Delta_{Ak}, \Delta_{Bk}) \quad \text{with } k \text{ denoting all sessions including both A and B} \quad (33)$$

424 **B.2 Standard errors on Δ_{47} differences between samples**

425 Consider two unknown samples A and B, whose session-averages compositions ($\overline{\delta_{47}}$, $\overline{\Delta_{47}^{raw}}$, and $\overline{\Delta_{47}}$) are re-
 426 spectively noted δ_A , δ_B , Δ_A^{raw} , Δ_B^{raw} , Δ_A , and Δ_B . Defining x as the difference ($\Delta_A - \Delta_B$), we can express its
 427 Jacobian J_x relative to the system of variables (Δ_A^{raw} , Δ_B^{raw} , a , b , c) as:

$$J_x = \frac{1}{a} [1, -1, \Delta_B - \Delta_A, \delta_B - \delta_A, 0] \quad (34)$$

428 and compute the variance of x using the same covariance matrix C as above:

$$\sigma_x^2 = J_x \times C \times J_x^T \quad (35)$$

$$\sigma_x^2 = \sigma_{47}^2 \left(\frac{1}{N_A} + \frac{1}{N_B} \right) + \frac{(\Delta_B - \Delta_A)^2 \sigma_a^2 + (\delta_B - \delta_A)^2 \sigma_b^2 + 2(\Delta_B - \Delta_A)(\delta_B - \delta_A)c_{ab}}{a^2} \quad (36)$$

429 **B.3 Standard errors on mean Δ_{47} values averaged over several samples**

430 As an example, we treat here the problem of a weighted average of three samples. Consider three unknown
 431 samples A, B, and C, whose session-averages compositions ($\overline{\delta_{47}}$, $\overline{\Delta_{47}^{raw}}$, and $\overline{\Delta_{47}}$) are respectively noted δ_A , δ_B ,
 432 δ_C , Δ_A^{raw} , Δ_B^{raw} , Δ_C^{raw} , Δ_A , Δ_B , and Δ_C . Defining W as the weighted average ($x_A \Delta_A + x_B \Delta_B + x_C \Delta_C$) and w as the
 433 weighted average ($x_A \delta_A + x_B \delta_B + x_C \delta_C$), we can express the Jacobian of W relative to the system of variables
 434 (Δ_A^{raw} , Δ_B^{raw} , Δ_C^{raw} , a , b , c) as:

$$J_W = \frac{1}{a} [x_A \quad x_B \quad x_C \quad -W \quad -w \quad -1] \quad (37)$$

435 and compute the variance of W using the same method as above:

$$\sigma_W^2 = J_W \times \begin{bmatrix} (\sigma_{47}^{raw})^2/N_A & 0 & 0 & 0 & 0 & 0 \\ 0 & (\sigma_{47}^{raw})^2/N_B & 0 & 0 & 0 & 0 \\ 0 & 0 & (\sigma_{47}^{raw})^2/N_C & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_a^2 & c_{ab} & c_{ac} \\ 0 & 0 & 0 & c_{ab} & \sigma_b^2 & c_{bc} \\ 0 & 0 & 0 & c_{ac} & c_{bc} & \sigma_c^2 \end{bmatrix} \times J_W^T \quad (38)$$

$$\sigma_W^2 = \sigma_{47}^2 \left(\frac{x_A^2}{N_A} + \frac{x_B^2}{N_B} + \frac{x_C^2}{N_C} \right) + \frac{W^2 \sigma_a^2 + w^2 \sigma_b^2 + \sigma_c^2 + 2(wWc_{ab} + Wc_{ac} + wc_{bc})}{a^2} \quad (39)$$

436 Note that the second term above is equal to the value of the standardization error field at the weighted
 437 barycenter of the samples in (δ_{47} , Δ_{47}) space.

438 **Appendix C: Monte Carlo assessment of the normality of Δ_{47} errors**

439 Because Δ_{47} is not a linear function of $(\Delta_{47}^{\text{raw}}, \delta_{47}, a, b, c)$, the propagation of standardization errors described in
 440 section 2.6 is an approximation. Here we used a Monte Carlo simulation based on the full InterCarb dataset
 441 to investigate how much autogenic and allogenic errors deviate from a Gaussian approximation. In each
 442 step of the simulation, we offset the original Δ_{47}^{raw} values observed in each of the 5329 analyses by random,
 443 independent, zero-centered Gaussian errors with a standard deviation equal to the session's σ_{47}^{raw} value. We
 444 then standardize all sessions of the modified dataset and record the final, session-averaged Δ_{47} values of each
 445 unknown sample ($N = 226$) for a total of 10^4 iterations. Each of these session averages is submitted to a
 446 Kolmogorov-Smirnov (KS) test of normality [Massey, 1951], comparing the distribution of these 10^4 values
 447 to a normal distribution centered on the original session-averaged value and whose width depends of the
 448 original propagated errors. Each of the 226 KS tests yields a p-value corresponding to the null hypothesis that
 449 the two distributions are identical. By design, if the Gaussian approximation of the propagated errors holds
 450 true, these p-values should be evenly distributed in the [0–1] interval. We may quantify how well they do so
 451 by performing a new KS test comparing the distribution of p-values to the uniform distribution, yielding a
 452 new, final p-value for the hypothesis that the errors in the InterCarb dataset follow Gaussian distributions.

453 We run this simulation in three different configurations, considering only autogenic errors, only standard-
 454 ization errors, or both. Initially, the random errors introduced in each iteration are scaled according to the
 455 Δ_{47} repeatability of each session (fig. 6A). We then repeat the simulations twice, by scaling the random errors
 456 according to a constant Δ_{47} repeatability of 50 ppm and 5 ppm, respectively (figs. 6B, 6C).

457 Predictably, based on eqs. (12-13), we find that autogenic errors behave in a Gaussian manner ($p = 0.81$),
 458 but this is clearly not the case for standardization errors ($p = 1 \times 10^{-28}$). Because the error propagation formula
 459 of eq. (8) is equivalent to a first-order Taylor expansion, the non-normality of standardization errors is expected
 460 worsen as Δ_{47} repeatability increases, as is the case in fig. 6B, and to become negligible when Δ_{47} repeatability
 461 is small enough (fig. 6C).

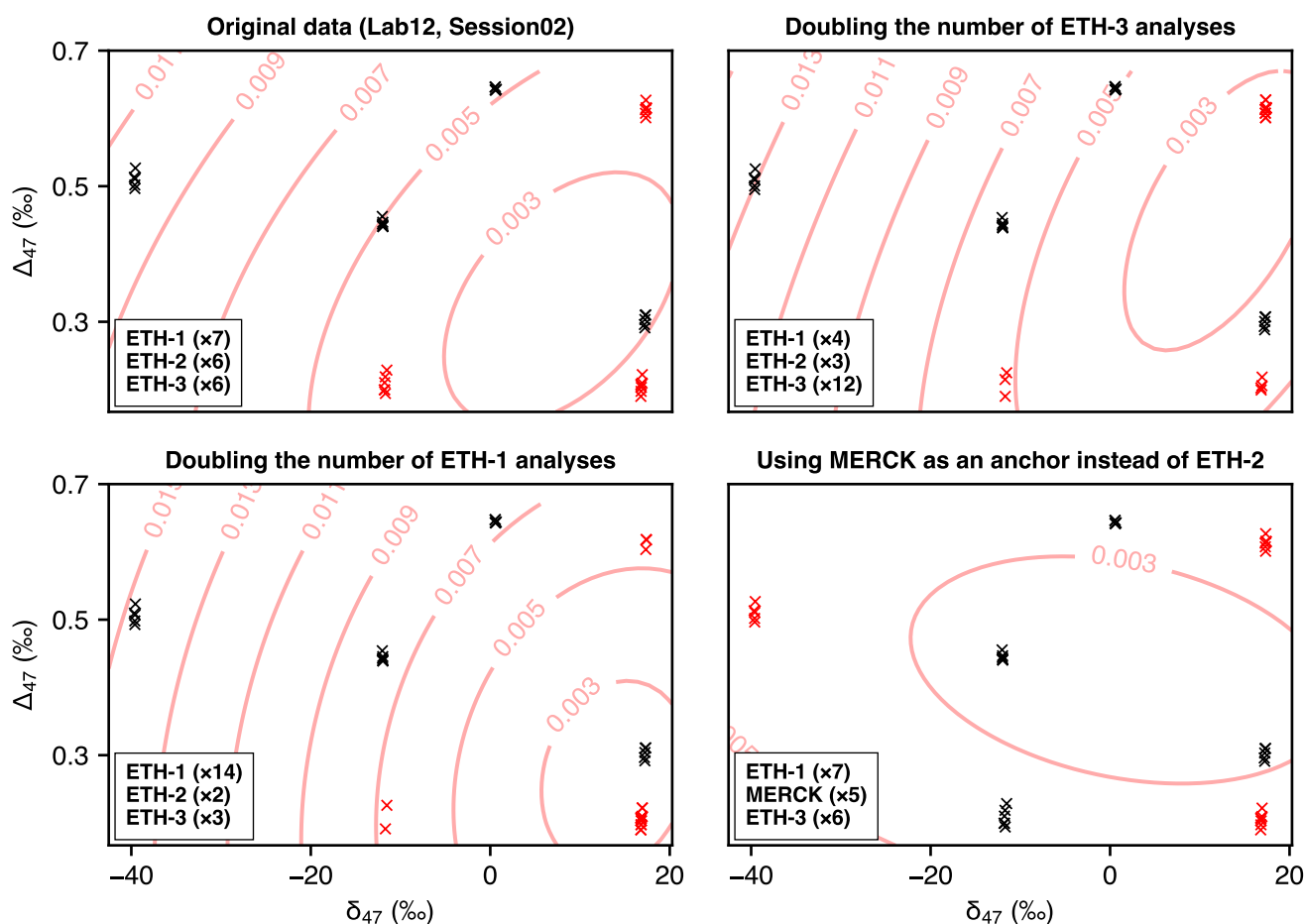


Figure 1: Properties of standardization errors. Upper left panel shows the unknown and anchor analyses (black and red crosses, respectively) and contours of the standardization error field (red lines) for Session #2 of Lab #12 in the InterCarb dataset. Upper right and lower left panels modify the original data by changing the distribution of anchor analyses between ETH-1, ETH-2, and ETH-3, keeping the total number of anchor analyses constant, illustrating that the error minimum coincides in $(\delta_{47}, \Delta_{47})$ space with the barycenter of anchor analyses. The lower right panel corresponds to the original data but treats ETH-2 as an unknown and MERCK as an anchor, highlighting the benefits of using isotopically extreme anchors.

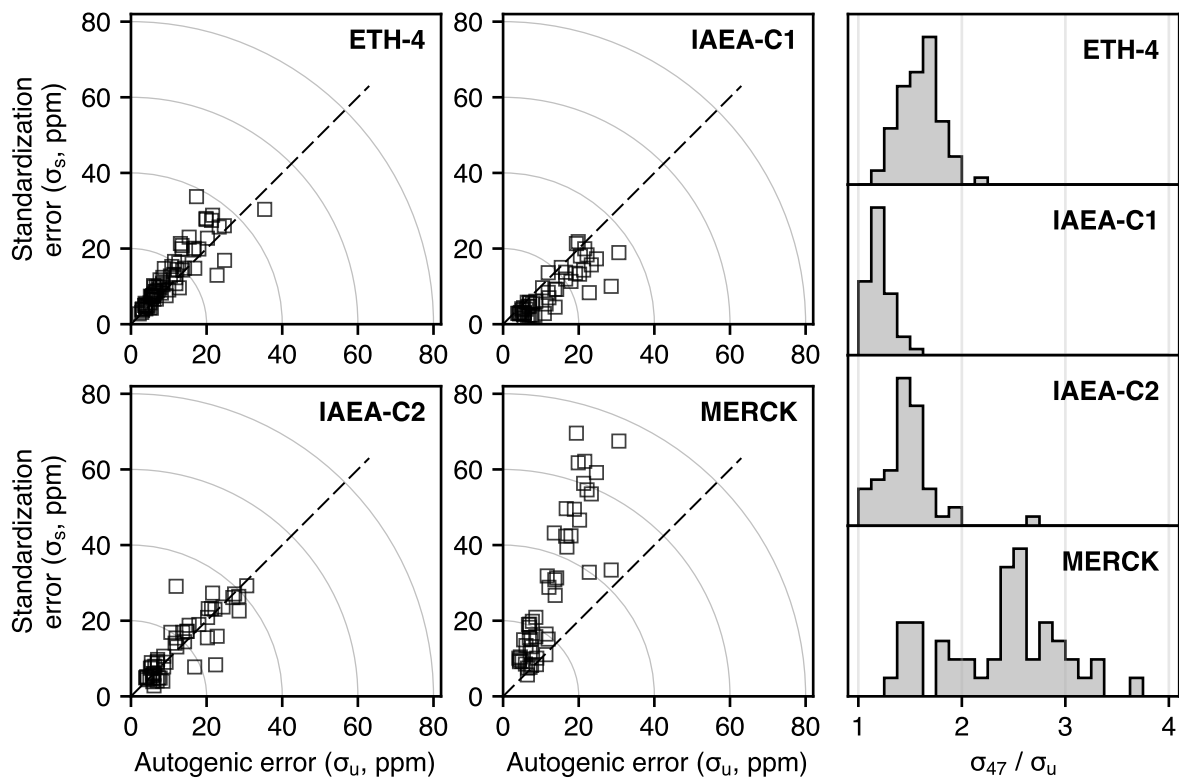


Figure 2: Autogenic versus standardization errors. Each square marker corresponds to the error components for the average Δ_{47} value of an unknown sample in each of the InterCarb sessions. Histograms characterize the ratios of total analytical error (σ_{47}) to autogenic errors (σ_u) for each of the unknown samples.

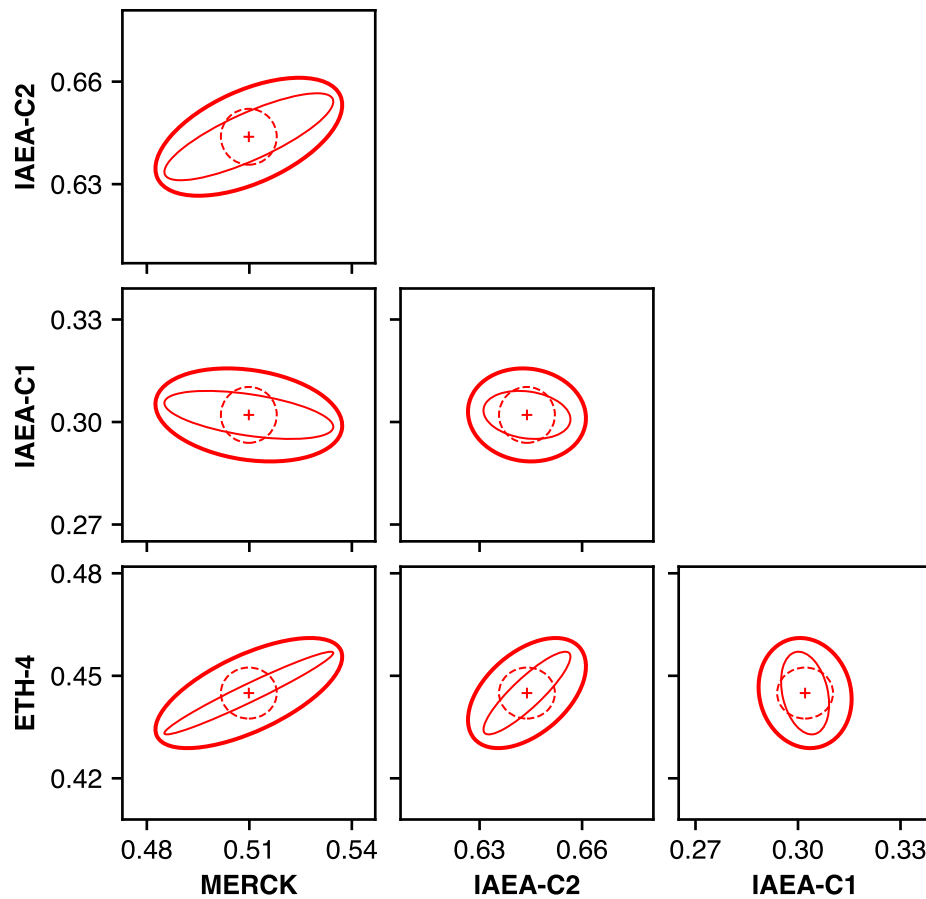


Figure 3: Covariance of errors in session-averaged Δ_{47} values for unknowns samples. Thick red lines correspond to joint 95 % confidence ellipses for the average Δ_{47} values of each unknown sample in Session02 of Lab12 (cf upper left panel of fig. 1). Thin red lines and dashed red lines correspond to joint 95 % confidence ellipses only taking into account standardization and autogenic errors, respectively.

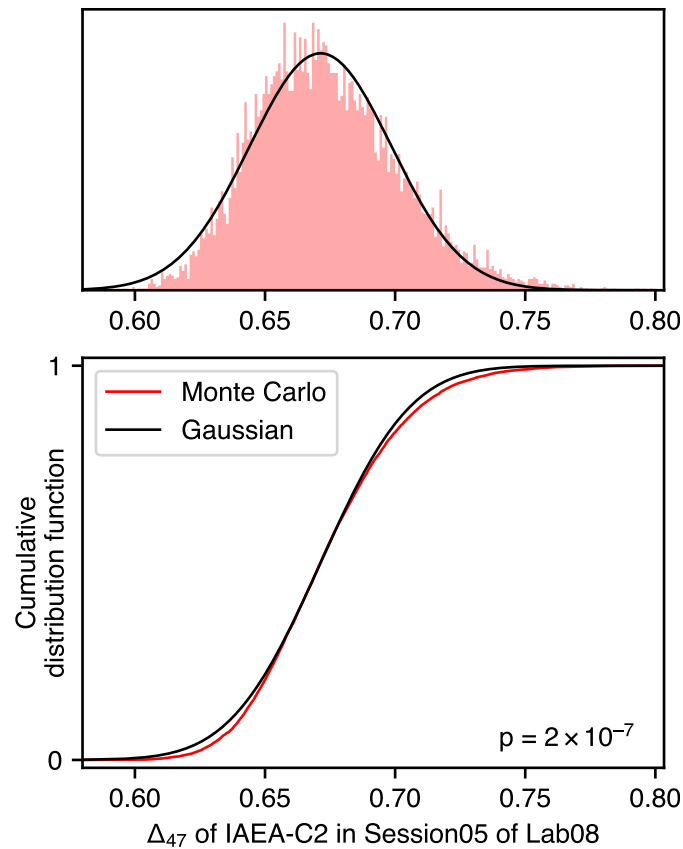


Figure 4: Error distribution for the “least Gaussian” average Δ_{47} value in the InterCarb dataset. Upper panel: red area corresponds to the Monte Carlo histogram of the average Δ_{47} value of IAEA-C2 in Session #7 of Lab #8, black line is the Gaussian probability distribution computed from eqs. (11-13). Lower panel: Monte Carlo (red line) and Gaussian (black line) cumulative distributions functions for this average Δ_{47} value. In the lower left corner, p is the Kolmogorov-Smirnov p -value for the null hypothesis that these two distributions are identical.

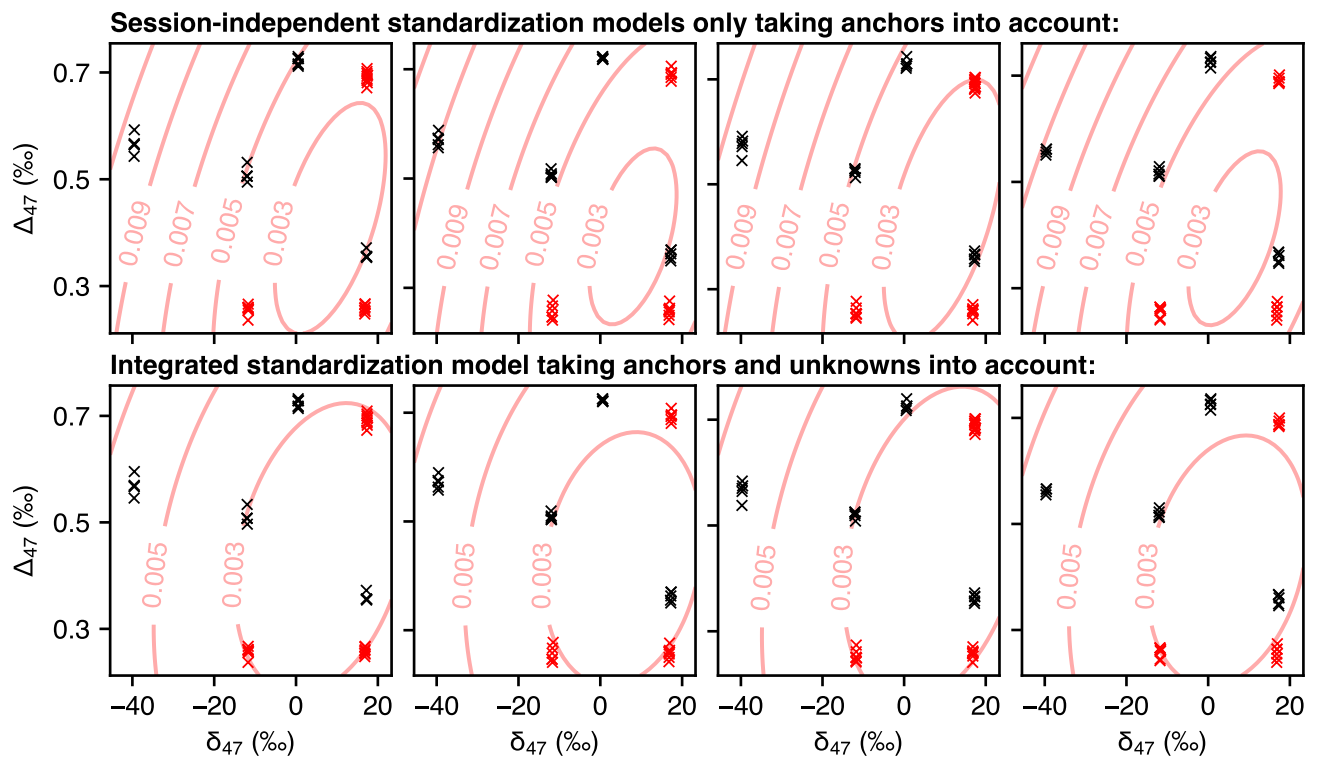


Figure 5: Benefits of a pooled standardization model. Unknown and anchor analyses (black and red crosses, respectively) and contours of the standardization error fields (red lines) for the four sessions of Lab #12 in the InterCarb dataset. Upper row: using four independent models only taking anchor analyses into account, with 20, 16, 24, and 16 degrees of freedom, respectively. Lower row: using a pooled standardization model with 153 degrees of freedom taking anchors and unknowns into account as described in section 3.6.

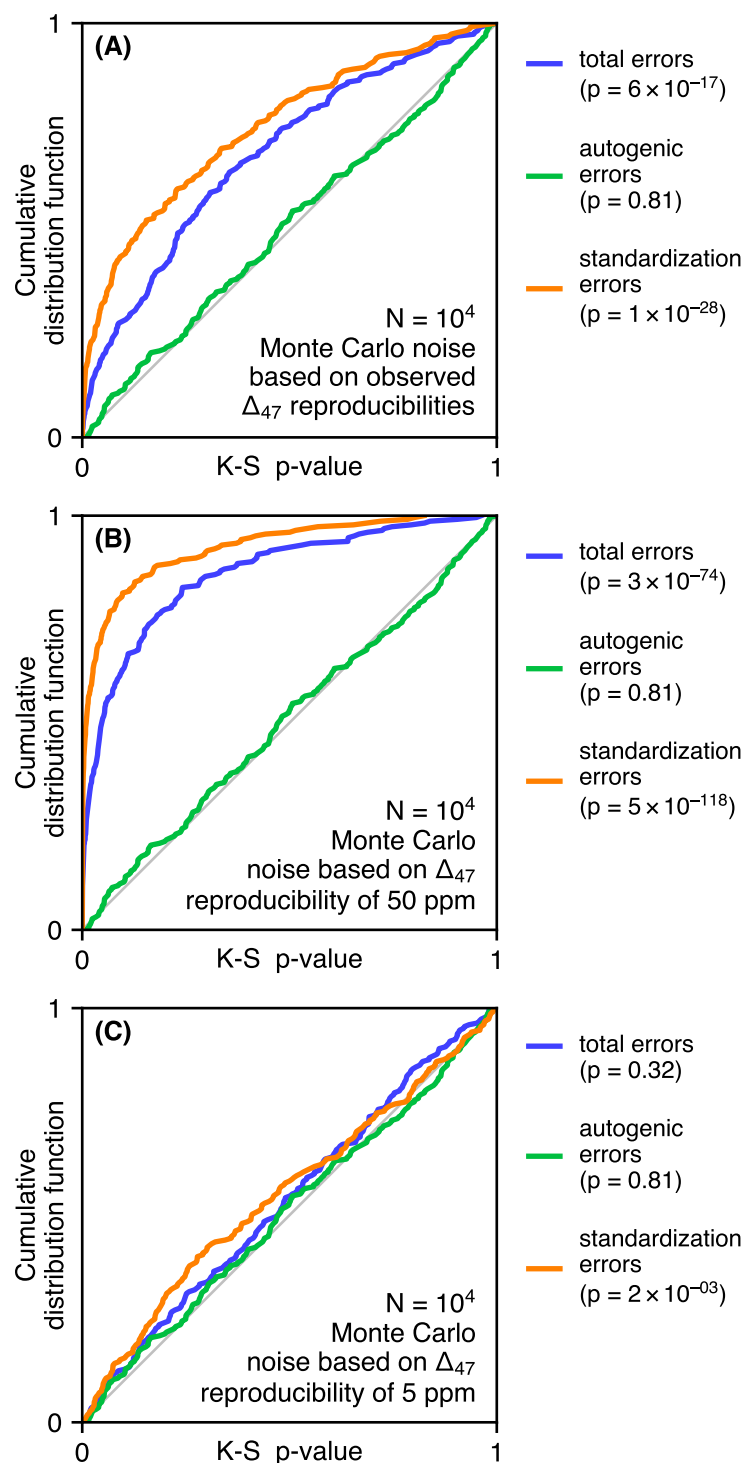


Figure 6: Monte Carlo simulation results All three panels display the cumulative distribution function of the p-values obtained from 10^4 Monte Carlo simulations of the full InterCarb dataset (see Appendix C for computational details). When the random offsets used by the simulation are scaled according to the original data (A), autogenic errors behave in a Gaussian manner, but the standardization (“allogenic”) errors do not, due to the limits of the first-order Taylor approximations used here for error propagation. As expected, greatly increasing (B) or decreasing (C) the random offsets used by the simulation results modulates the non-Gaussianity of standardization errors, while autogenic errors, despite being respectively increased or decreased (not shown here), remain Gaussian.

462 **References**

463 Bernasconi, S. M., Daëron, M., Bergmann, K. D., Bonifacie, M., Meckler, A. N., Affek, H. P., Anderson, N., Bajnai, D.,
 464 Barkan, E., Beverly, E., Blamart, D., Burgener, L., Calmels, D., Chaduteau, C., Clog, M., Davidheiser-Kroll, B., Davies,
 465 A., Dux, F., Eiler, J., Elliot, B., Fetrow, A. C., Fiebig, J., Goldberg, S., Hermoso, M., Huntington, K. W., Hyland, E.,
 466 Ingalls, M., Jaggi, M., John, C. M., Jost, A. B., Katz, S., Kelson, J., Kluge, T., Kocken, I. J., Laskar, A., Leutert, T. J.,
 467 Liang, D., Lucarelli, J., Mackey, T. J., Mangenot, X., Meinicke, N., Modestou, S. E., Müller, I. A., Murray, S., Neary,
 468 A., Packard, N., Passey, B. H., Pelletier, E., Petersen, S., Piasecki, A., Schauer, A., Snell, K. E., Swart, P. K., Tripathi, A.,
 469 Upadhyay, D., Vennemann, T., Winkelstern, I., Yarian, D., Yoshida, N., Zhang, N. & Ziegler, M. (2021). InterCarb: A
 470 community effort to improve inter-laboratory standardization of the carbonate clumped isotope thermometer using
 471 carbonate standards. In review (*Geochemistry, Geophysics, Geosystems*).

472 Bernasconi, S. M., Hu, B., Wacker, U., Fiebig, J., Breitenbach, S. F. M. & Rutz, T. (2013). Background effects on Faraday col-
 473 lectors in gas-source mass spectrometry and implications for clumped isotope measurements. *Rapid Communications*
 474 *in Mass Spectrometry* 27:(5), pp. 603–612. doi: [10.1002/rcm.6490](https://doi.org/10.1002/rcm.6490).

475 Bernasconi, S. M., Müller, I. A., Bergmann, K. D., Breitenbach, S. F. M., Fernandez, A., Hodell, D. A., Meckler, A. N.,
 476 Millan, I. & Ziegler, M. (2018). Reducing uncertainties in carbonate clumped isotope analysis through consistent
 477 carbonate-based standardization. *Geochemistry, Geophysics, Geosystems* 19. doi: [10.1029/2017GC007385](https://doi.org/10.1029/2017GC007385).

478 Bonifacie, M., Calmels, D., Eiler, J. M., Horita, J., Chaduteau, C., Vasconcelos, C., Agrinier, P., Katz, A., Passey, B. H.,
 479 Ferry, J. M. & Bourrand, J.-J. (2017). Calibration of the dolomite clumped isotope thermometer from 25 to 350 °C, and
 480 implications for a universal calibration for all (Ca, Mg, Fe)CO₃ carbonates. *Geochimica et Cosmochimica Acta* 200,
 481 pp. 255–279. doi: [10.1016/j.gca.2016.11.028](https://doi.org/10.1016/j.gca.2016.11.028).

482 Daëron, M., Blamart, D., Peral, M. & Affek, H.P. (2016). Absolute isotopic abundance ratios and the accuracy of Δ_{47} mea-
 483 surements. *Chemical Geology* 442, pp. 83–96. doi: [10.1016/j.chemgeo.2016.08.014](https://doi.org/10.1016/j.chemgeo.2016.08.014).

484 Daëron, M. & Blamart, D. (2016). Uncertainties and standardization in the absolute reference frame. Abstract, Fifth In-
 485 ternational Clumped Isotope Workshop, Saint Petersburg, USA. URL: <https://hal.archives-ouvertes.fr/hal-02492075>.

486

487 Dennis, K. J., Affek, H. P., Passey, B. H., Schrag, D. P. & Eiler, J. M. (2011). Defining an absolute reference frame for
 488 ‘clumped’ isotope studies of CO₂. *Geochimica et Cosmochimica Acta* 75, pp. 7117–7131. doi: [10.1016/j.gca.2011.](https://doi.org/10.1016/j.gca.2011.09.025)
 489 [09.025](https://doi.org/10.1016/j.gca.2011.09.025).

490 Eiler, J. M. (2011). Paleoclimate reconstruction using carbonate clumped isotope thermometry. *Quaternary Science Reviews*
 491 30, pp. 3575–3588. doi: [10.1016/j.quascirev.2011.09.001](https://doi.org/10.1016/j.quascirev.2011.09.001).

492 Eiler, J. M. & Schauble, E. (2004). ¹⁸O¹³C¹⁶O in Earth’s atmosphere. *Geochimica et Cosmochimica Acta* 68:(23), pp. 4767–
 493 4777. doi: [10.1016/j.gca.2004.05.035](https://doi.org/10.1016/j.gca.2004.05.035).

494 Fernandez, A., Müller, I. A., Rodríguez-Sanz, L., Dijk, J. van, Looser, N. & Bernasconi, S. M. (2017). A Reassessment of
 495 the precision of carbonate clumped isotope measurements: implications for calibrations and paleoclimate reconstruc-
 496 tions. *Geochemistry, Geophysics, Geosystems* 18:(12), pp. 4375–4386. doi: [10.1002/2017gc007106](https://doi.org/10.1002/2017gc007106).

497 He, B., Olack, G. A. & Colman, A. S. (2012). Pressure baseline correction and high-precision CO₂ clumped-isotope (Δ_{47})
 498 measurements in bellows and micro-volume modes. *Rapid Communications in Mass Spectrometry* 26, pp. 2837–2853.
 499 doi: [10.1002/rcm.6436](https://doi.org/10.1002/rcm.6436).

500 Huntington, K. W., Eiler, J. M., Affek, H. P., Guo, W., Bonifacie, M., Yeung, L. Y., Thiagarajan, N., Passey, B., Tripathi, A.,
 501 Daëron, M. & al., et (2009). Methods and limitations of “clumped” CO₂ isotope (Δ_{47}) analysis by gas-source isotope
 502 ratio mass spectrometry. *Journal of Mass Spectrometry* 44:(9), pp. 1318–1329. doi: [10.1002/jms.1614](https://doi.org/10.1002/jms.1614).

503 John, C. M. & Bowen, D. (2016). Community software for challenging isotope analysis: First applications of “Easotope” to
 504 clumped isotopes. *Rapid Communications in Mass Spectrometry* 30:(21), pp. 2285–2300. doi: [10.1002/rcm.7720](https://doi.org/10.1002/rcm.7720).

505 Kocken, Ilja J., Müller, Inigo A. & Ziegler, Martin (2019). Optimizing the Use of Carbonate Standards to Minimize Un-
 506 certainties in Clumped Isotope Data. *Geochemistry, Geophysics, Geosystems* 20:(11), pp. 5565–5577. doi: [10.1029/](https://doi.org/10.1029/2019gc008545)
 507 [2019gc008545](https://doi.org/10.1029/2019gc008545).

508 Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied*
 509 *Mathematics* 2:(2), pp. 164–168. doi: [10.1090/qam/10666](https://doi.org/10.1090/qam/10666).

510 Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for*
 511 *Industrial and Applied Mathematics* 11:(2), pp. 431–441. URL: <http://www.jstor.org/stable/2098941>.

512 Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*
 513 46:(253), pp. 68–78. doi: [10.1080/01621459.1951.10500769](https://doi.org/10.1080/01621459.1951.10500769).

514 Meckler, A. N., Ziegler, M., Millán, M. I., Breitenbach, S. F. M. & Bernasconi, S. M. (2014). Long-term performance of the
 515 Kiel carbonate device with a new correction scheme for clumped isotope measurements. *Rapid Communications in*
 516 *Mass Spectrometry* 28, pp. 1705–1715. doi: [10.1002/rcm.6949](https://doi.org/10.1002/rcm.6949).

517 Newville, M., Stensitzki, T., Allen, D. B. & Ingargiola, A. (2014). *LMFIT: Non-Linear Least-Square Minimization and Curve-*
 518 *Fitting for Python*. doi: [10.5281/zenodo.11813](https://doi.org/10.5281/zenodo.11813).

519 Olack, Gerard & Colman, Albert S. (2019). Modeling the Measurement: Δ_{47} , Corrections, and Absolute Ratios for Refer-
 520 ence Materials. *Geochemistry, Geophysics, Geosystems*. doi: [10.1029/2018gc008166](https://doi.org/10.1029/2018gc008166).

- 521 Petersen, S. V., Defliese, W. F., Saenger, C., Daëron, M., John, C. M., Huntington, K. W., Kelson, J. R., Bernasconi, S
 522 .M., Colman, A. S., Kluge, T., Olack, G. A., Schauer, A. J., Bajnai, D., Bonifacie, M., Breitenbach, S. F. M., Fiebig,
 523 J., Fernandez, A. B., Henkes, G. A., Hodell, D., Katz, A., Kele, S., Lohmann, K. C., Passey, B. H., Peral, M., Petrizzo,
 524 D. A., Rosenheim, B. E., Tripathi, A., Venturelli, R., Young, E. D., Wacker, U. & Winkelstern, I. Z. (2019). Effects of Im-
 525 proved ¹⁷O Correction on Interlaboratory Agreement in Clumped Isotope Calibrations, Estimates of Mineral-Specific
 526 Offsets, and Temperature Dependence of Acid Digestion Fractionation. *Geochemistry, Geophysics, Geosystems*. doi:
 527 [10.1029/2018gc008127](https://doi.org/10.1029/2018gc008127).
- 528 Schauer, A. J., Kelson, J., Saenger, C. & Huntington, K. W. (2016). Choice of ¹⁷O correction affects clumped isotope (Δ_{47})
 529 values of CO₂ measured with mass spectrometry. *Rapid Communications in Mass Spectrometry* 30:(24), pp. 2607–2616.
 530 doi: [10.1002/rcm.7743](https://doi.org/10.1002/rcm.7743).