A community effort to improve inter-laboratory standardization of the carbonate clumped isotope thermometer using carbonate anchors

Stefano Bernasconi[1,1,2,2], Mathieu Daëron[3,3,4,4], Kristin D Bergmann[5,5,5,5], Magali Bonifacie[6,6,6,6], and Anna Nele Meckler[7,7,7,7]

[1]ETH Zürich
[2]ETH Zurich
[3]Université Paris-Saclay
[4]LSCE Gif sur Yvette
[5]Massachusetts Institute of Technology
[6]Université de Paris, Institut de Physique du Globe de Paris
[7]University of Bergen

November 30, 2022

Posted on 30 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10504430/v4 — This a preprint and has not been peer reviewed. Data may be preliminary.

Abstract

Increased adoption and improved methodology in carbonate clumped isotope thermometry has greatly enhanced our ability to interrogate a suite of Earth-system processes. However, interlaboratory discrepancies in quantifying carbonate clumped isotope ($\Delta47$) measurements persist, and their specific sources remains unclear. To address interlaboratory differences as a clumped isotope community, we first provide community consensus values for four carbonate standards relative to heated and equilibrated gases with 1,447 individual measurements from nine laboratories. Then we analyzed the four carbonate standards and an additional three standards, spanning a broad range of $\Delta47$ and $\Delta47$ compositions, a total of 5,202 times on 26 mass spectrometers representing 23 unique laboratories. We use three standards to calculated values for the other four standards and find that the use of carbonate reference materials is a robust method for standardization that yields inter-lab discrepancies consistent with intra-lab analytical uncertainty. The use of carbonate reference materials, along with measurement and data processing practices described herein, provides the carbonate clumped isotope community with a robust way to achieve interlaboratory agreement as we continue to use and improve this powerful geochemical tool.

InterCarb: A community effort to improve inter-laboratory standardization of the carbonate clumped isotope thermometer using carbonate standards

S.M. Bernasconi[1], M. Daëron[2], K.D. Bergmann[3], M. Bonifacie[4], A.N. Meckler[5], H.P. Affek[6], N. Anderson[3], D. Bajnai[7], E. Barkan[6], E. Beverly[8†], D. Blamart[2], L. Burgener[9], D. Calmels[4†], C. Chaduteau[4], M. Clog[10], B. Davidheiser-Kroll[11], A. Davies[12††], F. Dux[13†], J. Eiler[14], B. Elliott[15], A. C. Fetrow[11], J. Fiebig[7], S. Goldberg[3], M. Hermoso[4.23], K.W. Huntington[16], E. Hyland[9], M. Ingalls[14†], M. Jaggi[1], C.M. John[12], A. B. Jost[3], S. Katz[8], J. Kelson[8], T. Kluge[12†], I.J. Kocken[17], A. Laskar[18], T.J. Leutert[5†], D. Liang[18], J. Lucarelli[15], T. J. Mackey[3†], X. Mangenot[4,14], N. Meinicke[5], S.E. Modestou[5], I.A. Müller[17], S. Murray[19], A. Neary[8], N. Packard[8], B.H. Passey[8], E. Pelletier[8], S. Petersen[8], A. Piasecki[5††], A. Schauer[16], K.E. Snell[11], P.K. Swart[20], A. Tripati[15], D. Upadhyay[15], T. Vennemann[21], I. Winkelstern[8††], D. Yarian[8], N. Yoshida[22,24], N. Zhang[22], M. Ziegler[17]

[1]Geological Institute, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland

[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, F-91191 Gif-sur- Yvette, France

[3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3†]now at Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico, USA

[4]Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France

[4†] now at Geosciences Paris Sud (GEOPS), Université Paris-Saclay, CNRS, 91405 Orsay, France

[5]Bjerknes Centre for Climate Research and Department of Earth Science, University of Bergen, Bergen, Norway

[5†]now at Max Planck Institute for Chemistry, Mainz, Germany; thomas.leutert@mpic.de

[5††]now at Department of Earth Sciences, Dartmouth College, Hanover, NH, USA

[6]Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel

[7]Institute of Geosciences, Goethe University Frankfurt, Frankfurt am Main, Germany

[8]Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA

[8†]now at Department of Earth and Atmospheric Sciences, University of Houston, Houston, Texas, USA

[8††]now at Geology Department, Grand Valley State University, Allendale, Michigan, USA

[9]Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, USA

[10]Scottish Universities Environmental Research Centre (SUERC), Glasgow, Scotland

[11]University of Colorado Boulder, Boulder, Colorado, USA

[12]Imperial College, London, UK
[12†]now at Karlsruher Institut für Technologie KIT
[12††]now at Stockholm University
[13]School of Geography, University of Melbourne, Australia
[13†]now at School of Earth and Life Sciences, University of Wollongong, Australia
[14]Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA
[14†]now at Department of Geosciences, The Pennsylvania State University, University Park, PA, USA
[15]Department of Earth, Planetary, and Space Sciences, University of California Los Angeles, Los Angeles, CA, USA
[16]University of Washington, Seattle, Washington, USA
[17]Department of Earth Sciences, University of Utrecht, Utrecht, Netherlands
[18]Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan
[19]Macquarie University, Australia
[20] Department of Marine Geosciences, Rostiel School of Marine and Atmospheric Sciences, University of Miami, Miami, FL, USA
[21]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland
[22]Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan
[23] Univ. Littoral Côte d'Opale, Univ. Lille, CNRS, Laboratoire d'Océanologie et de Géosciences (UMR 8187 LOG), 62930 Wimereux, France.
[24] National Institute of Information and Communications Technology, Tokyo, Japan

Corresponding authors: Stefano M. Bernasconi (stefano.bernasconi@erdw.ethz.ch) Mathieu Daëron (daeron@lsce.ipsl.fr)

Key Points:

- The exclusive use of carbonate reference materials is a robust method for the standardization of clumped isotope measurements
- Measurements using different acid temperatures, designs of preparation lines, and mass spectrometers are statistically indistinguishable
- We propose new consensus values for a set of 7 carbonate reference materials and updated guidelines to report clumped isotope measurements

Abstract

Increased use and improved methodology of carbonate clumped isotope thermometry has greatly enhanced our ability to interrogate a suite of Earth-system processes. However, inter-laboratory discrepancies in quantifying carbonate clumped isotope (Δ_{47}) measurements persist, and their specific sources remain unclear. To address inter-laboratory differences, we first provide consensus values from the clumped isotope community for four carbonate standards relative to heated and equilibrated gases with 1,819 individual analyses from 10 laboratories. Then we analyzed the four carbonate standards along with three additional standards, spanning a broad range of δ^{47} and Δ_{47} values, for a total of 5,329 analyses on 25 individual mass spectrometers from 22 different laboratories. Treating three of the materials as known standards and the other four as unknowns, we find that the use of carbonate reference materials is a robust method for standardization that yields inter-laboratory discrepancies entirely consistent with intra-laboratory analytical uncertainties. Carbonate reference materials, along with measurement and data processing practices described herein, provide the carbonate clumped isotope community with a robust approach to achieve inter-laboratory agreement as we continue to use and improve this powerful geochemical tool. We propose that carbonate clumped isotope data normalized to the carbonate reference materials described in this publication should be reported as Δ_{47} (I-CDES) values for Intercarb-Carbon Dioxide Equilibrium Scale.

1 Introduction

Carbonate clumped isotope (Δ_{47}) thermometry is the most developed branch of the rapidly evolving field of clumped isotope geochemistry. Given the broad range of applications in Earth Sciences (e.g. Affek & Eiler, 2006; Eagle et al., 2010; Ferry et al., 2011; Ghosh, Adkins, et al., 2006; Ghosh, Garzione, et al., 2006; Grauel et al., 2013; Guo & Eiler, 2007; Huntington et al., 2011; Passey & Henkes, 2012, Dale et al, 2014, Mangenot et al., 2018, Veillard et al, 2019) and the improvement of analytical methods including automation (Bernasconi et al., 2013, 2018; Defliese & Lohmann, 2015; Dennis et al., 2011; Fiebig et al., 2019; Ghosh, Adkins, et al., 2006; He et al., 2012; Hu et al., 2014; Huntington et al., 2009; Meckler et al., 2014; Müller, Fernandez, et al., 2017; Passey et al., 2010; Petersen et al., 2019; Petersen et al., 2016; Schmid & Bernasconi, 2010, Adlan et al, 2020), the last 5-10 years have seen an increasing number of laboratories implementing this technique. The great potential of this thermometer can only be fully exploited if precision and accuracy are sufficient to resolve differences of a few degrees in formation temperatures. In addition, widely available reference materials that match the sample matrices are necessary so that data can be robustly compared across laboratories (Meier-Augenstein & Schimmelmann, 2019). Currently the situation in the field of carbonate clumped isotope geochemistry is far from satisfactory. Published values for the ETH reference materials, the only carbonates that have been recently measured in many different laboratories worldwide, differ by up to 0.053 ‰ (see Bernasconi et al. 2018, Thaler et al. 2020 for recent comparisons). For paleoclimate applications, however, a repeatability across laboratoires of 0.01 ‰ or better is a necessary goal for meaningful data comparison. This clearly calls for better standardization procedures to improve laboratory comparability.

The data normalization scheme currently used in clumped isotope geochemistry of carbonates in many laboratories is based on the comparison of the composition of the CO_2 liberated from carbonates by reaction with phosphoric acid with that of a set of CO_2 gases with different bulk and clumped isotope compositions (Dennis et al., 2011). These gases are prepared either by heating CO_2 at 1000 °C (heated gases; HG) or by CO_2 equilibration with water at low temperatures (equilibrated gases at *e.g.* 25 °C, 50 °C; EG). By comparing the measured compositions with the theoretical predictions of the equilibrium thermodynamic abundance of multiply substituted isotopologues in heated and equilibrated gases (Wang et al. 2004 and updates in Petersen et al.

2019), the measurements are standardized to the scale that was named the "absolute reference frame" (ARF) by Dennis et al., (2011). In more recent publications, the ARF is often referred to as the "Carbon Dioxide Equilibration Scale" (CDES), a terminology introduced by Passey and Henkes (2012). This approach was designed to allow different laboratories to link their measurements to an internationally recognized scale firmly anchored to theory using relatively easy and established laboratory protocols to produce CO_2 standard gases of known isotopic composition. Early comparisons of $\Delta_{47\,\mathrm{CDES}}$ values for carbonates analyzed in different laboratories and corrected with HG/EG normalization were promising (Dennis et al., 2011). While Bonifacie et al. (2017) reported similar $\Delta_{47\,\mathrm{CDES}}$ values for nine dolomite samples covering a range of almost 0.4 ‰ measured both at Caltech and IPGP laboratories with HG/EG normalization, Spooner et al. (2016) found that carbonate standardization improved agreement between data they obtained on samples analyzed both at Caltech and WHOI laboratories, compared to when they were using HG/EG normalization. Such recurrent cases of poor inter-laboratory reproducibility (see also Bernasconi et al., 2018 and Thaler et al., 2020) suggest that there are still unexplained differences in the results among laboratories (see Petersen et al. (2019) for a recent review).

Apart from preservation problems, two known issues still limiting the reliability of this method to yield accurate temperature reconstructions are: [1] the lack of internationally recognized carbonate reference materials for a precise inter-laboratory calibration, and [2] that published Δ_{47}-temperature calibrations produced in different laboratories have differed in both temperature dependence (slope) and absolute value (intercept). Possible reasons for the differences in slopes and intercepts of the Δ_{47} temperature dependence have been widely discussed in the literature (e.g. Kluge et al, 2015; Bonifacie et al., 2017; Daëron et al., 2016; Fernandez et al., 2017; Katz et al., 2017; Kelson et al., 2017; Petersen et al., 2019; Schauer et al., 2016). Discrepancies have been attributed to analytical artefacts such as CO_2-acid re-equilibration at different acid digestion temperatures (see Wacker et al. 2013; Swart et al., 2019, for a recent discussion) and to slight pressure imbalances between sample and reference gas (Fiebig et al., 2016). Other factors proposed to influence the calculated slopes of the calibrations are the limitations of the datasets used in the individual studies, in particular in terms of the number of samples and replicates and of the temperature range covered by the available samples (Bonifacie et al., 2017; Fernandez et al., 2017). However, the discrepancies in the intercepts of the calibrations, for example between

Kelson et al. (2017) and Peral et al. (2018), and a generally poor laboratory comparability remain problems that could be mitigated by using a more robust standardization method.

Petersen et al. (2019), in a recent effort to resolve differences in calibrations, compiled raw data of a number of published temperature calibrations and recalculated them all in a consistent way using the revised IUPAC correction parameters to correct for the ^{17}O abundance (Daëron et al., 2016; Schauer et al., 2016). The goal was to test whether data processing differences and/or the use of consistent but incorrect ^{17}O- correction parameters in the calculations were the root causes of inconsistencies. The result of this study was that differences among calibrations were reduced but not eliminated by the recalculation, implying that other factors must be responsible for the remaining discrepancies. These differences have pushed many laboratories to use laboratory-specific calibrations performed with the same analytical approach, as they at least partially take into consideration possible procedural differences (Petersen et al., 2019). However, if a laboratory changes analytical procedures or has not generated a robust in-house calibration, this approach is problematic. Achieving an inter-laboratory reproducibility at the level of accuracy necessary for meaningful interpretations of the observed variations is a requirement for Δ_{47} thermometry to reach its potential as a mature analytical method with broad acceptance and quantitative usefulness.

While the definition of the CDES was a major milestone (Dennis et al., 2011), a known problem with this approach is that while the CO_2 standard gases equilibrated at known temperature (HG or EG) can be confidently used for correction of mass spectrometric fractionations/nonlinearities and for effects of the purification procedures, they cannot account for the effects of the phosphoric acid reaction on the composition of the produced CO_2. Among the factors responsible for discrepant calibrations and laboratory comparability, two important ones cannot be tested with a gas-based standardization: (1) the absolute value and temperature dependence of the phosphoric acid fractionation factor (see Petersen et al., 2019 for a recent compilation) and (2) possible CO_2 equilibration effects during acid digestion of the sample. Swart et al. (2019) presented evidence that equilibration of CO_2 with water or hot metal surfaces during phosphoric acid reaction and transfer of the CO_2 to the mass spectrometer could be a factor leading to the alteration of the apparent temperature dependence of clumped isotopes in carbonates and on the absolute value of calculated Δ_{47}. As many laboratories use custom built extraction lines with different designs and

volumes of tubing and of acid vessels, these factors are impossible to precisely quantify for each laboratory and may further contribute to inter-laboratory discrepancies.

We propose that these issues can be circumvented if carbonates, which undergo the same acid digestion as the samples, are used for normalization instead of or in addition to gases, consistent with the principle of identical treatment of sample and standards (Carter & Fry, 2013; Werner & Brand, 2001a). In addition, normalizing results to accepted carbonate reference material values, as is commonly done with conventional carbon and oxygen isotope analysis in carbonates, removes the requirement to precisely quantify acid fractionation factors at different temperatures (Bernasconi et al., 2018).

A carbonate standardization approach was introduced by Schmid & Bernasconi (2010) and improved by Meckler et al. (2014), with the following benefits: (1) the use of carbonates can more easily be fully automated, eliminating time-consuming and possibly error-prone manual preparation of CO_2 standard gases (equilibrated at known temperature) by individual users on separate extraction lines; (2) in some automated systems designed for the measurement of small carbonate samples (e.g. the Kiel Device), the heated and equilibrated gases had to be measured through a different capillary than the gases produced by acid digestion of carbonates with potential biases that would go unrecognized; and (3) in these same systems the equilibrated and heated gases are measured at constant ion beam intensity in bellow mode, whereas the samples are measured with decreasing ion beams in microvolume mode. These features argue in favor of carbonate standardization *a priori*, but it remains critical to assess *a posteriori* whether the results of this approach are as robust and accurate as expected and whether they significantly improve the inter-laboratory reproducibility of Δ_{47} measurements. Discussions at the Sixth International Clumped Isotope Workshop (Paris, 2017) led to the present inter-laboratory comparison exercise (InterCarb) to evaluate the benefits and drawbacks of a carbonate-based standardization approach as an alternative to the use of gas standards.

The primary goal of this study was to test whether the exclusive use of carbonate reference materials as a substitute for heated and equilibrated gases can minimize inter-laboratory discrepancies and provide an alternative to the measurement of heated and equilibrated gases for the entire community. This is particularly important because of the increasing number of

laboratories using commercial small-sample automated devices which cannot easily be standardized using the HG-EG approach. The InterCarb exercise also provides an opportunity to define the best community-derived consensus Δ_{47} values for the ETH standards of Meckler et al. (2014). Although these standards are already used in many laboratories, their current nominal Δ_{47} values are based on measurements from the ETH laboratory only. The InterCarb exercise can similarly establish community accepted values for other common carbonate reference materials, some of which have been in use for several years, in order to provide the community with a self-consistent set of carbonate reference materials with a broad range of bulk and clumped isotope compositions.

1.1. Nomenclature and data processing

Clumped isotope compositions are reported as an excess abundance of the CO_2 isotopologue of cardinal mass 47 (dominantly the isotopologues $^{13}C^{18}O^{16}O$) compared to a stochastic distribution according to the formula:

$$\Delta_{47} = R^{47} / R^{47*} - 1$$

where R^{47} is the ratio of the abundances of the set of minor isotopologues with mass 47 (mostly $^{13}C^{18}O^{16}O$ and trace amounts of $^{12}C^{17}O^{18}O$ and $^{13}C^{17}O_2$) divided by the abundance of the most abundant isotopologue with mass 44 ($^{12}C^{16}O_2$). The stochastic ratio R^{47*} is calculated using the measured abundance of ^{13}C and ^{18}O and measured or calculated abundance of ^{17}O in the sample (Affek & Eiler, 2006). According to the IUPAC guidelines the formula does not include the factor 1000 (Coplen, 2011; though Δ_{47} is commonly reported in units of per mil, which implies multiplication by a factor of 1000). Also, we omit here the classically included terms involving R45* and R46*, which are assumed to be zero by definition when computing $\delta^{13}C$ and $\delta^{18}O$, and in practice never exceed ±0.00002 ‰ in our calculations (Daeron et al. 2016). The measured abundance of isotopologues with *m/z* 47 in the sample with respect to the working gas (WG) in the mass spectrometer is reported in the traditional delta notation as:

$$\delta^{47} = R^{47} / R^{47}_{WG} - 1$$

The δ^{47} scale is a measure of the difference between the sample of interest and the WG of the specific instrument, therefore, it cannot be compared across laboratories. The same notation is used for masses 45, 46, 48 and 49.

The CO_2 gas-based standardization scheme for clumped isotope thermometry in carbonates relies on a set of CO_2 standard gases with different bulk compositions ($\delta^{13}C$ and $\delta^{18}O$, leading to different δ^{47}), preferably chosen by the user to encompass the δ^{47} values of unknown samples that have been (1) heated at 1000 °C to reach a near-stochastic distribution of all isotopologues, or (2) equilibrated with water at low temperature to reach equilibrium enrichments in the mass-47 isotopologues (Dennis et al., 2011). The heated gases, having a near-stochastic distribution of the heavy isotopes among all isotopologues, define the zero point of the CDES scale, through the assumption that at 1000 ˚C these gases achieve a Δ_{47} = 0.0266 ‰, and the water-equilibrated gases define a second, generally higher point on this scale (e.g., at 25 °C Δ_{47} = 0.9196 ‰). The theoretical values linking measurements to theory were calculated by Wang et al. (2004) and revised by Petersen et al. (2019). A wide range in δ^{47} values of gases used for normalization is generally chosen to allow for accurate correction for an apparent dependence of Δ_{47} on δ^{47}, which is caused by inaccurate pressure-dependent background corrections on the *m/z* 47 collector observed on many instruments (Bernasconi et al., 2013; He et al., 2012). The large range in Δ_{47} (i.e., 25 °C, 1000 °C), on the other hand, is necessary to correct for scale compression caused by processes of scrambling and molecule recombination in the source of the mass spectrometer or elsewhere in the sample preparation, transfer lines and/or the capillaries (Dennis et al., 2011, Swart et al., 2019). With properly chosen CO_2 standard gases with widely varying δ^{47} values it is possible to cover the entire range of natural carbonate compositions, avoiding extrapolations in the δ^{47} vs. Δ_{47} compositional space (Fig. 1). Note that with measurement errors (typically no better than 0.010 ‰) being relatively large compared to the natural compositional range (less than 0.5 ‰; Fig. 1), the large (0.9 ‰) difference in Δ_{47} of the CO_2 standard gases minimizes errors introduced by uncertainties resulting from the measurement of HG and EG.

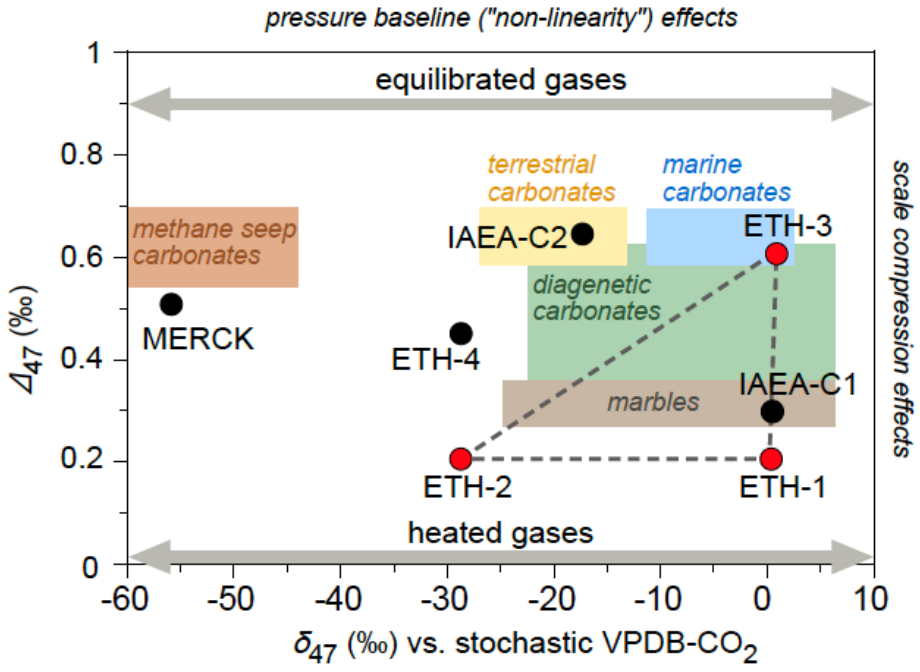


Figure 1. The δ^{47} vs. Δ_{47} values of carbonate standards (Δ_{47} on the I-CDES scale proposed here) and heated and equilibrated gases in comparison to the compositional ranges of typical natural carbonates. The observed range in measured clumped isotope compositions in natural carbonates can be completely bracketed by heated and equilibrated CO_2 standard gases from which δ^{47} values have been chosen by the user. The δ^{47} values for the anchor samples used in InterCarb (red) and the unknowns (black) are reported for a theoretical working gas with stochastic isotope distribution, derived from VPDB. Actual δ^{47} values will vary by laboratory depending on the composition of the working gas. Note the smaller achievable range in both δ^{47} and Δ_{47} values when using carbonate standards compared to heated and equilibrated gases and the large extrapolation necessary for the determination of the composition for MERCK. Heated and equilibrated CO_2 standard gases have a larger Δ_{47} range, allowing for more robust stretching calculations with identical numbers of standard:sample analyses.

Meckler et al. (2014) attempted to achieve a similar framework as the CO_2 gas-based standardization but with carbonate standards. They described four carbonates that were developed at ETH Zürich to serve as replacements for HG's and EG's and demonstrated that good long- and short-term reproducibility can be achieved using only carbonates for data correction. Bernasconi et. al (2018) discussed in detail these standards and postulated, based on a limited inter-laboratory dataset, that carbonate standardization should generally improve inter-laboratory data

comparability. This claim seems arguably strengthened by the results of Meinicke et al. (2020), Peral et al. (2018), Piasecki et al. (2019), Kele et al. (2015) as recalculated by Bernasconi et al. 2018), and Jautzy et al., (2020). The first three studies produced independent foraminifera-based the fourth a travertine and the fifth a synthetic carbonate-based Δ_{47}-temperature calibration anchored to the same set of carbonate standards. These studies yielded statistically indistinguishable slopes and intercepts despite the use of independent sample sets and in the case of Peral et al. (2018), a different analytical system. In addition, a reanalysis of samples from five previous calibrations by Anderson et al. (2021) using carbonate standardization revealed no significant differences in temperature dependence of Δ_{47} between the different sample sets. This, solved a long standing debate about variations in slope among calibrations.

A possible limitation of carbonate standardization is that available carbonates have a smaller range in δ^{47} and, perhaps more importantly, a smaller range in Δ_{47} values than what is achievable with heated and equilibrated gases. In some specific cases, standardization procedures require extrapolation to compositions that are not within the δ^{47}-Δ_{47} space created by carbonate standards (Fig. 1). In addition, the range of Δ_{47} values for carbonates is only on the order of 0.5 ‰ between 0 and 1000 °C. The smaller range in Δ_{47} compared to HG's and EG's requires higher precision and also a larger number of replicates of both standards and samples. Daëron (2021) and Kocken et al., 2019 suggest ~50:50 ratio of standard to sample replicates to keep standardization errors small.

1.2. InterCarb goals and design

InterCarb was designed with the aim to carefully evaluate the potential of carbonates to serve as a standardization scheme that improves inter-laboratory agreement for ‘unknown’ carbonates both inside and outside of the δ^{47}-Δ_{47} space defined by the anchor samples (Fig. 2). The main questions posed are:

1. Is it possible to produce consistent carbonate clumped isotope measurements across laboratories using carbonate reference materials exclusively? In other words, does the

observed inter-laboratory scatter in Δ_{47} values match that expected from intra-laboratory analytical precision?

2. How well does the carbonate standardization approach perform when extrapolating beyond the δ^{47}-Δ_{47} compositional space sampled by a set of carbonate reference materials?
3. Do carbonate reference materials fully correct effects arising from different reaction temperatures, sample preparation protocols, and analytical equipment?
4. Can we define a self-consistent set of widely available reference materials with community-agreed compositions accurately anchored to the CDES scale?
5. Does the use of carbonate reference materials for standardization improve the inter-laboratory reproducibility compared to using HG's and EG's?

1.3 Approach

Seven carbonate standards with a large range of δ^{47} and Δ_{47} values (Fig. 1) were distributed among participating laboratories and analyzed, treating three carbonates as "anchors" (whose Δ_{47} values are assigned *a priori*) and the remaining four as "unknowns" (whose Δ_{47} values are unknown, to be determined by comparison with the anchors). Due to their relatively widespread use in different laboratories, the three reference materials ETH-1, ETH-2 and ETH-3 (Meckler et al. 2014; Bernasconi et al. 2018) were chosen as anchors. They are still available today in relatively large quantities (>600 g), have been in use at ETH since 2013 and in many other laboratories for several years. Importantly, they have been thoroughly tested for homogeneity based on thousands of measurements in 80 to 150 µg aliquot sizes in different laboratories and no changes in composition have been noticed at ETH in the 7 years they have been in use.

The "unknown" InterCarb reference materials were chosen to cover a wide natural range in δ^{47} and Δ_{47} values. These samples had to be available in large quantities, inexpensive, and if possible distributed by an organization with a long-term perspective in order to ensure future data quality and availability for the increasing number of laboratories globally.

2 Materials and Methods

2.1. Sample description

The anchor samples ETH-1 (Carrara marble heated at 600 °C), ETH-2 (synthetic carbonate heated at 600 °C) and ETH-3 (Upper cretaceous chalk) are described in detail in Bernasconi et al. (2018).

IAEA-C1 (marble from Carrara, Italy) is distributed by the International Atomic Energy Agency (IAEA) as a mechanically crushed and milled product with grains ranging from 1.6 to 5 mm. All 50 g provided were ground and thoroughly homogenized in a ball mill at ETH Zürich to a grain size of less than 100 µm and transferred in 0.5 g aliquots to plastic vials for distribution. Nishida and Ishimura (2017) found that IAEA 603, which was produced from the same coarse marble as IAEA C-1, was isotopically inhomogeneous. Whitish grains (1–2 per 100 grains; grain weight, 8–63 µg) were significantly depleted in ^{18}O and ^{13}C compared to translucent grains. In this study we found no evidence of inhomogeneity in Δ_{47} for sample aliquots of 80-110 µg after the original material was ground in the ball mill.

IAEA-C2 is a freshwater travertine from Bavaria distributed by IAEA as a powder which was treated identically to IAEA-C1. XRD analysis shows it to be calcite (supplementary Fig. S1).

ETH-4 is a commercially available synthetic calcium carbonate (Riedel-De Haën; calcium carbonate Puriss. p.a.; Lot No. 30800) determined to be calcite by XRD (supplementary Fig. S2 with intermediate formation temperature and the same bulk isotope composition as ETH-2 (see Bernasconi 2018 for details).

MERCK (Catalog No. 1.02059.0050; lot no. B1164559 515) is an ultra-pure, commercially available synthetic calcium carbonate determined to be calcite by XRD (Müller et al., 2019) and was chosen for its very low $\delta^{13}C$ and $\delta^{18}O$ values of approximately -42.2 ‰ and -15.5 ‰ (VPDB), respectively. This sample represents an extreme case of extrapolation from the δ47-Δ47 space defined by the anchor materials (Fig. 2). The same product was recently used to prepare the carbon isotope reference material USGS44 by Qi et al. (2020) which, after careful determination of its Δ47 could be used as a substitute for the aliquots of MERCK distributed for this study.

2.2 Instrumentation

The reported data were produced with a variety of preparation systems including custom built (13 laboratories) and commercial systems (11 laboratories; Protium MS IBEX, ThermoFisher Scientific Kiel IV device and Nu Instruments Nucarb). Reaction temperatures were generally 90 °C for "large-sample" custom preparation systems and 70 °C for the Kiel and the NuCarb. Four mass spectrometer types were used: Thermo Fisher Scientific MAT253 and 253Plus, Nu Instruments Perspective, and Elementar Isoprime 100. All participants contributed results they considered to be of "publication-grade" quality, based on their existing quality-control procedures.

2.3 Clumped isotope compositions of the ETH anchor materials

The clumped isotope compositions of the four ETH reference materials relative to the CO_2 reference frame CDES were first reassessed based on new data provided by 10 laboratories that also provided HG and EG data measured during the same sessions as the ETH reference materials. The data were processed with the same Python script used for the carbonate data in order to avoid any differences in data processing (see section 2.4).

Although, strictly speaking, ^{13}C-^{18}O clumping in carbonate represents a mass-63 anomaly, the clumped isotope composition of carbonate minerals is reported as Δ_{47} i.e. as the mass-47 excess in the CO_2 produced by acid digestion of these minerals, including the respective temperature-

dependent isotopic fractionation. As initially all reactions were carried out at 25 °C (Ghosh, Adkins, et al., 2006), the Δ_{47} values have traditionally been reported for a 25 °C acid temperature. With the advent of automated extraction lines, reaction temperatures have been increased to 70 or 90 °C. To account for the temperature dependence of the acid fractionation factor (Guo et al., 2009, Petersen et al. 2019) and to project these results back to the original 25 °C acid reactions, various acid temperature correction values have been reported over time, based on experimental observations and/or theoretical predictions. Given that here seven out of ten laboratories reacted carbonates at 90 °C, two at 70 °C, and only one at 25 °C, our redetermination of the Δ_{47} values of ETH-1/2/3/4 relative to the CDES projected to 25°C would rely substantially on the accuracy of these acid temperature corrections (which typically range between 60–90 ppm). For this reason we report the Δ_{47} values of CO_2 produced by reacting ETH-1/2/3/4 at 90 °C. With this choice the numerical effect of poorly known acid corrections is minimized because the data from 70 °C and 25 °C reactions have relatively little influence on the final, error-weighted average Δ_{47} values (cf statistical weights in Fig. 2). We thus propose to break with tradition and define the nominal Δ_{47} values of the anchor standards as those of CO_2 produced at 90 °C, providing the most robust relationship to the CDES.

2.4 Data processing, correction and error assessment

It should be stressed that the InterCarb experiment, by design, is not intended to grade the analytical performance of individual laboratories. Each participating laboratory (or mass spectrometer, in the case of laboratories with several instruments) was thus randomly assigned an anonymous identifying number. Within each laboratory, analyses were grouped in different analytical sessions defined by the participants themselves. An analytical session is generally defined by a time in which the behavior of the analytical system (preparation system, source tuning, backgrounds, isotope scrambling in the source) is considered to be similar. The database record of each analysis consists of a laboratory identifier, a session identifier, an analysis identifier, the name of the analyzed sample, the mass spectrometer model, the acid reaction temperature, the mass of the reacted carbonate, and background-corrected δ^{45}, δ^{46} and δ^{47} values.

The only instrumental corrections to the raw data applied independently by each participating laboratory were background corrections ("Pressure Baseline Correction" or PBL) to the ion currents/voltages (Bernasconi et al., 2013; He et al., 2012, Fiebig et al., 2016; Fiebig et al., 2019). The PBL is strongly dependent on instrument design (it is not observed in some instruments) and configuration, and varies temporally depending on many factors. This correction, therefore, can only be carried out by each participating laboratory according to its own established procedures and monitoring.

To avoid artefacts arising from different calculation/standardization procedures, rounding errors, and ^{17}O correction parameters, raw data from all laboratories were processed by a single Python script (http://doi.org/10.5281/zenodo.4314448) based on data reduction, standardization and error propagation methods described in detail in the companion paper (Daëron, 2021). Here we briefly summarize these calculations.

Session-averaged, background-corrected δ^{45} and δ^{46} values for each of the three anchor samples were first used to calculate the bulk isotope composition of the working gas used in each session, based on (a) previously reported $\delta^{13}C_{VPDB}$ and $\delta^{18}O_{VPDB}$ values of ETH-1, ETH-2, and ETH-3 (Bernasconi et al., 2018), (b) the IUPAC ^{17}O correction parameters of Brand et al. (2010), and (c) a temperature-dependent oxygen-18 acid fractionation factor between CO_2 and calcite of Kim et al. (2015). This recalculation of working gas bulk compositions avoids (small) discrepancies potentially introduced by inaccuracies in the nominal compositions of the working gases.

Raw Δ_{47} values were computed according to:

$$\Delta_{47}^{raw} = R^{47} / R^{47*} - 1$$

Where R^{47} is the measured ratio and R^{47*} the calculated stochastic ratio of mass 47 over mass 44 of CO_2, assuming perfectly linear IRMS measurements and a stochastic working gas. Values are then normalized to "absolute" Δ_{47} values (noted Δ_{47}^{abs} in the equation below, and simply Δ_{47} thereafter) using session-specific relationships of the form:

$$\Delta_{47}^{raw} = a\,\Delta_{47}^{abs} + b\,\delta^{47} + c$$

For each session, the best-fit standardization parameters (a, b, c) are computed from an unweighted least-squares regression, treating Δ_{47}^{raw} as the response variable, only considering the three anchor samples ETH-1, ETH-2, and ETH-3. Note the advantage of this form over that in Dennis et al. 2011 is the ability to have three standards with distinct Δ_{47} values whilst being able to solve for b (compositional nonlinearity) (Daëron et al. 2016). Absolute Δ_{47} values are then computed for all replicates within that session. Standardization parameters for all sessions are listed in Table 2.

Throughout this study, the analytical error assigned to each individual raw Δ_{47} analysis is equal to the pooled "external" repeatability of raw Δ_{47} measurements of anchors and unknowns within each session. In the figures and tables, final measurement uncertainties are reported as standard errors and/or 95 % confidence limits, considering fully-propagated errors taking into account reference frame corrections. In Figures 2 and 4, different types of error bars are used to represent analytical errors only considering uncertainties in the analyses of a given sample or the full uncertainty considering standardization uncertainties (the "autogenic" errors of Daëron, 2021). In both cases, the analytical error assigned to each individual raw Δ_{47} analysis is equal to the pooled "external" repeatability of raw Δ_{47} measurements for all samples (anchors and unknowns) within each session. This treatment of error is a new approach that more fully accounts for error in both the sample measurement and reference frame.

3 Results and Discussion

3.1 Redetermination of nominal Δ47 values for the ETH standards relative to heated and equilibrated CO2 gases

The weighted averages of the 4 standards (projected to 90 °C for the reactions at 25 and 70°C using the acid temperature correction suggested by Petersen et al. 2019), comprising 873 analyses of the carbonate standards and 946 heated and equilibrated gases from 10 different laboratories, are reported in Table 1 and Fig 4. The large number of analyses and the appropriate consideration of the errors on the anchors (CO_2 gas analyses) distinguishes this effort from previous work and allow a robust redetermination of the accepted values of the ETH reference materials with 1SE uncertainties of 2 ppm or less.

	Laboratory	all	A	B	C	D	E	F	G	H	I	J
	N of sessions	**34**	4	4	11	7	1	2	1	1	1	2
	N of H/E CO2	**946**	44	193	257	85	47	21	38	192	13	56
ETH-1	**N of analyses**	**232**	34	14	5	54	4	4	19	58	8	32
	Δ_{47} (‰; 90 °C acid)	**0.2052**	0.2016	0.1926	0.2108	0.1940	0.1601	0.2013	0.2143	0.1932	0.2183	0.2152
	± 1SE	**0.0016**	0.0046	0.0058	0.0069	0.0042	0.0245	0.0107	0.0032	0.0045	0.0109	0.0036
	Statistical weight		0.118	0.074	0.053	0.146	0.004	0.022	0.241	0.124	0.021	0.197
ETH-2	**N of analyses**	**215**	23	13	11	51	4	4	18	51	8	32
	Δ_{47} (‰; 90 °C acid)	**0.2085**	0.2077	0.1840	0.2225	0.1978	0.1374	0.1650	0.2141	0.1968	0.2172	0.2170
	± 1SE	**0.0015**	0.0047	0.0070	0.0046	0.0050	0.0233	0.0101	0.0029	0.0043	0.0154	0.0033
	Statistical weight		0.105	0.047	0.108	0.092	0.004	0.023	0.272	0.125	0.010	0.213
ETH-3	**N of analyses**	**264**	55	15	20	54	4	5	15	59	8	29
	Δ_{47} (‰; 90 °C acid)	**0.6132**	0.6156	0.5975	0.6169	0.6102	0.5950	0.6143	0.6159	0.6094	0.6428	0.6124
	± 1SE	**0.0014**	0.0037	0.0056	0.0033	0.0038	0.0237	0.0099	0.0033	0.0042	0.0103	0.0035
	Statistical weight		0.140	0.062	0.175	0.134	0.003	0.020	0.179	0.110	0.018	0.158
ETH-4	**N of analyses**	**162**	10	12	5	55	4	4	12	47	7	6
	Δ_{47} (‰; 90 °C acid)	**0.4505**	0.4438	0.4230	0.4624	0.4506	0.4230	0.4454	0.4560	0.4414	0.4831	0.4646
	± 1SE	**0.0018**	0.0058	0.0071	0.0068	0.0049	0.0226	0.0095	0.0032	0.0042	0.0161	0.0057
	Statistical weight		0.093	0.064	0.068	0.133	0.006	0.035	0.314	0.177	0.012	0.097

Table 1. Newly determined nominal Δ_{47} values of the ETH standards projected to 90 °C acid reaction using acid correction factors of -0.088 ‰ and -0.022 ‰ for 25 °C and 70 °C reactions, respectively (Petersen et al. 2019). Reported standard errors represent analytical uncertainties associated both with reference frame errors (HG/EG) and carbonate sample reproducibility (Daëron, 2021).

When compared with Bernasconi et al. (2018), the average Δ_{47} values ETH-1 and ETH-2, projected back to 25 °C (+0.088 ‰), are respectively 0.035 and 0.040 ‰ more positive than the original values, whereas ETH-3 increases by 0.010 and ETH-4 by 0.031 ‰. A similar positive offset of Δ_{47} compared to the values reported in Bernasconi et al. (2018) has also been reported in Fiebig et al. (2019), Bajnai et al. (2020) and Thaler et al. (2020).

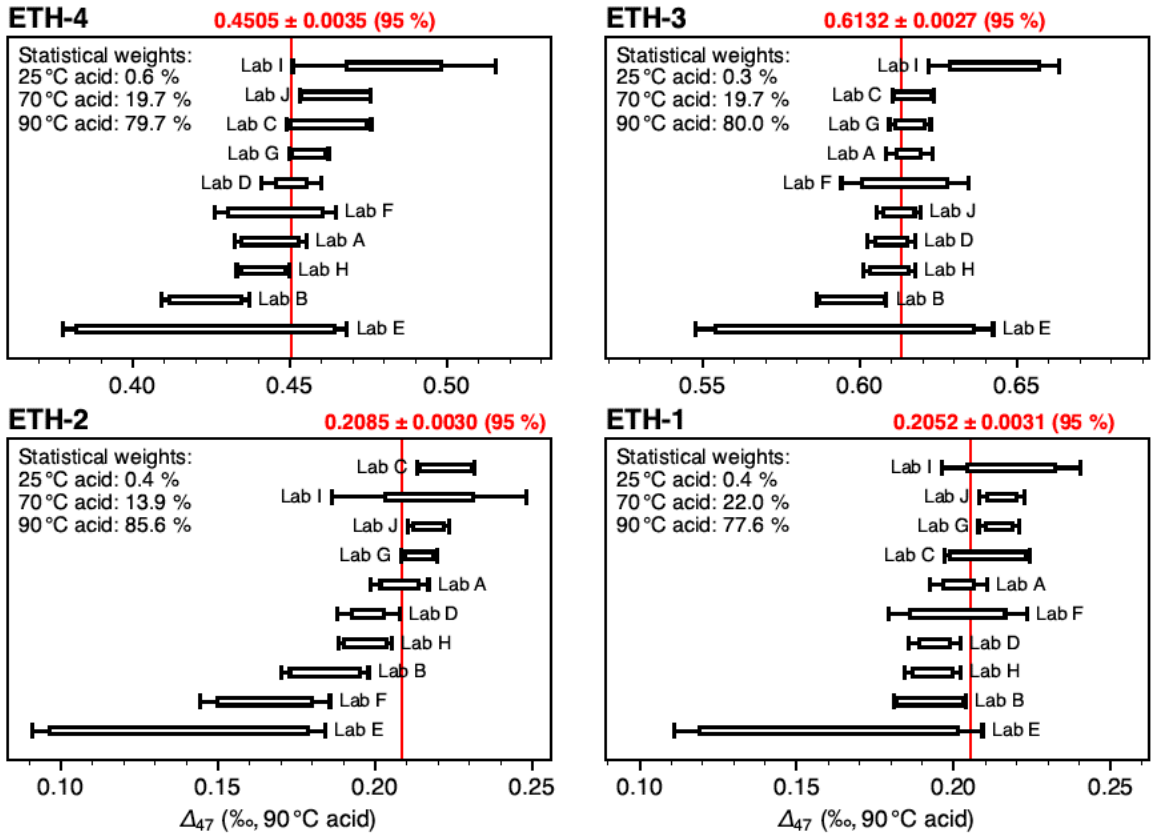


Figure 2. New determination of Δ_{47} values for the four ETH standards relative to the CDES using updated CO_2 equilibrium values. These measurements, using acid reaction temperatures of 90 °C, 70 °C or 25 °C, are projected to 90 °C using acid corrections of -0.088 ‰ and -0.022 ‰ for 25 °C and 70 °C reactions, respectively (Petersen et al. 2019). Error bars correspond to 95 % confidence limits taking into account fully propagated errors (i.e. taking into account errors in both unknown and anchor analyses). Boxes correspond to 95 % confidence limits not accounting for normalization errors (i.e. only taking into account errors in unknown analyses). Red numbers are the error-weighted average values (with statistical weights summarized in upper-left corners). All plots have the same horizontal scales for the different samples.

The observation that these changes in nominal values decrease as Δ_{47} increases suggests a simple hypothesis to explain this discrepancy: in the original study of Meckler et al. (2014), the carbonate samples and the heated/equilibrated CO_2 gases experienced different analytical procedures. The HGs were measured as large samples at constant beam intensity through a different capillary than the carbonates, which were measured using the microvolume and a decreasing beam. The potential effects of partial re-equilibration for the heated gases in the gas preparation line or in the capillaries

of the mass spectrometer could be significant whereas it would be minuscule for the gases equilibrated at 25°, leading to an overestimation of Δ_{47} scale compression and thus of the stretching applied to the Δ_{47} scale towards theoretical values. The observed changes in apparent ETH-1 and ETH-2 Δ_{47} values may therefore simply reflect partial re-equilibration of heated gases at the time of measurements at ETH (and reported in Meckler et al.. 2014), increasing their values in the original study by about 0.05 ‰ (Fig. 3).

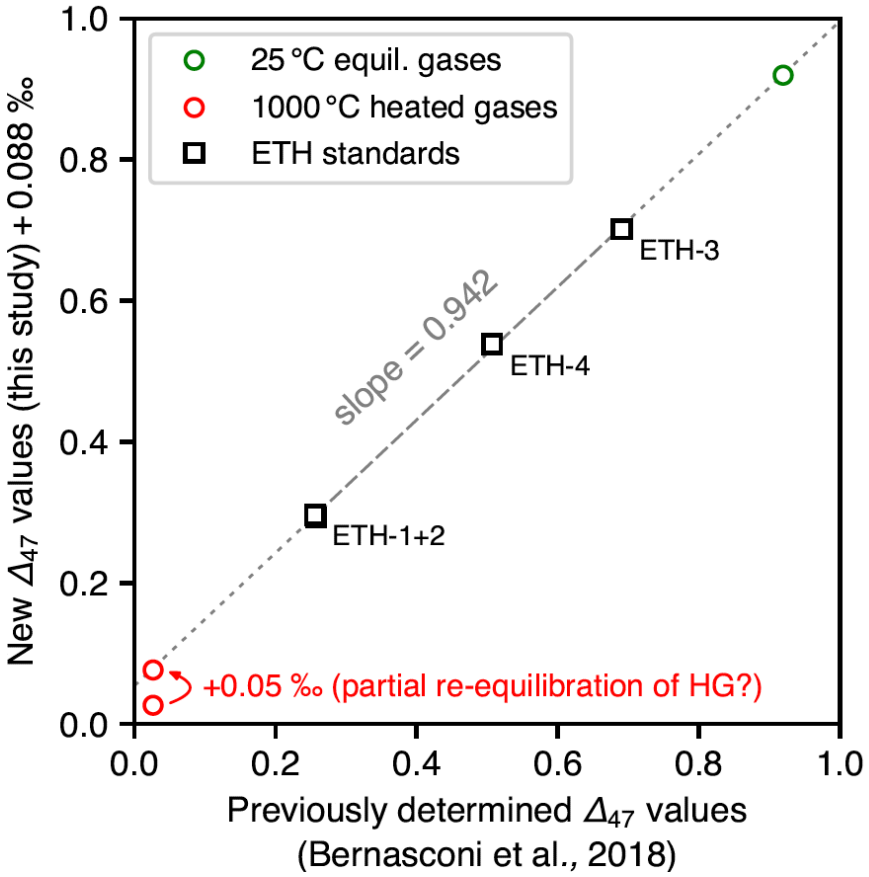


Figure 3. New nominal Δ_{47} values for the ETH standards compared to previously reported ones. The dashed gray line is a linear regression through the new versus old values of ETH-1/2/3/4, whose extrapolation coincides with 25 °C equilibrated CO_2 but not with heated gases. Apparent changes in the ETH-1/2/3/4 values thus scale linearly with the Δ_{47} difference between carbonate samples and 25 °C equilibrated CO_2, suggesting that Δ_{47} values of heated gases in the original study may have been biased by ~ +0.05 ‰ through partial re-equilibration at room temperature between the quenching of heated CO_2 and its ionization in the isotope-ratio mass spectrometer source.

It has been suggested previously that ETH-1 and ETH-2 should be indistinguishable in Δ_{47} and close to stochastic distribution (Müller, Violay, et al., 2017). This is because Δ_{47} values of ETH-1 and ETH-2, originally heated to 600 °C, were found to be higher by only around 0.006 ‰ from the same carbonates heated at 1000 °C to achieve stochastic distribution of the isotopes. However, additional test measurements in multiple laboratories of samples heated at >1000 °C are necessary to confirm this observation.

One laboratory (Laboratory F) did however observe a large difference in the value for ETH-1 and ETH-2, although their values of ETH-3 and ETH-4 are similar to other laboratories. The reason for these inconsistencies is probably due to the fact that ETH-1 was only measured four times with a limited number of HG/EG, and ETH-2 and ETH-4 were not measured in the same session. In addition, the laboratories with the smallest number of replicate measurements have uncertainties that are systematically larger (Table 1). These results highlight the importance of strict correction procedures in clumped isotope analysis. Sufficient replication of both standards and samples is critical and, if insufficient, offsets can arise when comparing results from different sessions. Due to these difficulties it is good practice to spread replicates of the same sample in different sessions over longer periods of time to obtain accurate results and follow a ~50:50 standard to sample replicate ratio.

Based on the results above, the difference between the average of ETH1/2 and ETH-3 is reduced by 0.028 ‰, thus leading to a compression of the scale by about 5.8 % compared to the values reported by Bernasconi et al. (2018). As a consequence, the slopes of published temperature calibrations produced with carbonate standardization (Kele et al, 2015; Bernasconi et al., 2018; Jautzy et al., 2021; Meinicke et al., 2020; Peral et al., 2018; Piasecki et al., 2019) will become slightly shallower, with more positive y-intercepts. If Δ_{47} results from previous publications are also recalculated with the new standard values (see section 3.4), however, changes in calculated formation temperatures will be negligible. For this reason, when comparing data from publications using old accepted values of the ETH standards for standardization (either those published by Meckler et al. (2014) or those recalculated with the IUPAC parameters by Bernasconi et al. (2018) to newer data, it is recommended to directly compare the reconstructed temperatures rather than recalculating Δ_{47}. Full recalculation of old measurements usually requires the availability of the

entire dataset including standards and the same correction procedures (e.g., averaging methods) used in the original publications (but see Appendix A for an alternative calculation method).

3.2 InterCarb results

Results for the unknown carbonate samples were obtained from 25 mass spectrometers in 22 laboratories. The Δ_{47} values of the 4 unknown samples were normalized to the new community-derived values of ETH-1, ETH-2, and ETH-3 of Table 1, then averaged per individual analytical session and mass spectrometer (Tables 2 and 3). Mean Δ_{47} values obtained for each sample in each mass spectrometer are shown in Figure 4. The details of each analytical session, including the number of samples and standards measured, the isotopic composition of the working standard, the scaling parameters and the internal reproducibilities (as 1SD) of the individual sessions are listed in Table 2. Some laboratories reported data for only a subset of the unknown samples, and both replication level and analytical reproducibility vary greatly from laboratory to laboratory (Table 2).

To clearly distinguish Δ_{47} values normalized to the carbon dioxide reference scale (CDES) using carbonates rather than heated and equilibrated gases, we propose the new acronym (I-CDES), short for InterCarb-CDES, to reflect the use of the proposed InterCarb reference materials for data standardization (see section 3.5 for more details).

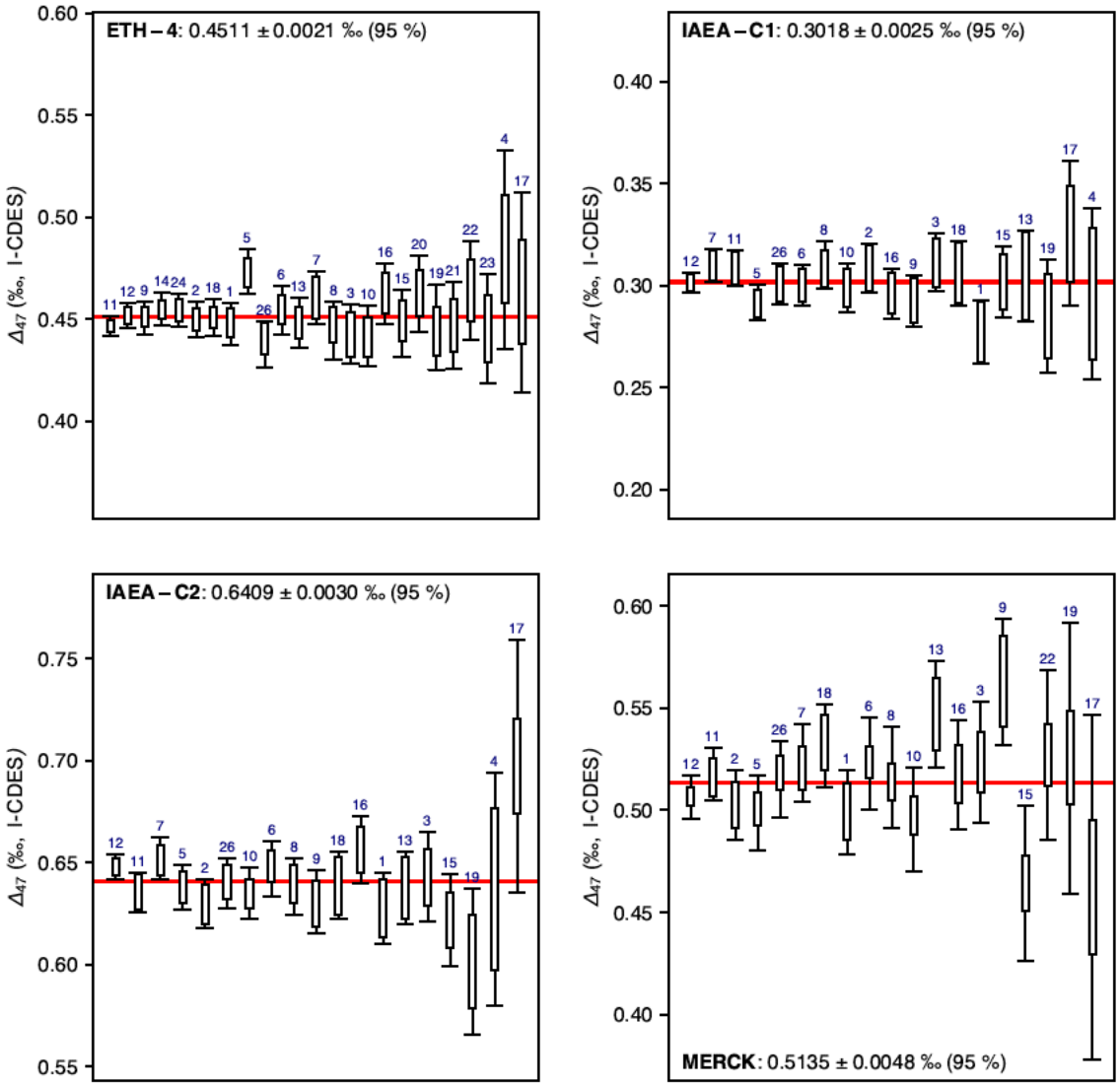


Figure 4. Final InterCarb results by laboratory. Error bars correspond to fully propagated 95 % confidence limits, taking into account errors in both unknown and anchor analyses. Boxes correspond to 95 % confidence limits not accounting for normalization errors (i.e., only taking into account errors in unknown analyses). Results are sorted by increasing analytical errors, and laboratories are identified by number. Overall error weighted average Δ_{47} values are displayed as solid red lines and reported in each panel. All plots have the same vertical scale.

The laboratory averages for the four unknowns show standard deviations of 0.011 ‰ for ETH-4 and IAEA-C1, 0.018 ‰ for IAEA-C2 and 0.024 ‰ for MERCK, the most extreme case of extrapolation (Table 3). Qualitatively, laboratories with stronger analytical constraints (i.e. better

intra-laboratory repeatability of Δ_{47} measurements and/or greater number of analyses) generally converge towards the overall mean value for each sample (Fig. 4). This suggests that the observed inter-laboratory variability is largely due to random errors that can be alleviated by replication, even for laboratories with relatively large analytical errors on individual measurements. It is also notable that fully propagated analytical errors that take into account uncertainties in the standardization procedure can be substantially larger than the errors based on the uncertainty associated with sample analyses alone, which is what is generally reported in the literature. The increase in error is also related to intra-laboratory repeatability and the number of standards measured. In addition, the error increases for unknown samples whose compositions lie outside the "anchor triangle" defined by ETH-1/2/3. This is illustrated by the increased scatter and errors associated with MERCK, the carbonate farthest from the "anchor triangle", consistent with the models of Daëron (2021, see also Kocken et al., 2019).

As seen in Table 2, there are stark differences in the total number of replicate analyses and the typical Δ_{47} reproducibility achieved in different laboratories. As a result, final uncertainties in the average Δ_{47} values of unknown samples vary considerably (Fig. 4). Inter-laboratory variability is smaller among laboratories with small analytical uncertainties, and larger among laboratories with few replicate analyses and/or poor analytical repeatability. If we chose only laboratories that have provided data with average standard errors below 0.01 ‰ (Table 3), which is within the shot-noise limits of modern IRMS instruments, inter-laboratory standard deviation (1SD) becomes $\leq$ 9 ppm for ETH-4 (N=22), IAEA-C1(N=15) and IAEA-C2 (N=13) and $\leq$ 0.015 ‰ for MERCK (N=11; with SE<0.0135). We note that this does not significantly change the average value of the unknowns, and highlights the importance of sufficient sample replication to obtain accurate results.

		Number of analyses								Working gas		Standardization parameters			Reproducibility (ppm)		
Lab	Session	El	E2	E3	E4	Cl	CZ	M	Nf	$\delta^{13}C_{VPDB}$	$\delta^{18}O_{VSMOW}$	*a*	*b*	*c*	$\delta^{13}C_{VPDB}$	$\delta^{18}O_{VSMOW}$	Δ_{47}
01	01	16	17	10	7	0	0	0	46	-3.58	25.38	0.91	(6.0 x10^{-4})	- 0.893	41	91	31.5
	02	6	5	3	1	0	0	0	11	-3.52	25.58	0.89	-2.1x10^{-3}	- 0.765	34	64	22.8
	03	150	146	65	72	19	21	22	488	-3.63	25.22	0.98	(-2.9x10^{-5})	- 0.965	33	74	33.5
02	01	19	24	20	18	4	5	4	87	-36.89	8.76	0.99	-5.6x10^{-4}	- 0.955	17	92	13.0
	02	6	8	5	4	2	3	2	23	-36 .88	8.83	0.98	(-5.5 x10^{-4})	- 0.931	25	77	16.1
03	01	37	24	17	9	0	0	0	83	-10.44	31.64	0.98	(-1.6 x10^{-4})	- 0.917	22	56	27.9
	02	29	32	12	14	17	13	11	121	-3.65	25.28	1.00	(-1.7 x10^{-4})	- 0.917	46	93	25.2
04	01	6	9		9	6	4		35	-6.57	27.18	0.97	5 0 x10^{-3}	-1.022	259	562	40.6
05	01	3	3	5	2	3	2	2	13	-10 .43	31.31	0.95	l. 7 x10^{-3}	- 0.970	15	27	8.6
	02	13	13	13	12	10	11	8	73	-3.62	25.05	0.99	(3.8x10^{-4})	- 0.968	15	24	20.9
	03	7	10	10	8	5	4	4	41	-3.63	25.06	0.90	1.1 x10^{-3}	- 0.901	42	113	17.3
06	01	6	3	5	3	3	3	3	19	-2.95	25.52	0.83	(-3.8 x10^{-4})	-0.920	22	25	21.0
	02	6	6	6	6	0	0	0	20	-2.98	24.93	0.92	(-9.9 x10^{-5})	- 0.920	14	71	13.3
	03	3	3	3	3	3	3	3	14	-3.01	24.90	0.88	(3.6 x10^{-4})	-0.932	10	43	9.4
	04	6	6	6	0	6	6	6	30	-2.95	25.28	0.90	(-1.4 x10^{-4})	- 0.926	18	61	17.3
07	01		4	4		4	4		19	-11.64	35.75	0.87	3.5 x10^{-3})	- 0.836	91	303	23.9
08	01	5	6	9	4	4	4	4	29	-2.68	25.86	0.94	(-9.2×10^{-4})	- 0.686	13	25	28.4
	02	5	4	14	6	4	5	4	35	-2.64	25.96	0.94	(8.6×10^{-4})	- 0.741	83	88	33.2
	03	4	4	13	4	3	5	6	32	-2.64	25.91	0.93	(-1.7×10^{-4})	-0.728	15	33	33.2
	04	4	5	9	5	4	4	4	28	-2.67	25.85	0.85	(1.3×10^{-4})	-0.629	17	51	44.5
	OS	3	6	8	4	4	4	4	26	-2 .70	25.79	0.87	(1.3×10^{-3})	-0.660	16	56	43.3
	06	4	4	16	6	6	6	4	39	-2.63	25.90	0.92	(3.9×10^{-4})	- 0 .693	85	54	37.8
	07	3	4	16	6	6	4	6	38	-2.66	25.90	0.96	(-1.9×10^{-3})	-0.709	19	52	48.8
	08	4	4	16	4	4	4	4	33	-2.66	25.89	1.03	(3.9×10^{-5})	- 0.806	12	46	42.7
	09	5	6	8	4	4	3	4	27	-2.67	25.84	0.92	(1.6×10^{-4})	- 0.722	19	25	46.7
	10	6	6	6	4	4	2	4	25	-2.63	25.91	0.97	(4.4×10^{-4})	- 0.767	36	39	40.5
	11	6	5	8	4	4	3	4	27	-2.67	25.87	0.97	(2.5×10^{-4})	- 0.760	11	31	49.5
	12	6	6	8	3	4	4	4	28	-2.66	25.86	1.02	(7.9 ×10^{-4})	-0.767	58	40	61.3
	13	4	6	8	6	4	4	6	31	-2.63	25.93	0.89	(1.3×10^{-3})	- 0.685	19	38	41.0
	14	5	7	5	4	4	4	4	26	-2.59	25.90	0.90	(-3.6×10^{-4})	-0.665	76	104	27.4
	15	6	4	8	4	4	4	4	27	-2.68	25.79	0.95	-2.0×10^{-3}	- 0.685	21	52	36.0
	16	2	2	10	5	4	2	4	22	-2.63	25.89	0.96	(-5.4×10^{-4})	-0.765	40	39	38.8
09	01	4	4	5	6	0	0	0	15	-3.60	25.36	0.89	3.8×10^{-3}	- 0.856	22	74	28.3
	02	26	19	16	24	0	0	0	81	-3.36	19.94	0.90	5.2×10^{-3}	-0.928	46	98	18.4
	03	21	17	13	19	0	1	0	66	-3.53	24.49	0.92	-l.0×10^{-2}	- 0.968	72	1667	22.4
	04	19	16	13	16	8	7	2	74	-3.60	25.27	0.98	-9.6×10^{-3}	-0.994	44	56	16.0
10	01	7	7	8	2	0	11	0	30	-7.43	32.38	0.98	l.9×10^{-3}	- 1.077	24	38	35.1
	02	15	15	21	15	11	20	11	101	-7.41	32.42	0.93	(-2.0×10^{-4})	-0.877	25	44	23.0
	03	17	18	25	9	22	31	20	135	-7.43	32.37	0.96	(-2.8×10^{-4})	- 0.900	31	92	30.0
11	01	24	24	28	28	0	0	0	100	-3.63	25.37	0.99	(-8.1×10^{-5})	-0.974	23	91	19.1
	02	20	18	15	15	0	0	0	64	-3.60	25.53	0.98	(3.5×10^{-4})	- 0.996	35	270	28.9
	03	69	62	74	66	13	13	8	298	-3.02	24.99	0.91	(-2.2×10^{-4})	- 1.065	34	89	25.0
	04	36	34	34	35	6	4	8	150	-3.01	25.08	1.00	(-3.1 ×10^{-4})	- 1.088	87	210	33.7
	OS	90	83	92	78	1 2	10	9	367	- 2.76	25.78	0.98	(-5.0×10^{-4})	- 1.088	97	317	19.3
12	01	7	7	9	5	5	6	5	37	-3.75	25.15	0.89	3.7×10^{-3}	-0.904	7	41	10.2
	02	7	6	6	6	5	5	5	33	-3.74	25.18	0.87	4.6×10^{-3}	- 0.897	8	so	9.3
	03	8	7	12	5	5	5	5	40	-3.74	25.17	0.88	5.5×10^{-3}	-0.909	9	51	9.7
	04	6	7	6	5	5	5	4	31	-3.74	25.17	0.88	5.3×10^{-3}	- 0.908	7	51	8.7
13	01	58	51	59	47	6	12	9	235	-10.29	33 .18	0.98	-3.7×10^{-4}	-0.993	176	239	26.8
14	01	4	7	10	10	0	0	0	27	-3.63	24.95	0.93	(1.3×10^{-4})	- 0.972	42	159	19.3
	02	10	11	8	7	0	0	0	32	-3.61	25.04	0.97	(5.8×10^{-4})	- 1.021	40	128	30.0
	03	6	4	4	3	0	0	0	13	-10.38	31.93	0.84	-1.7×10^{-3}	- 0.747	39	59	20.5
	04	2	2	2	2	0	0	0	4	-10 .40	31.92	0.86	-l.0×10^{-3}	- 0.794	20	29	9.2
	05	4	4	3	4	0	0	0	11	-10 .40	31.92	0.91	-l. 6×10^{-3} 3	- 0.807	27	60	11.0
	06	5	6	6	7	0	0	0	20	-10.43	31.84	0.99	(1.3×10^{-4})	- 0.908	39	53	22.4
	07	3	5	2	1	0	0	0	7	-10 .41	31.85	0.97	(-1.7×10^{-4})	- 0.877	51	43	12.8
	08	11	7	3	5	0	0	0	22	-10.47	31.66	0.94	-7.8×10^{-4}	- 0.920	61	84	23.4
	09	4	2	3	4	0	0	0	9	-10 .43	31.82	0.95	(-4.8×10^{-4})	- 0.907	55	83	12.0
	10	4	4	1	3	0	0	0	8	-10.49	31.73	0.99	(1.7×10^{-4})	- 0.926	40	71	13.3
15	01	4	4	4	4	0	0	0	12	-32 .89	36 .92	0.96	-2.5×10^{-3}	- 0.887	87	70	14.6
	02	4	4	4	4	4	4	4	21	-3.72	24.98	1.02	4.6×10^{-3}	- 1.027	59	41	14.0
16	01		6	6	4	4			23	-10.49	31.56	0.99	-4.1×10^{-3}	- 0.979	47	109	10.1
17	01		5			6	6		23	-9.73	23.81	0.81	(6.3×10^{-4})	- 0.940	65	204	29.3
18	01	168	147	172	169	20	20	25	714	-3.45	25.25	0.81	(1.5×10^{-4})	- 0 .722	65	11 0	37.7
	0 2	1 7	14	1 7	13	4	4	4	66	- 3 .4 1	25 .4 2	0.83	(1.6×10^{-5})	- 0.761	21	52	45.7
	03	11	12	13	14	2	4	2	51	-3.52	25.12	0.96	(6.0×10^{-4}) 1	- 0.835	23	45	40.5
19	01	4	4	5	7	5	4	4	26	-24.48	25.66	0.99	(2.0×10^{-4})	- 0.970	69	193	23.4
	02	7	8	10	7	0	0	0	28	5.03	38.66	0.99	(2.0×10^{-4})	- 0.962	164	416	22.5
20	01	9	6	6	6	0	0	0	23	-3.63	28.89	0.93	-2.1×10^{-3}	- 0.921	11	s o	14.3
21	01					0	0	0	8	-3.62	25.20	0.90	l. 0×10^{-3}	- 0.886	65	139	11.4
22	01	8	8	8		0	0		33	-3.54	25.37	0.98	9.9×10^{-3}	- 0.951	155	443	20.5
23	01	6	6	6	6	0	0		20	-10 .77	31.02	1.00	4.4×10^{-3}	- 0.948	47	91	20.5
24	01	19	18	15	12	0	0	0	60	-4.40	25.32	0.98	(2.1×10^{-4})	- 0.955	42	107	9.9
26	01	4	4	4		3			19	-40.04	5.51	0.89	(2.2×10^{-4})	- 0.998	96	14 5	15.0
	02	6	7	6		3			24	-40 .03	5.40	0.92	(-1.1 ×10^{-4})	- 1.014	50	88	8.7

Table 2. Summary of all InterCarb analyses. Nf is the number of degrees of freedom when estimating pooled analytical repeatabilities and standardization model uncertainties. Standardization parameters a, b and c refer to the scrambling factor in the source, the compositional slope due to positive or negative backgrounds in the collectors and the working gas offset, respectively (see section 2.4 and Daëron, 2021). Values of standardization parameter *b* which are statistically indistinguishable from zero at 95 % confidence level are reported in parenthesis. Reproducibility is reported as 1 SD

	ETH-4		IAEA-C1		IAEA-C2		MERCK	
MS	**$\Delta_{47(\text{I-CDES})}$ (‰ ± 1SE)**	**N**	**$\Delta_{47(\text{I-CDES})}$ (‰ ± 1SE)**	**N**	**$\Delta_{47(\text{I-CDES})}$ (‰ ± 1SE)**	**N**	**$\Delta_{47(\text{I-CDES})}$ (‰ ± 1SE)**	**N**
1	0.4477 ± 0.0052	80	0.2773 ± 0.0080	19	0.6275 ± 0.0088	21	0.4991 ± 0.0105	22
2	0.4499 ± 0.0044	22	0.3086 ± 0.0060	6	0.6299 ± 0.0061	8	0.5025 ± 0.0089	6
3	0.4430 ± 0.0074	23	0.3114 ± 0.0073	17	0.6427 ± 0.0112	13	0.5235 ± 0.0152	11
4	0.4841 ± 0.0248	9	0.2959 ± 0.0215	6	0.6368 ± 0.0291	4		
5	0.4734 ± 0.0055	22	0.2916 ± 0.0044	18	0.6378 ± 0.0057	17	0.4987 ± 0.0094	14
6	0.4545 ± 0.0060	12	0.3004 ± 0.0051	12	0.6471 ± 0.0069	12	0.5229 ± 0.0116	12
7	0.4607 ± 0.0066	8	0.3099 ± 0.0042	16	0.6520 ± 0.0052	15	0.5231 ± 0.0098	8
8	0.4442 ± 0.0072	73	0.3099 ± 0.0060	67	0.6383 ± 0.0071	62	0.5159 ± 0.0127	70
9	0.4505 ± 0.0041	65	0.2926 ± 0.0064	8	0.6309 ± 0.0078	8	0.5630 ± 0.0158	2
10	0.4416 ± 0.0075	26	0.2987 ± 0.0060	33	0.6348 ± 0.0065	62	0.4954 ± 0.0130	31
11	0.4468 ± 0.0025	222	0.3085 ± 0.0043	31	0.6354 ± 0.0050	27	0.5175 ± 0.0066	25
12	0.4521 ± 0.0032	21	0.3015 ± 0.0026	20	0.6479 ± 0.0032	21	0.5064 ± 0.0054	19
13	0.4484 ± 0.0062	47	0.3048 ± 0.0113	6	0.6376 ± 0.0091	12	0.5470 ± 0.0135	9
14	0.4548 ± 0.0041	46						
15	0.4480 ± 0.0083	8	0.3016 ± 0.0090	4	0.6217 ± 0.0116	4	0.4642 ± 0.0195	4
16	0.4627 ± 0.0076	4	0.2962 ± 0.0063	4	0.6563 ± 0.0084	3	0.5176 ± 0.0136	2
17	0.4634 ± 0.0250	5	0.3254 ± 0.0181	6	0.6971 ± 0.0314	6	0.4623 ± 0.0429	3
18	0.4510 ± 0.0046	196	0.3060 ± 0.0079	26	0.6386 ± 0.0084	28	0.5317 ± 0.0104	31
19	0.4460 ± 0.0106	14	0.2851 ± 0.0142	5	0.6015 ± 0.0183	4	0.5256 ± 0.0339	4
20	0.4627 ± 0.0095	6						
21	0.4470 ± 0.0108	3						
22	0.4639 ± 0.0124	7					0.5269 ± 0.0213	7
23	0.4453 ± 0.0137	6						
24	0.4544 ± 0.0042	12						
26	0.4378 ± 0.0058	8	0.3008 ± 0.0051	6	0.6396 ± 0.0062	6	0.5152 ± 0.0095	6
w. avg	**0.4511 ± 0.0011**	**945**	**0.3018 ± 0.0013**	**310**	**0.6409 ± 0.0016**	**333**	**0.5135 ± 0.0024**	**286**
SD	**0.011**		**0.011**		**0.018**		**0.024**	

Table 3. Average Δ_{47} values (±1SE, fully propagated uncertainties) obtained by each mass spectrometer from the 22 laboratories. Note the larger standard deviation for the samples further from the calibration triangle defined by the anchors. The average Δ_{47} values for individual analytical sessions are reported in table 2.

Next we may assess whether inter-laboratory discrepancies are significantly larger than expected from intra-laboratory analytical uncertainties, i.e., whether we can detect the effects of hypothetical unrecognized sources of scatter beyond known analytical errors.

In order to do so, we compute the "number-of-sigma" deviation obtained by each laboratory for each unknown sample, relative to that sample's overall weighted average value. For example, the sigma-deviation for sample ETH-4 and Lab01 is equal to (0.4477 – 0.4511) / 0.0052 = –0.66 and that for MERCK and Lab13 is equal to (0.5470 – 0.5135) / 0.0135 = +2.48. If the analytical errors reported in Table 3 are reasonably accurate, we expect the population of sigma-deviations among all laboratories to be distributed as the canonical Gaussian distribution ($\mu = 0$; $\sigma = 1$), and we can test this prediction using established statistical methods such as a Kolmogorov–Smirnov test of normality (Massey, 1951). We carried out this test for two cases: only considering the error of sample replication (Fig. 5, upper row) and secondly including the normalization error, (i.e. the fully propagated error (Fig. 5, lower row). If we neglect uncertainties arising from standardization (the "allogenic" errors of Daëron, 2021), the sigma-deviations are no longer normally distributed ($p = 0.003$, Fig. 5 upper-left panel). When considering fully propagated analytical errors, as shown in the lower-left panel of Figure 5, the distribution of sigma-deviations for all laboratories and all samples is statistically indistinguishable from the expected normal distribution ($p = 0.19$). Figure 5 also illustrates that neglecting standardization errors does not strongly affect the normality of sigma-deviations for IAEA-C1, which has δ^{47} and Δ_{47} values within the range covered by the three anchor samples. By contrast, sigma-deviations for unknowns with "exotic" isotopic compositions (ETH-4 and IAEA-C2 but especially MERCK) are only normally distributed if standardization uncertainties are correctly accounted for.

Based on these tests, we conclude that the inter-laboratory scatter observed in the InterCarb data set is neither smaller nor larger than expected from the analytical uncertainties computed within each laboratory, *as long as standardization errors are taken into account*. This important finding implies that, at least for the time being, we can rule out any systematic inter-laboratory discrepancies in carbonate-standardized Δ_{47} measurements, which constitutes an important milestone in the progress of clumped isotope measurement techniques.

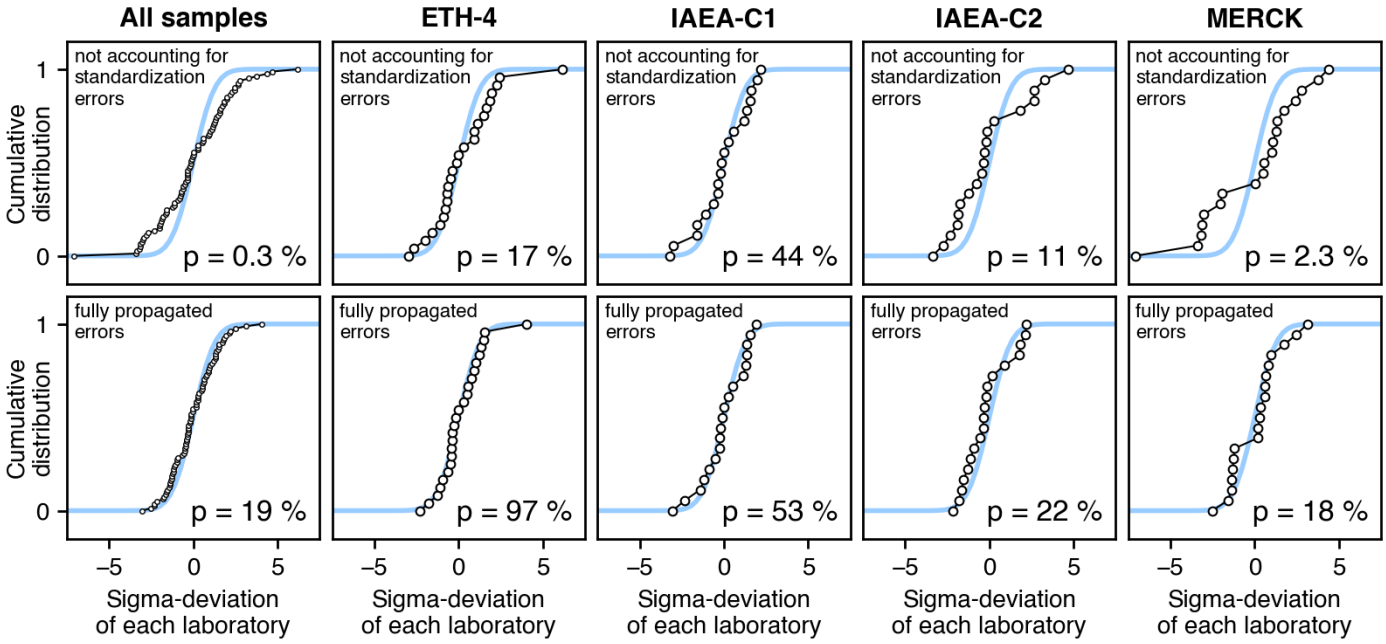


Figure 5. Kolmogorov–Smirnov tests of normality for the sigma-deviations obtained in each laboratory participating in the Intercarb effort (circular markers), either neglecting standardization uncertainty (upper row) or considering fully propagated analytical errors (lower row). Lower-right corner Kolmogorov–Smirnov p-values correspond to the null hypothesis that the sigma-deviations are normally distributed with a mean of zero and a standard deviation of 1. Blue lines correspond to the canonical Gaussian distribution ($\mu = 0$; $\sigma = 1$).

On demonstrating that we can fully account for inter-laboratory error using carbonate standardization, we revisit the results obtained for ETH1-4 using HG and EG (Fig. 2). Applying the same Kolmogorov–Smirnov test of normality yields unambiguous evidence that the inter-laboratory scatter observed here, using HG/EG standardization, is significantly greater than predicted from known analytical errors alone ($p = 0.005$, Fig. 6), contrary to the opposite finding for the InterCarb results, using carbonate-based standardization ($p = 0.19$, Fig. 5).

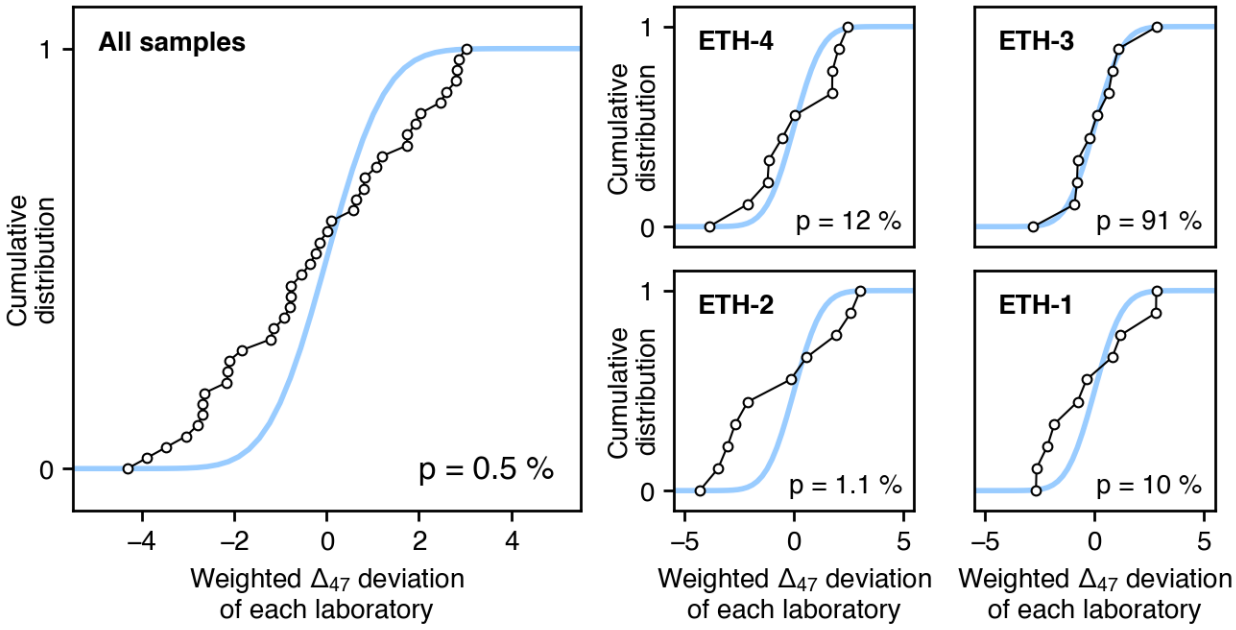


Figure 6. Kolmogorov–Smirnov tests of normality for the sigma-deviations, considering fully propagated analytical errors (accounting for uncertainties associated with conversion to the CDES reference frame), obtained in each laboratory participating in the ETH-1/2/3/4 determination using HG and EG (circular markers). Lower-right corner Kolmogorov–Smirnov p-values correspond to the null hypothesis that the sigma-deviations are normally distributed with a mean of zero and a standard deviation of 1. Blue lines correspond to the canonical Gaussian distribution ($\mu = 0$; $\sigma = 1$).

3.3 Effects of acid reaction temperature and IRMS models

Out of 25 sample preparation systems, 10 convert samples to CO_2 by acid reactions at 70 °C and 14 at 90 °C. To test for the possible effect of acid temperature, a commonly discussed cause for different slopes in the published temperature calibration curves (Came et al., 2014; Fernandez et al., 2014; Swart et al., 2019), we plot the Δ_{47} values of unknowns obtained by laboratories reacting at 70 °C vs. those obtained at 90 °C (Fig. 7, Table 4). Because acid fractionation effects equally affect anchors and unknowns, carbonate-standardized results can be compared directly without acid temperature correction. Δ_{47} values averaged by acid temperature are statistically indistinguishable for all of the unknowns. This implies that relative Δ_{47} differences between CO_2

evolved from different samples are independent of acid reaction temperature within the range of experimental conditions covered here, and for a very wide range of $\Delta_{47(\text{I-CDES})}$ values spanning 0.302 ‰ (marbles) to 0.641 ‰ (carbonates formed at ambient temperatures).

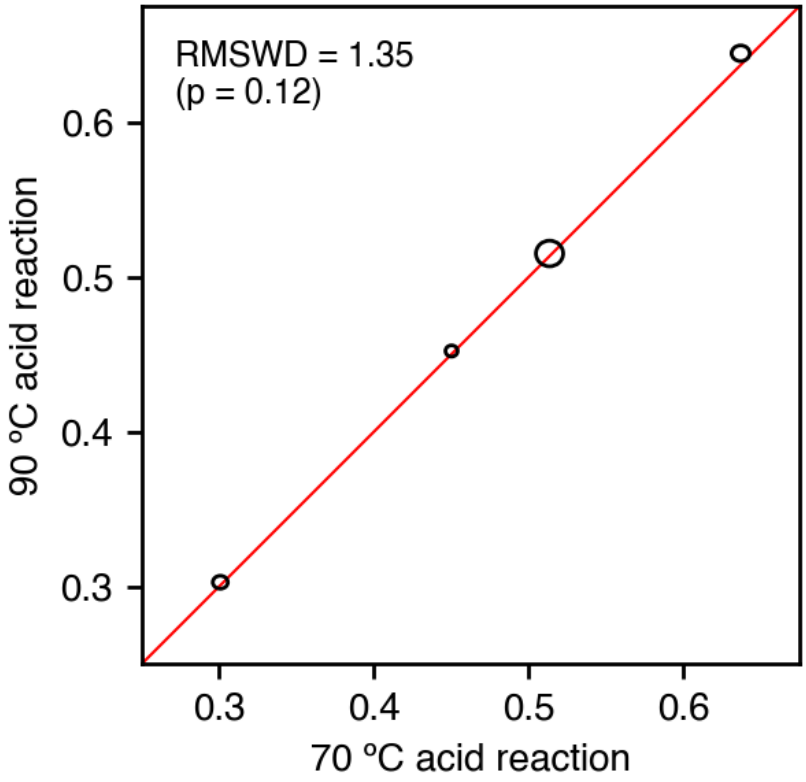


Figure 7. Error-weighted average Δ_{47}(I-CDES) values of unknowns obtained from acid reactions at 90 °C vs 70 °C. Solid black ellipses correspond to 95 % confidence limits. (see also Table 4).

	$\Delta_{47(\text{I-CDES})}$ **(70 °C reaction)**	$\Delta_{47(\text{I-CDES})}$ **(90 °C reaction)**	**Difference (±1SE)**
ETH-4	0.4501 ± 0.0016	0.4521 ± 0.0015	0.0020 ± 0.0022
IAEA-C1	0.3006 ± 0.0020	0.3026 ± 0.0017	0.0020 ± 0.0026
IAEA-C2	0.6369 ± 0.0024	0.6445 ± 0.0021	0.0076 ± 0.0032
MERCK	0.5134 ± 0.0036	0.5151 ± 0.0034	0.0017 ± 0.0049
average (all samples)			0.0033 ± 0.0017

Table 4. Error-weighted average $\Delta_{47(\text{I-CDES})}$ values (‰; ±1SE) for each unknown as a function of acid reaction temperature (see also Fig. 7).

The error-weighted results separated by mass spectrometer type and design of associated preparation lines, another postulated source of inter-laboratory disagreement (Swart et al., 2019), are shown in Figure 8 and Table 5. Out of the 22 participating laboratories, 8 use the Nu Perspective, 16 use versions of the Thermo MAT253, and one uses an Isoprime 100. Most results are statistically indistinguishable across instruments. Only IAEA-C2 yielded a significantly (>2σ) higher mean Δ_{47} value when measured on the Isoprime 100 ($\Delta\Delta_{47}$ of +0.0110 and +0.0081 ‰ vs Nu perspective and MAT253, respectively); but note that all of the Isoprime 100 data comes from a single laboratory. Inter-instrument differences averaged over all four samples (bottom row of Table 4) remain, however, indistinguishable from zero. Thus, any potential biases introduced by the use of different mass spectrometer models and/or the design of the preparation line which could cause partial equilibration of the produced CO_2 with the acid and/or heated metal surfaces (Swart et al., 2019) are undetectable when using carbonate standardization. Sample sizes used for individual measurements ranged from 90-120 µg for the Kiel IV to ~500 µg for the NuCarb individual acid vial preparation systems, and to 3-12 mg for samples reacted in common acid bath custom-built extraction lines. The fact that small sample measurements are carried out at 70 °C and large ones at 90 °C, also suggests that there is no significant effect of sample sizes and variations in sample to acid ratios in these results.

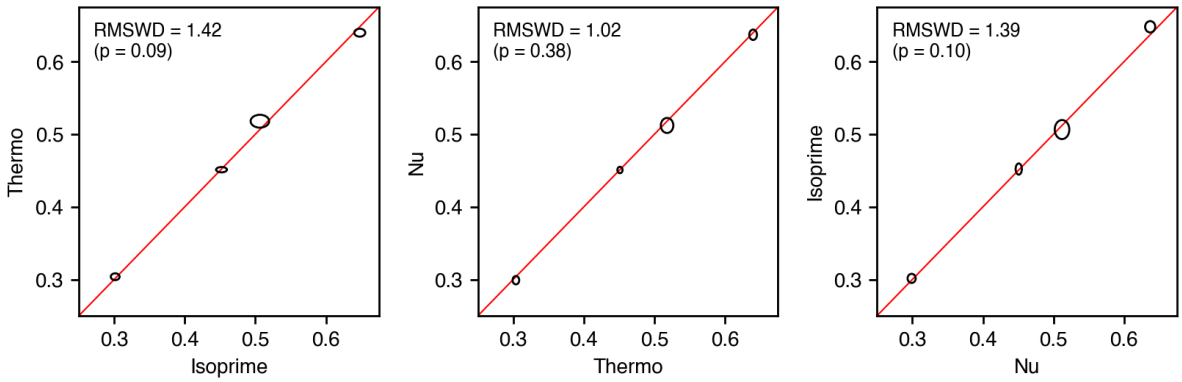


Figure 8. Error-weighted average $\Delta_{47(\text{I-CDES})}$ values of unknowns obtained using different mass spectrometer types. Solid black ellipses correspond to 95 % confidence limits.

	MAT 253 vs Isoprime 100	Nu Perspective vs MAT 253	Isoprime 100 vs Nu Perspective
ETH-4	-0.0009 ± 0.0035	-0.0004 ± 0.0024	0.0013 ± 0.0036
IAEA-C1	0.0023 ± 0.0032	-0.0048 ± 0.0030	0.0025 ± 0.0035
IAEA-C2	-0.0081 ± 0.0039	-0.0029 ± 0.0037	0.0110 ± 0.0043
MERCK	0.0115 ± 0.0065	-0.0059 ± 0.0056	-0.0056 ± 0.0068
average (all samples)	0.0012 ± 0.0022	-0.0035 ± 0.0019	0.0023 ± 0.0024

Table 5. Error-weighted average $\Delta_{47(i\text{-}CDES)}$ differences (±1SE) for each unknown as a function of mass spectrometer type.

3.4. Guidelines for minimizing uncertainties in clumped isotope analyses

The results of InterCarb strongly support the use of carbonate standardization for clumped isotope measurements and show that it is possible to reach excellent data quality and inter-laboratory consistency with instrumentation from all manufacturers and with both custom built and commercially available sample preparation systems.

When considering all laboratories, the standard deviation of the averages for the four unknowns range from 0.011 ‰ for ETH-4 to 0.024 ‰ for MERCK. The spread is still relatively large, and not significantly better than that obtained by HG-EG normalization if we consider either what has been reported on 4 carbonate standards for 4 laboratories in Dennis et al. (2011) or the scatter in the values reported by the 10 laboratories that provided data for the re-determination of the accepted values of ETH-1 to ETH-3 in this study. However, we can clearly state that the large scatter is dominated by random errors and is especially influenced by the laboratories with the largest errors in the individual sample reproducibility and a significantly magnified normalization error induced by a small number of replicates of anchors (Fig. 4). Based on these observations we present strategies to improve the repeatability within each laboratory.

If we consider only laboratories with reported errors < 0.010 ‰ (1SE) which is a desirable goal for the application of clumped isotopes in paleoclimate reconstructions, the standard deviation of the result is ≤ 0.009 ‰ for the samples with no or moderate extrapolation from the compositional triangle defined by the anchors. This correspond to uncertainties across laboratories of

approximately 3 °C at ambient temperatures. The laboratories with the smallest errors are those that generally analyzed a large number of samples and standards, as seen by the smaller 95 % CL errors and the small increase of the error when the normalisation error is included (Fig. 4). This observation underscores the necessity of sufficient replication to produce data of the quality that is required for meaningful interpretations. The number of necessary replicates to reach a target temperature unertainty can be reduced by improving the external reproducibility of the measurements (see also Daëron, 2021; Bonifacie et al., 2017; Fernandez et al., 2017; Kocken et al., 2019).

Spreading replicate sample measurements in time and over multiple analytical sessions should help avoid analytical biases. The number and distribution of standards in a measuring interval are also important parameters to improve reproducibility and reduce errors. This has been discussed in detail by Kocken et al. (2019) and Daeron (2021) who both concluded that carbonate standards with bulk and clumped-isotope compositions similar to those of unknowns should be analyzed with greater frequency than the other anchors, while preserving a minimal level of replication for each anchor. In this study, analyses were grouped in measurement intervals, and all data were processed assuming no short-term variation in the instrumentation. However, especially with "small sample approaches" (e.g. the Kiel device) relying on short (~30-45 min) measurements of many replicates, one can observe short-term variations (e.g. Bernasconi et al. 2018, Fig. 4). Thus a moving window correction with variable window size may be desirable in these cases, likely calling for specific error propagation procedures which remain yet to be defined.

Two important outcomes of this study are that acid reaction temperature and instrument and preparation line design are not a cause for differences among laboratories when standardization is based on carbonates. The lack of resolvable differences observed in our dataset indicates that if preparation line differences affect Δ_{47} measurements, standardization with carbonates corrects any such effects whereas HG-EG standardization may fail to do so (Swart et al. 2019). Thus, when using carbonate standardization, these factors can be ignored provided the carbonate standards cover a large range in Δ_{47}. Thus, following the principle of identical treatment of samples and standards (Werner & Brand, 2001b) clearly reduces uncertainties compared to the use of HG/EG standardization. We have to consider, however, that acid digestion conditions (e.g., reaction times, temperatures) and the temperature dependence of phosphoric acid fractionation (Defliese et al.

2015; Murray et al. 2016 ; van Djik et al. 2019) may differ with carbonate mineralogy, therefore possible effects on Δ_{47} could be mineral-specific. For this reason, it would be highly desirable to produce reference materials for dolomite, aragonite, magnesite and siderite. For dolomite, three samples were proposed by Müller et al. (2019) as possible reference materials and are available upon request. Ideally it would also be desirable to anchor the measurements in a carbonate absolute reference frame by creating some carbonates with independently known clumped isotope compositions.

For InterCarb, all distributed aliquots of IAEA-C1, IAEA-C2 and MERCK originated from single bottles. The IntrCarb results suggest that after milling, these carbonates were homogeneous within these bottles, but we strongly recommend verifying that additional bottles purchased from IAEA and MERCK are identical to the ones tested here. Nishida and Ishimura (2017) found that IAEA 603, which was produced from the same coarse marble as IAEA-C1, contained a minor mount of grains with different isotopic composition, thus it is recommended to thoroughly mill and homogenize IAEA-C1 before use, especially for use in small-sample preparation systems. Merk and ETH-4 are both very fine grained syntetic calcium carbonates, and it has been suggested that the oxygen isotope composition of such fine-grained carbonates could change with time due to exchange with atmospheric CO_2 (Qi et al. 2020). At the ETH laboratory different aliquots of ETH-4 have been in use since 2013 and no alterations of its oxygen or clumped isotope composition were observed. However we strongly recommend that all standards are stored in a dessiccator to reduce the chance of alteration.

Standardization errors could be reduced to some extent by increasing the range of bulk composition of the anchor samples (e.g., as illustrated by Fig. 1 of *Daëron*, 2021), especially when samples are measured that require significant extrapolation. A sample with an extreme bulk composition like MERCK would be a useful addition as an anchor, regardless of its Δ_{47} value. While with a three-anchor system, two heated standards for normalization are not strictly necessary, a “heated MERCK” anchor in combination with ETH-1 would furthermore allow verifying the PBL correction with greater confidence and with less replication than with ETH-2 (keeping in mind that small quadratic components to PBL correction might introduce a significant bias over a δ^{47} range of 60 ‰, e.g., Fig. 7 from He et al., [2012]).

With InterCarb, the nominal values of the ETH standards are robustly linked to the CDES, as they are now based on the average results of 10 laboratories, and are not only based on the values determined at ETH in 2013. Some laboratories may still want to continue measuring HG and EG to keep established laboratory procedures and/or to cover ranges in bulk compositions that require large extrapolations. However, the results of InterCarb, and the discussions in the literature (e.g. Petersen et al. 2019) show that with the HG/EG approach there are still poorly understood interlaboratory discrepancies (as suggested by Fig. 6) which are absent in the carbonate-based normalization (see Fig 5). For this reason, it is of paramount importance that several of the InterCarb reference carbonates are incorporated in the laboratory procedures to ensure interlaboratory data compatibility. The use of matrix-matched reference materials is necessary so that delta values can be unambiguously compared on a like-for-like basis (see Meier-Augenstein and Schimmelmann, 2019 for a recent discussion).

For laboratories using large sample common-acid bath methodologies and preferring gas-based data correction, we recommend that in addition to HG/EG a minimum of two of the InterCarb reference materials (or in-house standards with values calibrated to ETH standards) should be measured within the same analytical sessions as the samples and used in the data correction scheme. We recommend choosing two standards with a large difference in Δ_{47}, for example ETH-1 and ETH-3 or ETH-2 and IAEA-C2, depending on the bulk composition of the unknown samples (see Fig. 1). In the case of samples with very low δ^{47} values we additionally recommend the use of MERCK.

For laboratories with commercially available single acid vial, small sample preparation systems, we discourage the use of HG/EG, as the gases would most probably be measured under different conditions than the samples, an approach which is prone to error. In addition, InterCarb shows that carbonate-based standardization can provide robust and accurate data without the use of gases.

Achieving the best possible reproducibility and accuracy is especially important when reconstructing small temperature changes for the reconstruction of climate change and/or the study of high temperature processes where the sensitivity of the clumped isotope thermometer is low. We emphasize that the improvements in interlaboratory comparability that can be achieved with

carbonate standardisation, coupled with the reductions in the uncertainties of the temperature calibrations (e.g., Anderson et al. 2021), is decreasing the uncertainties in temperature reconstructions to levels comparable or better than other temperature proxies.

3.5 Reporting data normalized to carbonates: definition of the "InterCarb" Carbon Dioxide Equilibrium Scale (I-CDES)

There is a need for a community consensus on how to report clumped isotope measurements, both to promote data comparability and to reduce confusion stemming from the different scales used in the literature. Currently Δ_{47} data are reported for different temperatures of phosphoric acid digestion, mostly projected to 25 °C but also to 70 °C or 90 °C reactions, and generally labeled respectively as $\Delta_{47CDES25}$, $\Delta_{47CDES70}$ and $\Delta_{47CDES90}$, a terminology introduced in Bonifacie et al. (2017). In the literature, phosphoric acid correction factors used by different research groups to convert results from 90 to 25 °C reaction temperatures have varied between 0.069 ‰ (Wacker et al., 2014) to 0.092 ‰ (e.g., Bonifacie et al., 2017), thus representing a significant source of uncertainty and confusion.

The direct standardization to accepted values of solid phases, on the other hand, removes the need for a phosphoric acid correction, yielding results which are independent of the temperature at which the samples were reacted. In InterCarb this has only been tested for calcites, the mineralogy of all standards used here. Further studies are necessary in particular for dolomite and siderite, as these minerals require longer reaction times and there are contrasting findings in the literature on whether they require different phosphoric acid fractionations (Bonifacie et al., 2017; Müller et al. 2019, van Dijk et al., 2019; Petersen et al. 2019), possibly affecting the absolute values of samples with such mineralogies. Aragonite may also have a different phosphoric acid fractionation factor than calcite (Müller, Violay etl al. 2017). Nevertheless, the use of the InterCarb reference carbonates, allows for a consistent correction of instrumental effects and normalization to the I-CDES, independent of the mineralogy. Possible effects related to the longer reaction time remain to be tested, and require intercomparison samples of different mineralogy. In any case, for all

calcites, carbonate standardization eliminates different phosphoric acid correction factors as a source of uncertainty and provides a consistent framework to report Δ_{47} without uncertainties related to the reaction temperature.

For these reasons, we recommend that in the future, carbonate clumped isotope values should be reported relative to a carbonate reference frame uniquely defined by the absolute Δ_{47} values reported in Table 1 for ETH-1, ETH-2, and ETH-3. Unknown samples may be anchored to this reference frame either (a) directly by comparison to ETH-1/2/3, (b) indirectly by comparison with several of the four other carbonate standards used here (Figure 4), or (c) by comparison with a set of in-house laboratory standards whose composition is well-constrained relative to the materials reported here. To clearly distinguish this data normalization scheme from previous ones the denomination I-CDES (InterCarb - Carbon Dioxide Equilibration Scale) should be use, with the notation $\Delta_{47(\text{I-CDES})}$.. This approach is analogous to the change from the PDB to the VPDB scale, which was accomplished by assigning a consensus offset of +1.95 ‰ between the original PDB reference material and the NBS19 carbonate. This was subsequently, albeit temporarily, improved by defining a second anchor point with the L-SVEC lithium carbonate standard (Coplen et al., 2006). We note that because the carbonate Δ_{47} values in Table 1 are firmly anchored to the CDES scale via HG/EG measurements in multiple laboratories, the two scales are in principle equivalent. However, I-CDES has three major advantages: (1) it follows the principle of equal treatment of sample and standards, (2) it removes uncertainties related to fractionation effects due to different acid reaction temperatures and designs of the preparation lines and (3) it is based on traceable, stable materials (calcium carbonates) that are widely available to interested laboratories. Furthermore, the results summarized in Fig. 5 imply that I-CDES standardization yields consistent Δ_{47} values independent of laboratory and/or analytical protocols, so that inter-laboratory scatter is accurately predicted by the fully-propagated analytical uncertainties computed within each laboratory. By contrast, this does not always seem to be the case for gas-based standardization to the CDES reference frame (Fig. 6), suggesting the existence of poorly understood sources of inter-lab discrepancies in that approach. The broad availability of carbonate reference materials with widely varying bulk and clumped isotope compositions is an important step to help establish a worldwide equivalence among laboratories and help new laboratories establish and verify their analytical procedures.

3.6 Comparability with previously published data

The decision that $\Delta_{47(\text{I-CDES})}$ values are reported for an acid reaction temperature of 90 °C implies that the I-CDES value of any given sample will be almost 0.1 ‰ lower than its $\Delta_{47\text{CDES25}}$ values which is currently the most common convention used to report clumped isotopes. At first sight it may seem a disadvantage to lose the ability to intuitively compare new results to those obtained through different standardization approaches in the published literature. However, it will make it immediately obvious that I-CDES-normalized values cannot directly be compared to data standardized to the ETH-1/2/3/4 values reported by Bernasconi et al. (2018).

To directly compare previous carbonate-normalized Δ_{47} values and the already published calibration equations produced with ETH standard normalisation (Kele et al, 2015; Bernasconi et al., 2018; Jautzy et al., 2020; Meinicke et al., 2020; Peral et al., 2018; Piasecki et al., 2019) to data reported on the I-CDES scale, previous data have to be recalculated as described in detail in the Appendix. We emphasize again, however, that the temperatures calculated from the original calibrations using the original ETH standard values are directly comparable to temperatures calculated for samples normalized and calibrations recalculated to the I-CDES. This is because only the nominal values of the standards have changed, and thus the data normalization is internally consistent and traceable to the same solid standards.

In principle, I-CDES data are directly comparable to data produced by phosphoric acid reaction at 90°C using the HG/EG approach, with the important caveat that in absence of measurements of widely available carbonates, a direct comparison remains uncertain, especially for older data. A community effort to robustly anchor the composition of legacy standards measured in the original laboratories, may alleviate this problem in the near future.

4. Conclusions

- This study demonstrates that carbonate-based standardization of clumped isotope measurements solves many open questions that so far limited the application of carbonate clumped isotope thermometry as a mature and reliable tool in Earth sciences.
- Inter-laboratory discrepancies among 22 laboratories observed in this study are not greater than those predicted from intra-laboratory analytical uncertainties.
- We propose a set of 2 high-purity carbonate reagents and 5 widely available calcite reference materials for normalization of carbonate clumped isotope measurements with new community-accepted values: the 4 ETH standards, two samples distributed by the IAEA (C1 and C2), and a synthetic carbonate produced by MERCK. The ETH standards are available upon request from S.M. Bernasconi.
- Carbonate standardization removes the need to apply an acid digestion fractionation factor, eliminating uncertainties due to poorly known acid fractionation factors and different preparation systems and thus reduces differences between laboratories.
- In principle data expressed in the I-CDES are directly comparable to samples reacted at a temperature of 90°C normalized to the classical CDES with HG/EG. However, we emphasize that carbonate standardisation is preferred to pure HG/EG normalization because it is based on traceable carbonate samples that can be measured in every laboratory. Reporting the measured compositions of carbonate reference materials together with the samples is the only way to ensure inter-laboratory consistency.
- Robust standardization of clumped isotope measurements requires the analysis of a sufficient number of replicates of both samples and standard materials (either gases or carbonates) alongside unknowns to minimize error and obtain accurate measurements within a single laboratory.

Appendix:

A. Converting older, carbonate-anchored Δ_{47} values to the I-CDES

This section describes the steps necessary to convert existing Δ_{47} measurements to the I-CDES, provided that they were either standardized using carbonate anchors or analyzed simultaneously with several carbonate anchors. This mathematically exact approach is a simpler alternative to fully reprocessing the original raw data (with the caveat that this conversion will not provide, by itself, any error estimates).

By way of example, let us consider measurements originally standardized using ETH-1/2/3 with the nominal Δ_{47} values reported by Bernasconi et al. (2018). The "old" Δ_{47} values of these measurements are noted ${}^{old}\Delta_{47}$, and we wish to compute the "new" Δ_{47} values, noted ${}^{new}\Delta_{47}$, that would be obtained if the same data were standardized to the I-CDES.

Both old and new Δ_{47} values are derived from the same set of raw measurements using "linear" (more accurately: affine) transformations of the form:

(A.1) $$\Delta_{47}^{old} = x_1 + y_1\delta^{47} + z_1\Delta_{47}^{raw}$$

(A.2) $$\Delta_{47}^{new} = x_2 + y_2\delta^{47} + z_2\Delta_{47}^{raw}$$

We can rearrange the above equations to express Δ_{47}^{new} as a function of δ^{47} and Δ_{47}^{old} :

(A.3) $$\Delta_{47}^{new} = a + b\delta^{47} + c\Delta_{47}^{old}$$

Computing the numerical values of (a,b,c) is thus all that is required to compute Δ_{47}^{new} for any sample whose (δ^{47}, Δ_{47}^{old}) values are known. This computation only requires knowing (δ^{47}, Δ_{47}^{old}, Δ_{47}^{new}) for three different anchors, for instance ETH-1/2/3. In matrix form, eq. (A.3) then becomes:

(A.4)

Thus:
$$\begin{pmatrix} 1 & \delta^{47}_{(ETH1)} & \Delta^{old}_{47(ETH1)} \\ 1 & \delta^{47}_{(ETH2)} & \Delta^{old}_{47(ETH2)} \\ 1 & \delta^{47}_{(ETH3)} & \Delta^{old}_{47(ETH3)} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Delta^{new}_{47(ETH1)} \\ \Delta^{new}_{47(ETH2)} \\ \Delta^{new}_{47(ETH3)} \end{pmatrix}$$

$$\text{(A.5)} \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & \delta^{47}_{(ETH1)} & \Delta^{old}_{47(ETH1)} \\ 1 & \delta^{47}_{(ETH2)} & \Delta^{old}_{47(ETH2)} \\ 1 & \delta^{47}_{(ETH3)} & \Delta^{old}_{47(ETH3)} \end{pmatrix}^{-1} \begin{pmatrix} \Delta^{new}_{47(ETH1)} \\ \Delta^{new}_{47(ETH2)} \\ \Delta^{new}_{47(ETH3)} \end{pmatrix}$$

In this example, the Δ_{47}^{old} and Δ_{47}^{new} values are the old and new nominal values of ETH-1/2/3. If the δ^{47} values are defined as usual by reference to a working gas, (a,b,c) will vary when different working gases are used. However, the above equations remain valid if δ^{47} is defined instead by reference to a fixed, hypothetical CO_2 composition, e.g., stochastic VPDB-CO_2 ($\delta^{13}C_{VPDB} = 0$; $\delta^{18}O_{VSMOW} \approx 41.5$ ‰; $\Delta_{47} = 0$, $R^{47} = 4.834\times10^{-5}$). In that case, numerical values of the parameters (a,b,c) can be determined once and then used to compute Δ_{47}^{new} for any unknown sample based only on its Δ_{47}^{old} and δ^{47} values (the latter being defined relative to VPDB-CO_2 and computed by applying an acid $^{18}O/^{16}O$ fractionation factor of 1.01025 to said sample). Importantly, this particular transformation applies to all data sets originally standardized in the reference frame defined by the "old" nominal values of ETH-1/2/3.

For instance, the relationship linking the ETH-1/2/3 reference frame of Bernasconi et al. (2018) to the I-CDES is defined by:

(A.6)

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 0.010 & 0.258 \\ 1 & -28.375 & 0.258 \\ 1 & 0.538 & 0.691 \end{pmatrix}^{-1} \begin{pmatrix} 0.2052 \\ 0.2085 \\ 0.6132 \end{pmatrix} = \begin{pmatrix} -0.037997 \\ -0.000182 \\ 0.942590 \end{pmatrix}$$

Thus:

(A.7) $\Delta_{47}^{new} = -0.038039 - 0.000183\ \delta^{47} + 0.942603\ \Delta_{47}^{old}$

In this case, it is clear that the conversion from the old reference frame to the new one is not very sensitive to δ^{47} values: for unknown samples with δ^{47} values within ±6 ‰ of ETH-1 (or stochastic VPDB-CO_2), the effect of the second term in Eq. (A.7) is less than 1 ppm, and less than 3 ppm for unknowns within ±18 ‰ of ETH-1. In such cases, the conversion may be simplified as a simpler affine transformation, akin to a more traditional two-point normalization (e.g., VSMOW-VSLAP standardization):

(A.8) $\Delta_{47}^{new} = 0.942603\ \Delta_{47}^{old} - 0.038039$

For instance, to convert the Δ_{47}^{old} value of ETH-4 reported by Bernasconi et al. (2018) to the I-CDES, we only need to know that $\delta^{47}_{(ETH4)} = -28.8$ ‰ and $\Delta_{47}^{old} = 0.507 \pm 0.004$ ‰. The Δ_{47}^{new} value predicted by eq. (A.7) is then 0.445 ± 0.004 ‰, to be compared with the independently constrained values reported here in Table 1 (0.450 ± 0.002 ‰) and Table 3 (0.451 ± 0.001 ‰).

The above computation could also be performed using any arbitrary set of three carbonate materials whose (δ^{47}, Δ_{47}^{old}, Δ_{47}^{new}) values are known, provided that they span a wide enough range in δ^{47} and Δ_{47}. This is true even if the carbonates in question were not originally used to standardize the raw data, as would be the case for CO_2-standardized measurements. In the case where only two suitable carbonate standards X and Y are available, an acceptable approach would be to neglect

δ^{47} effects (equivalent to setting the value of b to zero in eq. (A.3)) by solving the following equation:

(A.9) $$\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1 & \Delta_{47(X)}^{old} \\ 1 & \Delta_{47(Y)}^{old} \end{pmatrix}^{-1} \begin{pmatrix} \Delta_{47(X)}^{new} \\ \Delta_{47(Y)}^{new} \end{pmatrix}$$

Acknowledgments, Samples, and Data

SMB acknowledges instrumentation funding from ETH Zürich and support from Swiss National Science Foundation grants 200021_143485, 200020_160046. MD acknowledges the clumped-isotope facility at LSCE is part of PANOPLY (Plateforme Analytique Géosciences Paris-Saclay) and was supported by the following institutions: Région Ile-de-France; Direction des Sciences de la Matière du Commissariat à l'Energie Atomique; Institut National des Sciences de l'Univers, Centre National de la Recherche Scientifique; Universtité de Versailles/Saint-Quentin-en-Yvelines. KDB and the MIT carbonate clumped isotope facility acknowledge support from the Packard Foundation, the Agouron Foundation and NASA Exobiology Grant 80NSSC19K0464. MB acknowledges the program Emergences Ville de Paris for early funding that allowed building her clumped isotope laboratory at Institut de Physique du Globe de Paris. MB also thanks IPGP and Université de Paris for support for the organization of the VIth Clumped Isotope Workshop in Paris in 2017. Funding was provided to CJ and the Imperial College Clumped Isotope Laboratory by Qatar Petroleum, Shell, and the Qatar Science and Technology Panel. Funding was provided to KH at University of Washington from NSF EAR grants 1933130, 1713275, and 1156134. TV acknowledges the support of the Swiss National Science Foundation grant 206021-164032. ANM acknowledges funding from the Trond Mohn Foundation and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 638467). Funding was provided to the Tripati Laboratory at UCLA for these analyses from DOE BES grant DE-FG02-13ER16402. Funding was provided to the Hyland Laboratory at NCSU for these analyses from NSF EAR-FRES grant 1925973. Funding was provided to the Yoshida Laboratory at Tokyo Tech. for these analyses from JSPS grant JP17H06105. Funding was provided to the CLIMB Lab at the University of Bergen for these

analyses by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 638467) and by the Trond Mohn Foundation. Funding was provided to the UU Clumped Laboratory at Utrecht University by the Netherlands Research Organization (NWO) through VIDI grant 016.161.365. Funding was provided to the CUBES-SIL at CU Boulder by startup funds from the University of Colorado Boulder and NSF EAR grant 1524785. The UM SCIPP Laboratory was supported by startup funds from the University of Michigan. JR Kelson was supported by NSF PRF grant 1854873. HPA acknowledges support by ERC (Grant no.724097) and ISF (Grant no. 171/16 and 1000/16). Funding was provided to the Eiler lab at Caltech from the DOE BES program, award number DE-SC0016561.

Data

All data analyzed as part of this study, along with all code used to process them, are available at https://github.com/mdaeron/InterCarb and http://doi.org/10.5281/zenodo.4314448.

Author contributions

The community designed the study at the Paris 2017 International Clumped Isotope Workshop. S.M.B. prepared materials, gathered results and wrote the first draft of the manuscript. M.D. created all code and conducted all analysis of the results. S.M.B., M.D., K.D.B., N.M., and M.B. organized the effort and contributed to early drafts. All authors contributed to the analyses and curation of data shared in this study. All authors contributed to the final version of the manuscript.

References

Adlan, Q., Davies, A.J., John, C.M., 2020. Effects of oxygen plasma ashing treatment on carbonate clumped isotopes, Rapid Communications in Mass Spectrometry, DOI: https://doi.org/10.1002/rcm.8802

Affek, H. P., & Eiler, J. M. (2006). Abundance of mass 47 CO2 in urban air, car exhaust, and human breath. *Geochimica et Cosmochimica Acta*, *70*(1), 1–12. https://doi.org/10.1016/j.gca.2005.08.021

Anderson N.T., Kelson J.R., Kele S., Daëron M., Bonifacie M., Horita, J., Mackey T.J., John C.M., Kluge T., Petschnig P., Jost, A.B., Huntington K.W., Bernasconi S.M. and Bergman K.D. (2021) A unified clumped isotope thermometer calibration (0.5-1100 °C) using carbonate-based standardization. *Geophysical Research Letters*, in press.

Bajnai, D., Guo, W., Spoetl, C. Coplen, T.B., Methner, K., Löffler, N., Krsnik, E., Gischler, E., Hansen, M., Henkel, D., Price, G.D., Raddatz, J., Scholz, D., & Fiebig, J. (2020). Dual clumped isotope thermometry resolves kinetic biases in carbonate formation temperatures. Nature Communications 11:4005.

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

Bernasconi, S. M., Müller, I. A., Bergmann, K. D., Breitenbach, S. F. M. M., Fernandez, A., Hodell, D. A., et al. (2018). Reducing uncertainties in carbonate clumped isotope analysis through consistent carbonate-based standardization. *Geochemistry, Geophysics, Geosystems*, *19*(9), 1–20. https://doi.org/10.1029/2017GC007385

Bonifacie, M., Calmels, D., Eiler, J. M., Horita, J., Chaduteau, C., Vasconcelos, C., et al. (2017). Calibration of the dolomite clumped isotope thermometer from 25 to 350 °C, and implications for a universal calibration for all (Ca, Mg, Fe)CO3 carbonates. *Geochimica et Cosmochimica Acta*, *200*, 255–279. https://doi.org/10.1016/j.gca.2016.11.028

Brand, W. a., Assonov, S. S., & Coplen, T. B. (2010). Correction for the 17O interference in δ(13C) measurements when analyzing CO2 with stable isotope mass spectrometry (IUPAC Technical Report). *Pure and Applied Chemistry*, *82*(8), 1719–1733. https://doi.org/10.1351/PAC-REP-09-01-05

Came, R. E., Brand, U., & Affek, H. P. (2014). Clumped isotope signatures in modern brachiopod carbonate. *Chemical Geology*, *377*, 20–30. https://doi.org/10.1016/j.chemgeo.2014.04.004

Carter, J. F., & Fry, B. (2013). Ensuring the reliability of stable isotope ratio data - Beyond the principle of identical treatment. *Analytical and Bioanalytical Chemistry*, *405*(9), 2799–2814. https://doi.org/10.1007/s00216-012-6551-0

Coplen, T. B. (2011). Guidelines and recommended terms for expression of stable-isotope-ratio and gas-ratio measurement results. *Rapid Communications in Mass Spectrometry*, 25(17), 2538–2560. https://doi.org/10.1002/rcm.5129

Coplen, T. B., Brand, W. A., Gehre, M., Gröning, M., Meijer, H. A. J., Toman, B., & Verkouteren, R. M. (2006). After two decades a second anchor for the VPDBδ13C scale. *Rapid Communications in Mass Spectrometry*, *20*(21), 3165–3166. https://doi.org/10.1002/rcm.2727

Daëron, M. (2021) Full propagation of analytical uncertainties in Δ_{47} measurements. Submitted to *Geochemistry, Geophysics, Geosystems (Companion Paper)*

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

Dale, A., John, C. M., Mozley, P. S., Smalley, P. C., & Muggeridge, A. H. (2014). Time-capsule concretions: Unlocking burial diagenetic processes in the Mancos Shale using carbonate clumped isotopes. Earth and Planetary Science Letters, 394. https://doi.org/10.1016/j.epsl.2014.03.004

Defliese, W. F., & Lohmann, K. C. (2015). Non-linear mixing effects on mass-47 CO<inf>2</inf> clumped isotope thermometry: Patterns and implications. *Rapid Communications in Mass Spectrometry*, *29*(9), 901–909. https://doi.org/10.1002/rcm.7175

Defliese, W. F., Hren, M. T., & Lohmann, K. C. (2015). Compositional and temperature effects of phosphoric acid fractionation on D47 analysis and implications for discrepant calibrations. Chemical Geology, 396, 51–60. https://doi.org/10.1016/j.chemgeo.2014.12.018

Dennis, K. J., Affek, H. P., Passey, B. H., Schrag, D. P., & Eiler, J. M. (2011). Defining an absolute reference frame for "clumped" isotope studies of CO2. *Geochimica et Cosmochimica Acta*, *75*(22), 7117–7131. https://doi.org/10.1016/j.gca.2011.09.025

Eagle, R. a, Schauble, E. a, Tripati, A. K., Tütken, T., Hulbert, R. C., & Eiler, J. M. (2010). Body temperatures of modern and extinct vertebrates from (13)C-(18)O bond abundances in bioapatite. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(23), 10377–10382. https://doi.org/10.1073/pnas.0911115107

Fernandez, A., Tang, J., & Rosenheim, B. E. (2014). Siderite 'clumped' isotope thermometry: A new paleoclimate proxy for humid continental environments. *Geochimica et Cosmochimica Acta*, *126*, 411–421. https://doi.org/10.1016/j.gca.2013.11.006

Fernandez, A., Müller, I. A., Rodríguez-Sanz, L., van Dijk, J., Looser, N., & Bernasconi, S. M. (2017). A Reassessment of the Precision of Carbonate Clumped Isotope Measurements: Implications for Calibrations and Paleoclimate Reconstructions. *Geochemistry, Geophysics, Geosystems*, *18*(12). https://doi.org/10.1002/2017GC007106

Ferry, J. M., Passey, B. H., Vasconcelos, C., & Eiler, J. M. (2011). Formation of dolomite at 40-80 ? °C in the Latemar carbonate buildup, Dolomites, Italy, from clumped isotope thermometry. *Geology*, *39*(6), 571–574. https://doi.org/10.1130/G31845.1

Fiebig, J., Hofmann, S., Löffler, N., Lüdecke, T., Methner, K., & Wacker, U. (2016). Slight pressure imbalances can affect accuracy and precison of dual inlet-based clumped isotope analysis. *Isotopes in Environmental and Health Studies* 52, 12-28.

Fiebig, J., Bajnai, D., Löffler, N., Methner, K., Krsnik, E., Mulch, A., & Hofmann, S. (2019). Combined high-precision Δ_{48} and Δ_{47} analysis of carbonates. *Chemical Geology*, *522* (Feb), 186–191. https://doi.org/10.1016/j.chemgeo.2019.05.019

Ghosh, P., Adkins, J., Affek, H., Balta, B., Guo, W., Schauble, E. A., et al. (2006). 13C-18O bonds in carbonate minerals: A new kind of paleothermometer. *Geochimica et Cosmochimica Acta*, *70*(6), 1439–1456. https://doi.org/10.1016/j.gca.2005.11.014

Ghosh, P., Garzione, C., & Eiler, J. (2006). Rapid Uplift of the Alitplano Revealed Through 13C-18O Bonds in Paleosol Carbonates. *Science*, *311*(January), 511–515. https://doi.org/10.1126/science.1119365

Grauel, A. L., Schmid, T. W., Hu, B., Bergami, C., Capotondi, L., Zhou, L., & Bernasconi, S. M. (2013). Calibration and application of the "clumped isotope" thermometer to

foraminifera for high-resolution climate reconstructions. *Geochimica et Cosmochimica Acta*, *108*, 125–140. https://doi.org/10.1016/j.gca.2012.12.049

Guo, W., & Eiler, J. M. (2007). Temperatures of aqueous alteration and evidence for methane generation on the parent bodies of the CM chondrites. *Geochimica et Cosmochimica Acta*, *71*(22), 5565–5575. https://doi.org/10.1016/j.gca.2007.07.029

Guo, W., Mosenfelder, J. L., Goddard, W. A., & Eiler, J. M. (2009). Isotopic fractionations associated with phosphoric acid digestion of carbonate minerals: Insights from first-principles theoretical modeling and clumped isotope measurements. Geochimica et Cosmochimica Acta, 73(24), 7203–7225. https://doi.org/10.1016/j.gca.2009.05.071

He, B., Olack, G. A., & Colman, A. S. (2012). Pressure baseline correction and high-precision CO2 clumped-isotope (*Δ*47) measurements in bellows and micro-volume modes. *Rapid Communications in Mass Spectrometry*, *26*(24), 2837–2853. https://doi.org/10.1002/rcm.6436

Hu, B., Radke, J., Schlüter, H. J., Heine, F. T., Zhou, L., Bernasconi, S. M., et al. (2014). A modified procedure for gas-source isotope ratio mass spectrometry: The long-integration dual-inlet (LIDI) methodology and implications for clumped isotope measurements. *Rapid Communications in Mass Spectrometry*, *28*(13), 1413–1425. https://doi.org/10.1002/rcm.6909

Huntington, K. W., Eiler, J. M., Affek, H. P., Guo, W., Bonifacie, M., Yeung, L. Y., et al. (2009). Methods and limitations of "clumped" CO2 isotope (*Δ*47) analysis by gas-source isotope ratiomass spectrometry. *Journal of Mass Spectrometry*, *44*(9), 1318–1329. https://doi.org/10.1002/jms.1614

Huntington, K. W., Budd, D. A., Wernicke, B. P., & Eiler, J. M. (2011). Use of Clumped-Isotope Thermometry To Constrain the Crystallization Temperature of Diagenetic Calcite. *Journal of Sedimentary Research*, *81*(9), 656–669. https://doi.org/10.2110/jsr.2011.51

Jautzy, J. J., Savard, M. ., Dhillon, R. ., Bernasconi, S. ., Lavoie, D., & Smirnoff, A. (2021). Clumped isotope temperature calibration for calcite: Bridging theory and experimentation. *Geochemical Perspective Letters*, (2020), 36–41. https://doi.org/10.7185/geochemlet.2021

Katz A., Bonifacie M., Hermoso M., Cartigny P., Calmels D. (2017) Laboratory-grown coccoliths exhibit no vital effect in clumped isotope (Δ_{47}) composition on a range of geologically relevant temperatures in *Geochimica Cosmochimica Acta*, 208, 335–353. https://doi.org/10.1016/j.gca.2017.02.025

Kele, S., Breitenbach, S. F. M., Capezzuoli, E., Meckler, A. N., Ziegler, M., Millan, I. M., Kluge, T., Deák, J., Hanselmann, K., John, C. M., Yan, H., Liu, Z., & Bernasconi, S. M. (2015). Temperature dependence of oxygen- and clumped isotope fractionation in carbonates: A study of travertines and tufas in the 6-95°C temperature range. Geochimica et Cosmochimica Acta, 168. https://doi.org/10.1016/j.gca.2015.06.032

Kelson, J. R., Huntington, K. W., Schauer, A. J., Saenger, C., & Lechler, A. R. (2017). Toward a universal carbonate clumped isotope calibration: Diverse synthesis and preparatory methods suggest a single temperature relationship. *Geochimica et Cosmochimica Acta*, *197*, 104–131. https://doi.org/10.1016/j.gca.2016.10.010

Kim, S. T., Coplen, T. B., & Horita, J. (2015). Normalization of stable isotope data for carbonate minerals: Implementation of IUPAC guidelines. *Geochimica et Cosmochimica Acta*, *158*, 276–289. https://doi.org/10.1016/j.gca.2015.02.011

Kluge, T., John, C. M., Jourdan, A.-L., Davis, S., & Crawshaw, J. (2015). Laboratory calibration of the calcium carbonate clumped isotope thermometer in the 25-250°C temperature range. G*eochimica et Cosmochimica Acta*, 157, 213-227. https://doi.org/10.1016/j.gca.2015.02.028

Kocken, I. J., Müller, I. A., & Ziegler, M. (2019). Optimizing the Use of Carbonate Standards to Minimize Uncertainties in Clumped Isotope Data. *Geochemistry, Geophysics, Geosystems*, *20*(11), 5565–5577. https://doi.org/10.1029/2019GC008545

Mangenot X., Gerdes A., Gasparrini M., Bonifacie M., Rouchon V. (2018). An emerging thermo-chronometer for carbonate-bearing rocks : Δ_{47}/(U-Pb). *Geology* 46 (12), 1067-1070.

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

Meckler, A. N., Ziegler, M., Millan, M. I., Breitenbach, S. F. M., Bernasconi, S. M., Millán, M. I., et al. (2014). Long-term performance of the Kiel carbonate device with a new

correction scheme for clumped isotope measurements. *Rapid Communications in Mass Spectrometry*, *28*(15), 1705–1715. https://doi.org/10.1002/rcm.6949

Meier-Augenstein, W., & Schimmelmann, A. (2019). A guide for proper utilisation of stable isotope reference materials *. *Isotopes in Environmental and Health Studies*, *55*(2), 113–128. https://doi.org/10.1080/10256016.2018.1538137

Meinicke, N., Ho, S. L., Hannisdal, B., Nürnberg, D., Tripati, A., Schiebel, R., & Meckler, A. N. (2020). A robust calibration of the clumped isotopes to temperature relationship for foraminifers. *Geochimica et Cosmochimica Acta*, *270*, 160–183. https://doi.org/10.1016/j.gca.2019.11.022

Müller, I. A., Fernandez, A., Radke, J., Dijk, J. Van, Bowen, D., Schwieters, J., et al. (2017). Carbonate clumped isotope analyses with the long-integration dual-inlet (LIDI) workflow: scratching at the lower sample weight boundaries. *Rapid Communications in Mass Spectrometry*, *31*(12), 1057–1066. https://doi.org/10.1002/rcm.7878

Müller, I. A., Violay, M. E. S., Storck, J.-C. C., Fernandez, A., van Dijk, J., Madonna, C., & Bernasconi, S. M. (2017). Clumped isotope fractionation during phosphoric acid digestion of carbonates at 70 °C. *Chemical Geology*, *449*, 1–14. https://doi.org/10.1016/j.chemgeo.2016.11.030

Müller, I. A., Rodriguez-Blanco, J. D., Storck, J.-C., do Nascimento, G. S., Bontognali, T. R. R., Vasconcelos, C., et al. (2019). Calibration of the oxygen and clumped isotope thermometers for (proto-)dolomite based on synthetic and natural carbonates. *Chemical Geology*, *525*(December 2018), 1–17. https://doi.org/10.1016/j.chemgeo.2019.07.014

Murray, S. T., Arienzo, M. M., & Swart, P. K. (2016). Determining the *Δ*47 acid fractionation in dolomites. *Geochimica et Cosmochimica Acta*, 174, 42–53. https://doi.org/10.1016/j.gca.2015.10.029

Nishida, K., & Ishimura, T. (2017). Grain-scale stable carbon and oxygen isotopic variations of the international reference calcite, IAEA-603. *Rapid Communications in Mass Spectrometry*, 31(22), 1875–1880. https://doi.org/10.1002/rcm.7966

Passey, B. H., & Henkes, G. A. (2012). Carbonate clumped isotope bond reordering and geospeedometry. *Earth and Planetary Science Letters*, *351–352*, 223–236. https://doi.org/10.1016/j.epsl.2012.07.021

Passey, B. H., Levin, N. E., Cerling, T. E., BROWN, F. H., Eiler, J. M., & Turekian, K. K. (2010). High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11245–11249. https://doi.org/10.1073/pnas.1001824107

Peral, M., Daëron, M., Blamart, D., Bassinot, F., Dewilde, F., Smialkowski, N., et al. (2018). Updated calibration of the clumped isotope thermometer in planktonic and benthic foraminifera. *Geochimica et Cosmochimica Acta*, *239*, 1–16. https://doi.org/10.1016/j.gca.2018.07.016

Petersen, S. V., Defliese, W. F. F., Saenger, C., Daëron, M., Huntington, K. W. W., John, C. M. M., et al. (2019). Effects of Improved 17 O Correction on Inter-laboratory Agreement in Clumped Isotope Calibrations, Estimates of Mineral-Specific Offsets, and Temperature Dependence of Acid Digestion Fractionation. *Geochemistry, Geophysics, Geosystems*, *20*, 3495–3519. https://doi.org/10.1029/2018GC008127

Petersen, Sierra V., Winkelstern, I. Z., Lohmann, K. C., & Meyer, K. W. (2016). The effects of Porapak^{TM} trap temperature on $\delta^{18}O$, $\delta^{13}C$, and $\Delta 47$ values in preparing samples for clumped isotope analysis. *Rapid Communications in Mass Spectrometry*, *30*(1), 199–208. https://doi.org/10.1002/rcm.7438

Piasecki, A., Bernasconi, S. M., Grauel, A., Hannisdal, B., Ho, S. L., Leutert, T. J., et al. (2019). Application of Clumped Isotope Thermometry to Benthic Foraminifera. *Geochemistry, Geophysics, Geosystems*, 2018GC007961. https://doi.org/10.1029/2018GC007961

Qi, H., Moossen, H., Meijer, H. A. J., Coplen, T. B., Aerts-Bijma, A. T., Reid, L., et al. (2021). USGS44, a new high-purity calcium carbonate reference material for $\delta^{13}C$ measurements . *Rapid Communications in Mass Spectrometry*, 35(4). https://doi.org/10.1002/rcm.9006

Schauer, A. J., Kelson, J., Saenger, C., & Huntington, K. W. (2016). Choice of 17O correction affects clumped isotope (Δ_{47}) values of CO_2 measured with mass spectrometry. *Rapid Communications in Mass Spectrometry*, *30*(24), 2607–2616. https://doi.org/10.1002/rcm.7743

Schmid, T. W., & Bernasconi, S. M. (2010). An automated method for "clumped-isotope" measurements on small carbonate samples. *Rapid Communications in Mass Spectrometry*, *24*(14), 1955–1963. https://doi.org/10.1002/rcm.4598

Spooner, P. T., Guo, W., Robinson, L. F., Thiagarajan, N., Hendry, K. R., Rosenheim, B. E., & Leng, M. J. (2016). Clumped isotope composition of cold-water corals: A role for vital effects? *Geochimica et Cosmochimica Acta*, *179*, 123–141. https://doi.org/10.1016/j.gca.2016.01.023

Swart, P. K., Murray, S. T., Staudigel, P. T., & Hodell, D. A. (2019). Oxygen Isotopic Exchange between CO_2 and Phosphoric Acid: Implications for the Measurement of Clumped Isotopes in Carbonates. *Geochemistry, Geophysics, Geosystems*, (1950), 2019GC008209. https://doi.org/10.1029/2019GC008209

Thaler C., Katz A., Bonifacie M., Ménez B., Ader M. Oxygen isotope composition of waters recorded in carbonates in strong clumped and oxygen isotopic disequilibrium (2020). *Biogeosciences*, 17, 1–14, 2020. DOI 10.5194/bg-17-1-2020

van Dijk, J., Fernandez, A., Storck, J. C., White, T. S., Lever, M., Müller, I. A., et al. (2019). Experimental calibration of clumped isotopes in siderite between 8.5 and 62°C and its application as paleo-thermometer in paleosols. Geochimica et Cosmochimica Acta, 254, 1–20. https://doi.org/https://doi.org/10.1016/j.gca.2019.03.018

Veillard, C. M. A., John, C. M., Krevor, S., & Najorka, J. (2019). Rock-buffered recrystallization of Marion Plateau dolomites at low temperature evidenced by clumped isotope thermometry and X-ray diffraction analysis. *Geochimica et Cosmochimica Acta,* 252, 190–212. https://doi.org/10.1016/J.GCA.2019.02.012

Wacker, U., Fiebig, J., & Schoene, B. R. (2013). Clumped isotope analysis of carbonates: comparison of two different acid digestion techniques. Rapid Communications in Mass Spectrometry : RCM, 27(14), 1631–42. https://doi.org/10.1002/rcm.6609

Wacker, U., Fiebig, J., Tödter, J., Schöne, B. R., Bahr, A., Friedrich, O., et al. (2014). Empirical calibration of the clumped isotope paleothermometer using calcites of various origins. *Geochimica et Cosmochimica Acta*, *141*, 127–144. https://doi.org/10.1016/j.gca.2014.06.004

Wang, Z., Schauble, E. A., & Eiler, J. M. (2004). Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. *Geochimica et Cosmochimica Acta*, *68*(23), 4779–4797. https://doi.org/10.1016/j.gca.2004.05.039

Werner, R. A., & Brand, W. A. (2001a). Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Communications in Mass Spectrometry*, *15*(7), 501–519. https://doi.org/10.1002/rcm.258

Werner, R. A., & Brand, W. A. (2001b). Referencing strategies and techniques in stable isotope ratio analysis. *Rapid Communications in Mass Spectrometry*, *15*(7), 501–519. https://doi.org/10.1002/rcm.258