RESEARCH ARTICLE

Carbonate clumped isotope analysis (Δ_{47}) of 21 carbonate standards determined via gas-source isotope-ratio mass spectrometry on four instrumental configurations using carbonate-based standardization and multiyear data sets

Deepshikha Upadhyay[1] | Jamie Lucarelli[1] | Alexandrea Arnold[1] |
Randy Flores[1] | Hayley Bricker[1] | Robert N. Ulrich[1] | Gregory Jesmok[1,2] |
Lauren Santi[1,3] | William Defliese[1,4] | Robert A. Eagle[1] | Hannah M. Carroll[1] |
Jesse Bloom Bateman[1,5] | Victoria Petryshyn[1,6] | Sean J. Loyd[1,7] | Jianwu Tang[1] |
Antra Priyadarshi[1] | Ben Elliott[1] | Aradhna Tripati[1]

[1]Department of Earth, Planetary, and Space Sciences, Department of Atmospheric and Oceanic Sciences, Institute of the Environment and Sustainability, Center for Diverse Leadership in Science, University of California, Los Angeles, California

[2]Department of Geological Sciences, California State University, Northridge, California

[3]GSI Environmental Inc, Irvine, California

[4]School of Earth and Environmental Sciences, The University of Queensland, St Lucia, Queensland, Australia

[5]Biological Sciences Department, State University of New York, Cortland, New York

[6]Environmental Studies Program, Department of Environmental Studies, University of Southern California, Los Angeles, California

[7]Geological Sciences Department, California State University Fullerton, Fullerton, California

Correspondence
D. Upadhyay and A. Tripati, Department of Earth, Planetary, and Space Sciences, Department of Atmospheric and Oceanic Sciences, Institute of the Environment and Sustainability, Center for Diverse Leadership in Science, University of California, Los Angeles, CA 90290, USA.
Email: deepshikha1936@gmail.com; atripati@g.ucla.edu

Rationale: Clumped isotope geochemistry examines the pairing or clumping of heavy isotopes in molecules and provides information about the thermodynamic and kinetic controls on their formation. The first clumped isotope measurements of carbonate minerals were first published 15 years ago, and since then, interlaboratory offsets have been observed, and laboratory and community practices for measurement, data analysis, and instrumentation have evolved. Here we briefly review historical and recent developments for measurements, share Tripati Lab practices for four different instrument configurations, test a recently published proposal for carbonate-based standardization on multiple instruments using multi-year data sets, and report values for 21 different carbonate standards that allow for recalculations of previously published data sets.

Methods: We examine data from 4628 standard measurements on Thermo MAT 253 and Nu Perspective IS mass spectrometers, using a common acid bath (90°C) and small-sample (70°C) individual reaction vessels. Each configuration was investigated

by treating some standards as anchors (working standards) and the remainder as unknowns (consistency standards).

Results: We show that different acid digestion systems and mass spectrometer models yield indistinguishable results when instrument drift is well characterized. For linearity correction, mixed gas-and-carbonate standardization or carbonate-only standardization yields similar results. No difference is observed in the use of three or eight working standards for the construction of transfer functions.

Conclusions: We show that all configurations yield similar results if instrument drift is robustly characterized and validate a recent proposal for carbonate-based standardization using large multiyear data sets. Δ_{47} values are reported for 21 carbonate standards on both the absolute reference frame (ARF; also refered to as the Carbon Dioxide Equilibrated Scale or CDES) and the new InterCarb-Carbon Dioxide Equilibrium Scale (I-CDES) reference frame, facilitating intercomparison of data from a diversity of labs and instrument configurations and restandardization of a broad range of sample sets between 2006, when the first carbonate measurements were published, and the present.

1 | INTRODUCTION

The field of clumped isotope geochemistry is based on the abundances of multiple heavy isotope substitutions in molecules, including CO_2 (e.g., $^{13}C^{18}O^{16}O$), dissolved inorganic carbon species, carbonate minerals, and methane.[1–4] Equilibrium fractionations of multiply-substituted molecular species are governed by temperature-dependent homogeneous isotope exchange reactions.[3,5–7] The most widely studied system in clumped isotope geochemistry is based on the proportion of ^{13}C and ^{18}O isotopes bound to each other within carbonate ion ($^{13}C^{18}O^{16}O_2^{2-}$) groups in carbonate-containing minerals, relative to a stochastic distribution of isotopes.[1–3,6,7] Carbonate clumped isotope thermometry has been used to probe the temperature history of surface[8] and subsurface environments,[9] paleo-physiology,[10] biomineralization processes,[11] and the alteration history of meteorites,[12–14] as well as the origin of possible kinetic effects.[3,15]

However, a range of instrumentation and techniques have been employed by different laboratories. Several studies comparing a breadth of instrumentation have identified factors that contribute to interlaboratory offsets and proposed advances to address intercomparability.[16–20] Community-wide changes in standardization methods have occurred,[16] in addition to laboratory-specific practices.[17] Building on this work, we implement recent proposals from Bernasconi et al[17,21] of utilizing carbonate standardization for reference frame construction (as initially described by Dennis et al[16]). We utilize and build on this suggested standardization approach on a multiyear data set (n = 4628) from multiple instruments and digestion systems to explore the efficacies of different implementations.

The Tripati laboratory at the University of California, Los Angeles (UCLA), is uniquely situated to perform this as it employs both the pioneering instrumentation that is widely used (Thermo MAT 253) and the newest generation of mass spectrometers (Nu Perspective IS).

The Thermo MAT 253 is paired with a common acid bath (CAB) digestion system for large samples containing 3–6 mg of pure carbonate.[22] The Nu Perspective IS has features such as secondary electron suppression and continuous pressure balancing, which allow for the use of both a small (0.25–0.5 mg of pure carbonate) single-sample digestion system (the Nu Carb device) and a CAB sample digestion system designed and fabricated in our laboratory.

This study presents standard data and shows that although considerable instrument drift can occur on timescales of more than a month, with careful standardization the different instruments yield consistent results, following on a recent study comparing synthetic carbonates run on these same instruments by Defliese and Tripati.[18] We find that carefully chosen measurement windows associated with frequent analysis of standards are necessary to yield results that are comparable across multiple instruments. No statistically significant differences in data accuracy or precision are discernable using different machine configurations in this study, using a proposal from Bernasconi et al.[17,21] We show that convergent results can be achieved using all acid digestion setups, acid temperatures, sample sizes, mass spectrometers, and standardization procedures. We report clumped isotope values for 21 internal and international carbonate standards sourced from laboratories in China, Switzerland, and the United States. We have run these standards on multiple mass spectrometer systems and anchored these results to both the absolute reference frame (ARF or CDES) and the new InterCarb-Carbon Dioxide Equilibrium Scale (I-CDES) reference frame. These results are promising for future widespread applications of clumped isotope analyses, as we expand the number of instrumental configurations beyond what has been utilized previously. The breadth of standards reported allows for intercomparison of data from a diversity of laboratories and instrument configurations and the restandardization of a broad range of sample sets.

2 | BACKGROUND

2.1 | Explanation of the carbonate clumped isotope thermometer

In carbonate minerals and dissolved inorganic carbon (DIC) species, the heavy isotopes, ^{13}C and ^{18}O, are rare relative to the light isotopes, ^{12}C and ^{16}O. Combinations of these isotopes form isotopologues of carbonate ions, with masses varying from 60 to 67 Da. Acid digestion of carbonate minerals with these isotopologues produces CO_2 isotopologues of masses 44–49 Da, with abundances related to their abundances in the solid before acid digestion (Table S1 [supporting information]).[1] Multiply-substituted isotopologues of $CaCO_3$, DIC, and other molecules possess distinct properties due to distinguishable differences in zero-point energy.[1,6] Therefore, they serve as tracers of equilibrium and kinetic processes. While similar to the use of singly-substituted isotopes in many geochemical applications, the carbonate clumped isotope thermometer is based on internal isotope exchange reactions occurring within a single phase. Thus, carbonate minerals that form in equilibrium have a clumped isotopic composition that is influenced only by temperature.[1–3,6,7,13,23–25]

For carbonate minerals, statistical thermodynamics predicts that the bonding or "clumping" of heavy isotopes ^{13}C and ^{18}O is negatively correlated with temperature.[6] Heavy isotope bonding between ^{13}C and ^{18}O will conform to isotopic equilibrium constants for reactions such as the following:

$$X^{12}C^{18}O^{16}O_2 + X^{13}C^{16}O_3 \rightarrow X^{13}C^{18}O^{16}O_2 + X^{12}C^{16}O_3^2 \quad (1)$$
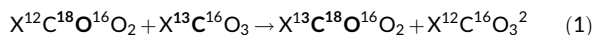
where X refers to cations such as Ca^{2+}, Mg^{2+}, Sr^{2+}, Ba^{2+}.[6–8] The $^{13}C-^{18}O$ bond abundance in carbonate minerals can be measured using isotope-ratio gas-source mass spectrometry on CO_2 liberated from carbonates by reaction with phosphoric acid.[1]

2.2 | Notation

$^{13}C-^{18}O$ bond ordering in CO_2 is described using the quantity Δ_{47}: the deviation in the abundance of CO_2 ions with m/z 47 from the abundance predicted by stochastic mixing. Several papers discuss the calculation of isotope ratios in detail.[7,22,26,27]

Briefly, the quantity Δ_{47} is dependent on the abundance ratios of isotopologues $^{47}CO_2$, $^{46}CO_2$, and $^{45}CO_2$ relative to the major isotopologue $^{44}CO_2$,[22] denoted by R_{47}, R_{46}, and R_{45}. Because R_{13} and R_{18} are the ratios for $^{13}C/^{12}C$ and $^{18}O/^{16}O$, respectively, the isotopologue ratios R_{47}, R_{46}, and R_{45} can be defined as follows:

$$R_{47} = (2 \cdot R_{13} \cdot R_{18}) + (2 \cdot R_{17} \cdot R_{18}) + \left(R_{13}(R_{17})^2\right) \quad (2)$$

$$R_{46} = (2 \cdot R_{18}) + (2 \cdot R_{13} \cdot R_{17}) + (R_{17})^2 \quad (3)$$

$$R_{45} = R_{13} + (2 \cdot R_{17}) \quad (4)$$

Given that Δ_{47} is defined as the excess of $^{13}C^{18}O^{16}O$ in a sample relative to a stochastic abundance (denoted by "*"), it can also be described as follows:

$$\Delta_{47} = [(R_{47}/R_{47}*-1) - (R_{46}/R_{46}*-1) - (R_{45}/R_{45}*-1)] \cdot 1000 \quad (5)$$

Expanded to include all terms, the quantity Δ_{47} is given as follows:

$$\Delta 47 = \frac{R^{47}}{(2 \cdot R_{13} \cdot R_{18}) + (2 \cdot R_{17} \cdot R_{18}) + \left(R_{13}(R_{17})^2\right)} - \frac{R^{46}}{(2 \cdot R_{18}) + (2 \cdot R_{13} \cdot R_{17}) + (R_{17})^2} - \frac{R^{45}}{R_{13} + (2 \cdot R_{17})} + 1$$

Δ_{63} refers to the deviation in the abundance of the multiply-substituted CO_3^{2-} isotopologue $^{13}C^{18}O^{16}O_2$ of m/z 63 from the abundance predicted by random mixing. Δ_{47} of a sample is related to Δ_{63} of the same sample by correcting for isotopic fractionation that occurs during acid digestion, such that

$$\Delta_{47} = \Delta_{63} + y \quad (6)$$

where y is the fractionation factor.

The bulk compositions of ions with m/z 47 are reported relative to a reference gas standard, denoted by δ_{47}:

$$\delta_{47} (‰) = \left(R_{47-sample\ gas}/R_{47-reference\ gas} - 1\right) \cdot 1000 \quad (7)$$

Given the range of reference gases utilized by different laboratories as working gases, measured values of δ_{47} are not directly comparable between laboratories.

3 | A REVIEW OF ANALYTICAL MEASUREMENTS AND DATA HANDLING

The mass spectrometric procedures that were used 11 years ago in the first laboratory that pioneered the measurement are reported in depth in Huntington et al.[26] Since then, there have been some modifications. Therefore, here we describe briefly analytical measurements and data handling, as well as relevant literature for reference. We describe in Section 4 the methods used in this study.

In brief, mass spectrometric measurements of clumped isotope measurements of carbonates involve the initial conversion of carbonate minerals to CO_2 by *ortho*-phosphoric acid digestion at a constant temperature. The product gases are then purified to remove water, noncondensable gases such as N_2 and SO_2, and organic compounds that are considered contaminants. Although there are some variations, most methods use cryogenic purification to remove water and noncondensable gases, a trap or column with Porapak Q for organics and silver wool for sulfur.[1,22,26,28] The abundances of CO_2 isotopologues 44–47 are determined using a gas-source isotope-ratio mass spectrometer.[1] Subsequent data handling includes corrections for ^{17}O and the bulk isotopic composition.[29] Outputs

from software associated with commercial mass spectrometers are processed using a diverse array of approaches, which may include Excel, scripts in R or other languages,[19] and the community freeware "Easotope."[30] Results are standardized using equilibrated gas standards, carbonate standards, or both.[1,16,17]

3.1 | ^{17}O, $^{18}O/^{16}O$, and $^{13}C/^{12}C$ definitions

The impact of using different parameter sets for the calculation of absolute isotope ratios (i.e., the $^{13}C/^{12}C$ ratio of Vienna Pee Dee Belemnite (VPDB), the $^{17}O/^{16}O$ and $^{18}O/^{16}O$ ratios of Vienna Standard Mean Ocean Water (VSMOW) or VPDB-CO_2, and the slope of the triple oxygen isotope line [λ]) was initially explored by Daëron et al[29] and Schauer et al,[31] who showed that this was a potentially important source of interlaboratory offsets for some laboratories. Their findings were partially validated for some instrumental configurations but may not be as relevant to other configurations.[19] The impact of the choice of parameter set depends in part on the composition of the samples analyzed and the working gas used.[19,29] Best practices for these corrections include the use of the "IUPAC" (International Union of Pure and Applied Chemistry) parameter set,[32] that are reviewed in depth by Daëron et al[29] and Petersen et al.[19]

3.2 | Using gas and/or carbonate standards for mass spectrometric corrections

Initially, clumped isotope measurements utilized 1000°C equilibrated gases of different bulk compositions that had stochastic distributions as working standards to implement a linearity correction,[1,29] an analytical artifact observed on the Thermo 253 that yields a changing dependence of measured Δ_{47} on δ_{47}. Subsequent standardization approaches included the use of multiple sets of heated and water equilibrated gases,[16,18,19] carbonate standards,[16–18] or a mixed gas-and-carbonate standards approach[18,22] to implement both a linearity correction and a scale compression/stretching correction.

Mass spectrometric methods have changed since the first clumped isotope measurements of CO_2 were published 17 years ago,[23] including with the implementation of different processes to correct for nonlinearities in the measurement of multiply-substituted isotopic species. It is now understood that clumped isotope collectors on mass spectrometers have negative backgrounds resulting from the presence of secondary electrons. These negative backgrounds are largely controlled by the flux rate of CO_2 to the ion source and the resultant intensities of the ion beams.[33,34] These negative backgrounds are typically statistically irrelevant in conventional analyses of singly-substituted isotopologues that yield $\delta^{18}O$ and $\delta^{13}C$ values but are large compared with measured clumped isotope signals. Changes in mass spectrometer linearity due to pressure-dependent backgrounds can occur over short or long timescales (from weeks to months to years) and manifest as a dependence on the clumped isotope signal ($\Delta_{47\text{-}WG}$) on bulk isotopic

composition ($\delta_{47\text{-}WG}$). To date, multiple methods have been used for implementing these linearity corrections. Equilibrated gases are used by some laboratories to monitor the effects of changes in the negative background,[1,16,23,25,26,35] whereas others use measurements of the pressure-dependent background made by off-peak measurements.[33,34] Instruments with secondary electron suppression reduce the scale of these effects.[18]

Investigations of interlaboratory offsets led to the proposal of an absolute reference frame with a scale for Δ_{47} defined by the composition of gases equilibrated at two different temperatures (e.g., 25°C and 1000°C) that were pinned to theoretical calculations for thermodynamic equilibrium.[16] It had been observed that on the stochastic reference frame, scale compression/stretching could occur, not only between instruments but also on a single instrument, as noticed by the presence of drift in carbonate standard values over a timescale of months to years (e.g., see Figure S4 in Passey et al[22]). Thus, running equilibrated gases of two different bulk compositions and temperatures would allow for a linearity correction and provide the basis for an additional scale compression/stretching correction.[16] This work showed that the use of an absolute reference frame improved interlaboratory reproducibility. In addition, Dennis et al[16] suggested that carbonate standards could be defined in the absolute reference frame and used to develop secondary transfer functions (i.e., standardization based solely on carbonate standards) or a mixed transfer function (i.e., standardization based on a mixture of equilibrated gases and carbonate standards). Passey et al[22] utilized a gas-based standardization scheme while using carbonate standards to normalize sample data via a polynomial fit minimizing standard residuals. The analysis of carbonate standards also circumvents the challenge of treating samples and standards differently, which results only when gases are used for reference frame calculations.

More recent work has suggested the use of carbonate standards for both linearity and scale compression corrections. Carbonates with a stochastic distribution and distinct bulk $\delta^{18}O$ and $\delta^{13}C$ compositions can be used for linearity corrections, such as the widely used carbonate standards ETH-1 and ETH-2 produced in heating experiments.[17] This carbonate standards–based approach could be used instead of, or in addition to, the analysis of equilibrated gases. An additional carbonate standard with a much higher Δ_{47} value (e.g., the unheated version of these standards) could be used to constrain scale compression/stretching, such as ETH-3 or ETH-4,[17] or other carbonate standards. Accurately pinning these values to the composition of equilibrated gases is critical for projection into the absolute reference frame.

3.3 | Acid digestion corrections

Acid digestion corrections are utilized to account for the range of different reaction temperatures used for clumped isotope measurements of carbonates (from 25°C to 110°C). Differences are observed between measured $\Delta_{47\text{-}25°C}$ and $\Delta_{47\text{-}70°C}$ (or digestion at any other temperature), and corrections are necessary to account for

systematic offsets between materials digested at different temperatures. For oxygen isotopes, small but significant variations in acid digestion fractionation occur due to temperature, as well as digestion apparatus, acid density, grain size, mineral chemistry and structure, and acid:mineral ratio.[35] These same factors are likely to be important for acid digestion fractionations of clumped isotopes, and thus acid digestion fractionations are the subject of several studies.[1,13,28,36–38]

A temperature-dependent acid fractionation factor (AFF) is usually added as a final step in the data reduction process[16] to help ensure that values of Δ_{47} are comparable between laboratories.[1,13,28,36–38] Furthermore, accounting for differences between Δ_{47} (determined on CO_2) and Δ_{63} (the value for the carbonate ion group) facilitates comparison with theory. $\Delta_{47} - \Delta_{63}$ offsets can be empirically determined using material from heating experiments,[1,13,17,39] or theoretical values can be used.[13]

We note that an acid digestion correction is, in principle, necessary only when equilibrated gases are used to build the absolute reference frame. Carbonate-based standardization can eliminate the need for acid digestion corrections, as the AFF is built into the reference frame. However, this is the case only if carbonate standard values are accurately defined in the absolute reference frame.

4 | METHODS

We used two different types of mass spectrometers for clumped isotope measurements of CO_2 at UCLA: (a) the Thermo 253 (Thermo Fisher Scientific, Bremen, Germany), the most widely used instrument for clumped isotope measurements; and (b) the Nu Perspective IS (Nu Instruments Ltd, Wrexham, UK) with secondary electron suppression. We also have two different types of acid digestion and sample purification systems, and in all cases, reaction times, freeze-down temperatures, and transfer timings need to be determined. The different instrument configurations used in this study are described in Table 1.

4.1 | Phosphoric acid digestion and gas purification

4.1.1 | Common acid bath

The first type of digestion system used is a CAB system modeled after what is described in Passey et al.[22] Typical sample sizes range from 3 to 8 mg of pure $CaCO_3$, with larger amounts of sample required as the $CaCO_3$ content decreases. The CAB systems that we used were built in-house and modeled after what is used at Caltech and at Johns Hopkins University.[22,40] They include an automated, online digestion and CO_2 purification system (also referred to as an autoline). Briefly, an autoline consists of the following:

1. A zero blank autosampler (Costech, Valencia, CA, USA) or custom-built Fairbanks device made of stainless steel that is capable of pulling high vacuum and that can hold multiple samples.
2. A CAB for phosphoric acid digestion of samples that are typically reacted for 20 min at a mineral-specific temperature (in our laboratory, typically 90°C but ranging from 70°C to 110°C). The CO_2 gas liberated during digestion remains in contact with the acid until proceeding in the line toward cryogenic traps.
3. Cryogenic traps for removal of water and other gases with low vapor pressures, with one trap containing ethanol surrounded by dry ice and a second trap containing liquid nitrogen.
4. An in-line elemental-silver wool column (Sigma-Aldrich, St. Louis, MO, USA) to remove sulfur compounds from the gas mixture.
5. A gas chromatography (GC) column using ultra-high-purity helium carrier gas and Porapak Q 50/80 mesh column packing material that maintains −20°C during the gas transit to separate CO_2 from the remaining components of the produced gas mixture.
6. A final cryogenic purification stage before transfer of CO_2 into the bellows of the mass spectrometer.

The use of a GC column packed with Porapak Q and helium as a carrier gas enables this system to handle larger samples with significant organic content and smaller concentrations of pure

TABLE 1 Description of each instrument configuration in this study

	Configuration 1	Configuration 2	Configuration 3A	Configuration 3B
Acid digestion system	Common acid bath	Common acid bath	Nu Carb	Nu Carb
Acid temperature	90°C	90°C	70°C	70°C
Mass spectrometer model	Thermo 253	Nu Perspective (2014 model)	Nu Perspective (2014 model)	Nu Perspective (2016 model)
Regularly analyze equilibrated gases?	Yes	Yes	No	No
Mass-44 beam ion intensity	16 V	80 nA	80 nA for large samples, 80–30 nA for small samples	80 nA for large samples, 80–30 nA for small samples
Integration time	720 s	1600s	1600 s for large samples, 1200 s for small samples	1600 s for large samples, 1200 s for small samples

carbonate than with other systems. Custom software was written in Labview to control the autoline (the autosampler, all valves, the GC, and Dewar lifters) and is coupled to the mass spectrometer control software (Thermo Fisher Scientific Isodat software interface for the MAT 253; Perspective Stable Gas Control Software [Nu Instruments] for the Nu Perspective).

4.1.2 | Nu Carb sample preparation system

The second type of acid digestion system that we use is a commercial Nu Carb sample digestion system (Nu Instruments Ltd) that can react samples typically of 0.25–0.5 mg of pure $CaCO_3$. The controlling software, Perspective Stable Gas Control Software, was also developed by Nu Instruments. Samples were reacted for 20 min at 70°C in individual reaction vials that eliminate potential memory effects from a CAB. This is similar to a Thermo Fisher Scientific Kiel device, in which both the Nu Carb and the Kiel device add acid to carbonate within individual reaction vessels to facilitate the digestion process. The produced gases are then purified immediately after digestion in a series of temperature-controlled, liquid-nitrogen-cooled cold fingers and the Adsorption Trap (AdTrap), an in-line, short GC column packed with Porapak Q 50/80 and silver wool, before being introduced into the mass spectrometer dual inlet. This system operates entirely under vacuum without a carrier gas for the GC column gas separation step. The column is relatively small compared with that on the autoline and is held at −30°C for the duration of the CO_2 transit. The short-length column allows only for use of relatively small samples of mostly pure carbonate.

4.1.3 | Phosphoric acid preparation and storage

Several different recipes are used for producing phosphoric acid; here we describe our laboratory procedures. Acid digestion was facilitated using 105 wt% phosphoric acid (H_3PO_4). To achieve this concentration, phosphorus pentoxide (P_2O_5) was added to a beaker of 85 wt% phosphoric acid and continuously stirred on a hot plate set at 300°C. The P_2O_5 reacted with water in 85 wt% H_3PO_4, hydrolyzing the P_2O_5 and creating more phosphoric acid. This reaction eliminated the remaining water in the acid and produced phosphoric acid molecules, thus increasing the concentration of phosphoric acid. The concentration of the acid was checked by measuring its density at room temperature. Our target density was 1.94–1.95 g/mL. A ratio of 1.25 kg of P_2O_5 to 1 L of 85% ortho-phosphoric acid produced the desired 105 wt% H_3PO_4. This high nominal concentration was achieved due to the polymerization of the phosphate under anhydrous conditions into multimeric acids such as the dimer, pyrophosphoric acid, and the trimer, triphosphoric acid. A density below the target range indicates the presence of a significant proportion of water in the acid and the need to add more P_2O_5. A density above the target range indicates an overconcentration, with formation of significant amounts of the higher multimers of the acid

phosphate, and is remedied by decreasing the hot plate temperature and allowing some water from the atmosphere to be absorbed by the acid. Once the proper density was attained, the acid was stored in Pyrex glass jars inside dehydrated desiccators for future use. Batches that are to be used immediately for digestion are maintained at 90°C on a stirring hot plate.

4.2 | Mass spectrometers

4.2.1 | Thermo 253 isotope-ratio mass spectrometer

The Thermo MAT 253 is the first commercially available high-precision gas-source, sector mass spectrometer that is sufficiently sensitive to be suitable for applications of carbonate clumped isotope thermometry.[2,23] The instrument has a dual inlet system and is specifically configured to measure multiply-substituted isotopologues of CO_2 via an array of Faraday cups capable of simultaneously measuring masses 44, 45, 46, 47, 48, and 49. Detectors for masses 44–46 are registered through 3×10^8, 3×10^{10}, and 10^{11} Ω resistors, respectively, whereas detectors for masses 47–49 are registered with 10^{12} Ω resistors. Electro-polished nickel capillaries are used to deliver sample and reference gas from the bellows to the source instead of the initially outfitted stainless-steel capillaries (Thermo Fisher Scientific).

Given the low abundance of multiply-substituted isotopologues, relatively large sample sizes and long count times are needed.[1,2,26] At the beginning of each acquisition, the gas pressures in the sample and reference bellows are adjusted to achieve a signal of 16 V across the mass 44 Faraday cup, resulting in 2–3 V across the mass 47 cup. Each sample of CO_2 gas is measured between 7 and 10 gas-working cycles. Data are taken in nine blocks of 10 cycles, with each cycle consisting of 8 s of integration and 16 s of changeover delay, for a total integration time of 720 s per sample. Each block consists of a pressure balance adjustment, peak centering, determination of background, and integration time. Measurements range from 16 to 15.7 V on mass 44 over the course of each block, with a pressure adjustment between blocks to return the beams to 16 V at the start of each block. Typical durations for sample analysis are 2–2.5 h.

4.2.2 | Nu Perspective IS

The Nu Perspective isotope-ratio mass spectrometer is a recent mass spectrometer design featuring secondary electron suppression. An energy filter fitted in front of the Faraday collector array drives the suppression and results in a linearity correction that is one to two orders of magnitude less than that on the Thermo 253. The detectors for masses 44, 45, and 46 are registered through 3×10^8, 3×10^{10}, and 10^{11} Ω resistors, respectively, whereas the channels for masses 47–49 are registered with 10^{12} Ω resistors. Detectors for masses 47–49 are shielded by secondary electron suppressors.

TABLE 2 Isotope ratios of standards used in this study

Standard	Mineralogy	Origin[a]	n	Δ47 I-CDES[b] (‰)	Δ47 I-CDES[c] (‰)	Δ47 I-CDES[c] SE	Δ47 ARF[d] (‰)	Δ47 ARF[d] SD	Δ47 ARF[d] SE	δ47 (vs. WG, ‰)	δ18O (VPDB, ‰)	δ18O SD (VPDB)	δ13C (VPDB, ‰)	δ13C SD (VPDB)
102-GC-AZ01	Calcite	Vein carbonate from Grand Canyon (CIT)	3	0.620			0.699	0.005	0.003	2.040	−14.0	0.046	0.6	0.035
Carmel Chalk (CC)	Calcite	Chalk, abbreviated as CC (CIT)	97	0.587			0.666	0.019	0.002	11.190	−4.0	0.094	−2.2	0.068
Carrara Marble (CM)	Calcite	Collected in Carrara, Tuscany, Italy; abbreviated as CM (CIT)	105	0.314			0.377	0.005	0.006	16.612	−1.6	0.039	2.0	0.016
CMTile	Calcite	Homogenized version of Carrara Marble (UCLA)	12	0.303			0.366	0.022	0.006	17.714	−1.5	0.060	2.0	0.024
Coral Std[e]	Aragonite	Deep-sea coral (CIT)	6	0.661			0.745	0.012	0.005	18.538	2.7	0.096	−1.8	0.121
ETH-1	Calcite	Carrara Marble, heated to 600°C at 155 MPa for 10 h (ETHZ)	81	0.207	0.205	0.002	0.264	0.023	0.003	16.772	−2.2	0.078	2.0	0.031
ETH-2	Calcite	Reagent-grade synthetic, subjected to same treatment as ETH-1 (ETHZ)	74	0.213	0.209	0.002	0.265	0.019	0.002	−11.936	−18.7	0.078	−10.2	0.052
ETH-3	Calcite	Calcareous chalk from northern Germany (ETHZ)	70	0.613	0.613	0.001	0.694	0.021	0.003	17.256	−1.8	0.107	1.7	0.038
ETH-4	Calcite	Same reagent grade synthetic as ETH-2 but unheated (ETHZ)	66	0.452	0.451	0.002	0.519	0.017	0.002	−11.863	−18.8	0.103	−10.2	0.034
IAEA-C1	Calcite	Carrara Marble (IAEA)	51	0.300	0.302	0.001	0.363	0.025	0.005	17.515	−2.3	0.072	2.5	0.034
IAEA-C2	Travertine	Collected in Bavaria (IAEA)	43	0.639	0.641	0.002	0.719	0.023	0.005	0.706	−8.9	0.072	−8.1	0.042
ISTB-1[e]	Calcite	Speleothem from Yichang (Hubei Province), China (CUG)	10	0.606			0.683	0.041	0.013	−1.515	−8.5	0.023	−10.7	0.049
MallinckrodtCal	Calcite	Synthetic, from Mallinckrodt Baker, Inc.	13	0.464			0.526	0.039	0.011	−39.128	−21.9	0.112	−40.4	0.091
MERCK	Calcite	Synthetic (IAEA)	48	0.522	0.514	0.002	0.588	0.028	0.006	−39.100	−15.6	0.065	−42.0	0.039
NBS 19	Calcitic marble	Carrara Marble (NBS)	7	0.314			0.378	0.024	0.009	22.075	−2.0	0.119	1.9	0.066
Spel 2-8-E	Calcite	Speleothem from Yichang (Hubei Province), China (CIT)	10	0.611			0.689	0.031	0.010	1.406	−6.4	0.101	−9.2	0.165

(Continues)

TABLE 2 (Continued)

Standard	Mineralogy	Origin[a]	n	Δ_{47} I-CDES[b] (‰)	Δ_{47} I-CDES[c] (‰)	Δ_{47} I-CDES[c] SE	Δ_{47} ARF[d] (‰)	Δ_{47} ARF[d] SD	Δ_{47} ARF[d] SE	δ_{47} (vs. WG, ‰)	$\delta^{18}O$ (VPDB, ‰)	$\delta^{18}O$ SD (VPDB)	$\delta^{13}C$ (VPDB, ‰)	$\delta^{13}C$ SD (VPDB)
SRM 88B	Dolomitic limestone	Collected from mine site near Skokie, Illinois, USA (NIST)	10	0.499			0.573	0.007	0.002	13.124	−7.4	0.091	2.2	0.017
TB-1[e]	Calcitic marble	Marine-sourced marble from Quyang (Hebei Province), China (CUG)	4	0.308			0.369	0.009	0.004	5.438	−11.6	0.344	1.8	0.120
TB-2[e]	Calcite	Hydrothermally formed calcite from Yanji city (Jilin Province), China (CUG)	2	0.322			0.380	0.036	0.026	−14.746	−24.4	0.014	−5.9	0.099
TV01	Calcite	Travertine tile (CIT)	3	0.641			0.723	0.011	0.006	9.977	−8.2	0.191	2.7	0.125
TV03	Calcite	Travertine tile (CIT)	56	0.620			0.700	0.015	0.014	11.166	−8.2	0.038	3.1	0.034
VeinStrom	Vein calcite	Shallow carbonate vein collected from Tempiute Mountain, Nevada (UCLA)	95	0.633			0.712	0.019	0.002	−1.510	−12.6	0.092	−6.2	0.054

Notes. Mean clumped (Δ_{47}) and bulk ($\delta^{13}C$, $\delta^{18}O$, and δ_{47}) values are determined using mean values from the Nu Perspective + 90°C common acid bath (configuration 2, analysis B), with eight standards used for constructing the empirical transfer function. Δ_{47} values for standards are shown both on the new I-CDES scale (Intercarb - Carbon Dioxide Equilibrated Scale) and ARF (Absolute Reference Frame). Abbreviations for standards are shown in parentheses.

Abbreviations: CIT, California Institute of Technology; I-CDES, InterCarb-Carbon Dioxide Equilibrium Scale; SE, standard error; SD, standard deviation.

[a]CIT, Eiler Lab, Caltech; ETHZ, Bernasconi Lab, ETH Zurich; UCLA, Tripati Lab, UCLA; CUG, China University of Geosciences; IAEA, International Atomic Energy Agency; NBS, National Bureau of Standards; NIST, National Institute of Standards and Technology.

[b]Calculated using standardization procedures for ETH-1, ETH-2, ETH-3, and ETH-4, and the scaling procedure from Bernasconi et al[21]: $^{new}\Delta_{47} = 0.048529 - 0.0001165 \times \delta_{47} + 0.944081 \times ^{old}\Delta_{47}$ (equation Y), where $^{old}\Delta_{47}$ refers to the Δ_{47} values from analysis B of this study. Projected to 90°C acid reaction using acid correction factor of $-0.088‰$ (Petersen et al[19]) following Bernasconi et al.[21]

[c]Values from Bernasconi et al[21] projected onto the 90°C reference frame using acid correction factors of $-0.088‰$ (Petersen et al[19]).

[d]Calculated using the values for ETH-1, ETH-2, ETH-3, and ETH-4 from Bernasconi et al[17] and on the ARF.

[e]Standards were not run on this instrumental configuration (configuration 2, analysis B), and thus mean values are calculated using the average of values determined in this study using other instrumental configurations.

Large samples (>3 mg pure $CaCO_3$) are measured in bellows mode, with four blocks of 20 cycles for a total of 80 cycles of sample-standard comparison with an 8 s changeover delay and 20 s of integration per cycle for a total integration time of 1600 s. There is continuous pressure adjustment using a Newtonian zeroing technique, with balancing to achieve 16 V (or 80 nA) on mass 44 at every acquisition rather than only at the beginning of the block.

Small samples (0.25–0.5 mg pure $CaCO_3$) are measured in microvolume mode, with precisely matched sample and working gas volumes, allowing the gas to deplete at precisely matched rates and increasing the efficacy of gas usage. The microvolumes allow for a full hour-long measurement to take place on a single sample. Data are taken in three blocks of 20 cycles, with each cycle consisting of 20 s of integration and 8 s of changeover delay, for a total integration time of 1200 s per sample. Measurements range from 80 to 30 nA on mass 44 over the course of each sample's acquisition.

4.3 | Standards

We measured 21 carbonate standards from four machine configurations and analyzed data using two methods. Standards and data for this study were measured from 2013 to 2019. Our working gas was obtained from Oztech (Safford, AZ, USA) and has a typical $\delta^{13}C$ value of -3.6‰ (VPDB) and a typical $\delta^{18}O$ value of 25.0‰ (VSMOW). Carbonate standard names and abbreviations are presented in Table 2. Using configurations 1 and 2, the CAB systems were used to measure equilibrated gas standards, typically daily. We use a 10-standard moving average, with multiple standards typically analyzed each day.

4.3.1 | Equilibrated gas standards

We utilize four endmember equilibrated gases that differ in their δ_{47} and Δ_{47} values. Gases with two distinct δ_{47} compositions that differ by ~60‰ are used for linearity correction. The two gases bracket the δ_{47} of our Oztech IRMS working gas and our samples. The low δ_{47} endmember consists of gas drawn from a factory (Airgas, Radnor, PA, USA) CO_2 gas cylinder, equilibrated with 5–10 mL of DI ("deionized," with no charge) water at 25°C in a water bath. The heavy isotope-enriched δ_{47} endmember consists of gas produced by acid digestion of a Carrara Marble carbonate standard and 105% phosphoric acid, which is subsequently equilibrated with evaporated DI water in the same bath. Our evaporated DI water is produced by boiling DI water for 2 days. Water bath equilibrations are carried out for at least 3 days before use, and these gases are referred to as unheated gases. Unheated gases are stored in Pyrex or quartz breakseals.

Aliquots of the two unheated gases are equilibrated at a second temperature, 1000°C, to randomize the distribution of isotopes among isotopologues, and theoretically they should differ in their Δ_{47} values from the unheated 25°C equilibrations by 0.9‰.[16,41] This equilibration is performed by heating aliquots of cryogenically purified

25°C gas that are in quartz breakseals using a muffle furnace set to 1000°C for at least 2 h followed by flash cooling to room temperature.

4.3.2 | Carbonate standards

Data for 20 calcite standards and 1 dolomite standard are reported (Table 2). These include ETH-1, ETH-2, ETH-3, ETH-4, IAEA-C1, IAEA-C2, and MERCK, as described in Bernasconi et al,[17] and NBS-19, SRM 88B (dolomite), Mallinckrodt Calcium Carbonate, 102-GC-AZ01, Carmel Chalk, Carrara Marble, CMTile (another homogenized Carrara Marble), Spel 2-8-E, TV01, TV03, and VeinStrom. Carbonate standards obtained from the California Institute of Technology (CIT) and China University of Geosciences (CUG) include Coral Std, ISTB-1, TB-1, and TB-2.[42] These carbonates span a range of δ_{47} that is more than 30‰ and ~0.5‰ in Δ_{47}. Raw data and analyzed results are provided in the supporting information.

4.4 | Sample-standard bracketing

Typically, nine analyses are conducted each day on our CAB systems, of which three to four are carbonate standards. From 2013 to 2016, we cycled through seven carbonate standards (a combination of 102-GC-AZ01, Carmel Chalk, Carrara Marble, and the ETH standards) (Figure 7; Figures S2 and S4 [supporting information]). From 2017 to the present, we have cycled through 12 carbonate standards, including a combination of Carmel Chalk, Carrara Marble, CMTile, TV03, and VeinStrom, with a focus on the InterCarb standards (ETH-1, ETH-2, ETH-3, ETH-4, IAEA-C1, IAEA-C2, and MERCK) (Figure 7; Figures S2 and S4 [supporting information]). As of 2017, daily runs have been initiated with an equilibrated gas standard, followed by one of the carbonate standards mentioned, one to two additional carbonate standards midway through the run, and ending with a carbonate standard. Immediately after machine downtime or maintenance, all analyses conducted are of equilibrated gas and carbonate standards to reestablish reference frames. Sample analysis is resumed once ~10 standards measured at "expected" values are obtained, to more robustly construct a nonlinearity correction with a moving average of 10 standards on either side of a given standard, for a 20-point moving average.

For the Nu Carb sample preparation systems, we measure only carbonate standards, with 14–15 analyses per day, of which 5–6 are carbonate standards. From 2017 to the present, we have cycled through 12 carbonate standards, including a combination of Carmel Chalk, Carrara Marble, CMTile, TV03, and VeinStrom, with a focus on the InterCarb standards (ETH-1, ETH-2, ETH-3, ETH-4, IAEA-C1, IAEA-C2, and MERCK) (Figure 7; Figures S2 and S4 [supporting information]). Daily sequences are initiated with two of the carbonate standards mentioned, one to two additional carbonate standards midway through the run, and ending with two carbonate standards. Following machine downtime or maintenance, all analyses are of

carbonate standards. The same construction of the nonlinearity correction is followed as is applied to the CAB systems, where sample analysis is resumed once ~10 standards measured at "expected" values are obtained so as to correct using a 20-point moving average.

To ensure a range of standards with different values that should bracket most samples, we cycle through each of these aforementioned carbonate standards on all configurations (see Figure 1 for δ_{47} and Δ_{47} ranges). Outliers in Δ_{47} are excluded at the 3σ level (0.075‰), greater than the recommended screening factor of 0.06‰ used in Meckler et al.[43]

4.4.1 | Corrections

Over 2 years, several laboratory group members entered data from 6 years of instrumental analyses into Easotope.[30] Standardization intervals are primarily identified based on times when the source was cleaned due to a contaminant, the source was retuned, or there were other major changes (e.g., a digestion system or gas chromatograph being repaired). All data are reported using the IUPAC parameter set[29,32] on the absolute reference frame[16] so as to reduce internal differences that are associated with various standardizations, methodologies, corrections, and calibrations.[17,19]

We analyzed data in several ways over multiple years; this study presents the results of two sets of analyses that are most relevant to the community, referred to as analyses A and B. These two analyses differ in the number of standards used for drift corrections and transfer functions, termed "working standards" throughout this study (three and eight, respectively; see Table S2 [supporting information]). The remaining standards, processed as unknowns, have been termed "consistency standards."
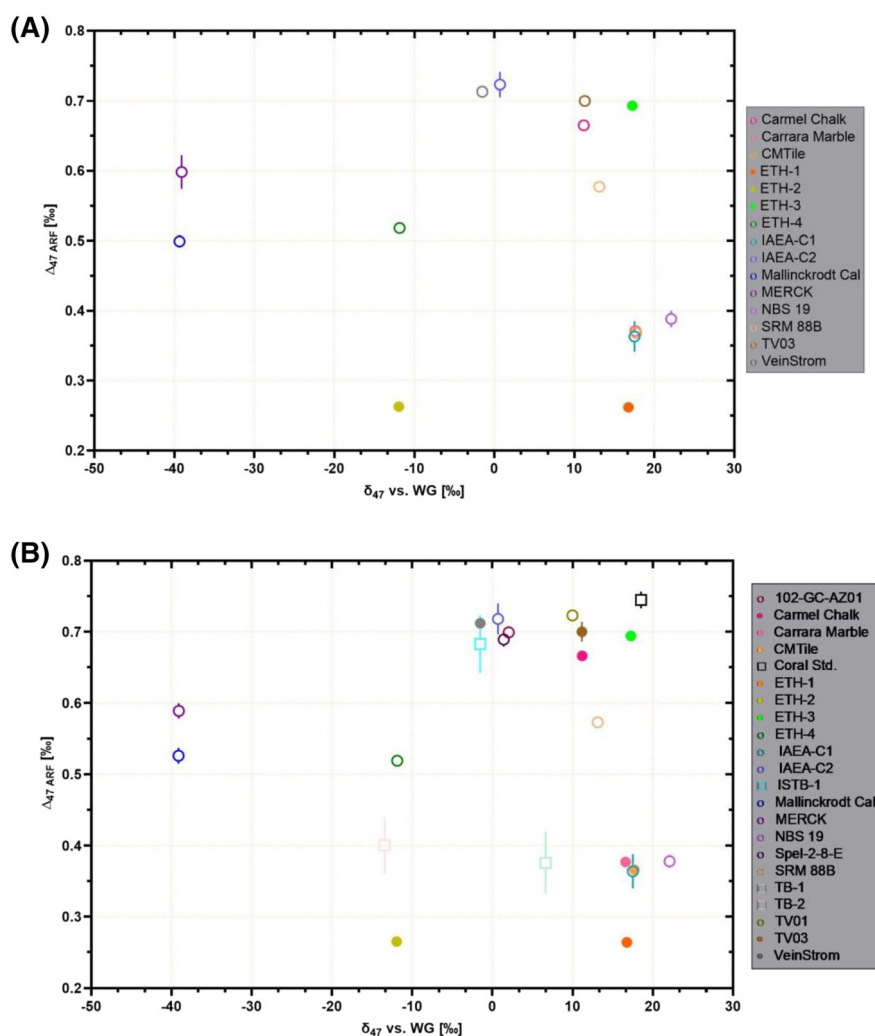

FIGURE 1 Mean Δ_{47} (‰; CDES [Carbon Dioxide Equilibrium Scale]) and standard error versus δ_{47} (‰; WG) for 21 standards (filled symbols, working standards; open symbols, consistency standards) measured in the Tripati Lab. A, analysis A (three working standards); B, analysis B (eight working standards)

For analysis A, $\delta^{13}C$ and $\delta^{18}O$ drift corrections utilize ETH-1, ETH-2, and ETH-3 as working standards. Linearity corrections for all instrument configurations are based on measurements of the carbonate standards ETH-1 and ETH-2 that were produced during heating experiments by the ETH Zürich group, with values from Bernasconi et al.[17] For instrument configurations that use a CAB system (configurations 1 and 2), linearity corrections use equilibrated gases in addition to carbonate standards ETH-1 and ETH-2. The transfer functions used for analysis A are based on the proposal from Bernasconi et al[17] as well as recommendations from Kocken et al[44] to use values from three carbonate standards (ETH-1, ETH-2, and ETH-3) for corrections; these values are taken from Bernasconi et al[17] to evaluate the reproducibility of the remaining carbonate standards (Table S2 [supporting information]).

For analysis B, we assess the impact of utilizing a larger number of carbonate standards as working standards for corrections. $\delta^{13}C$ corrections utilize ETH-1, ETH-2, ETH-3, Carmel Chalk, Carrara Marble, CMTile, and VeinStrom, whereas $\delta^{18}O$ drift corrections utilize these and also TV03. Linearity corrections are the same as for analysis A. Transfer functions utilize eight standards: ETH-1, ETH-2, ETH-3, Carmel Chalk, Carrara Marble, CMTile, VeinStrom, and TV03 (Table S2 [supporting information]).

4.5 | Statistical methods

Statistical analyses were performed in R version 4.1.0.[45] To assess whether there were differences in measurements between machines, analyses, or configurations, we created linear mixed effects models for repeated measures in package *nlme* version 3.1-152[46] followed by adjustment for multiple comparisons and confidence interval estimation in package *emmeans* version 1.6.0.[47] Pooled degrees of freedom (dfs) are reported from adjustments in *emmeans*; for all models, df = 98 for pairwise comparisons of individual standards by analysis and configuration except where noted, specifically, Carmel Chalk in analysis A for configuration 1, where df = 121. For analysis by configuration comparisons, the pooled df = 6992. Several standards (ETH-1, ETH-2, ETH-3, and ETH-4) were measured from multiple aliquots of the standard. No differences were found between aliquots, and they were therefore pooled for analyses. Analytical uncertainty was included in models as a random effect in the form of the measurement standard error. Package *emmeans* implements a marginal mean calculation that differs slightly (~0.005‰) from the more typical weighted means approach. However, any discrepancies are within the bounds of calculated uncertainties.

To assess precision, we implemented a linear regression approach using a 75/25 training/test split of observed Δ_{47} values with prediction post resampling in package *caret* version 6.0-88.[48] The three CIT standards, ISTB-1, TB-1, and TB-2, were excluded, as they are rarely measured in our laboratory and do not provide accurate information about instrumental precision. As findings indicate that there is no measurable difference in standard values between analyses A and B (see Section 5.4), we chose to consider linear

models, including standard only as a predictor of Δ_{47} in addition to standard plus configuration, standard plus analysis, and standard plus configuration and analysis. Rosner's test for outliers in measurement error was performed in package *EnvStats* version 2.4.0.[49] The data and R code used in analyses are publicly available at https://github.com/Tripati-Lab/Upadhyay-et-al.

5 | RESULTS AND DISCUSSION

5.1 | Baselines and linearity corrections

Figure 2 shows baselines for the different instrument configurations. Our Thermo MAT 253 has a more pronounced negative baseline than the Nu Perspective, which results in a linearity correction for Δ_{47} with a relatively steep slope in the third decimal place. The Nu Perspective IS has lower baselines and relatively flat linearity corrections for Δ_{47} that vary in the fourth decimal place. This is because the presence of an energy filter and quadratic lenses fitted in front of the Faraday collector arrays suppresses secondary electrons, resulting in an almost-negligible linearity correction.

For both types of instruments, we found evidence for baseline stability over timescales of a month or greater. Monitoring of the pressure baseline on each instrument followed a procedure similar to that outlined in Meckler et al.[43] To determine peak shapes, the accelerating voltage is changed, and the resultant beam intensities are recorded. For the Thermo MAT 253, weekly/monthly scans show a long-term temporal stability between consecutive samples/standards within a correction interval, which are typically 1 month or longer in length. Weekly peak scans at *m/z* 44 intensities ranging from 10 to 20 V on the MAT 253 yield baselines that ranged from −5 to −40 mV for *m/z* 47 (Figure 2). For the Nu Perspective IS, bimonthly peak scans show relative stability typically over a 3 month or longer interval. Scans at *m/z* 44 yielded intensities ranging from 40 to 100 nA (equivalent to 12–30 V) on the Nu Perspectives, with baselines around 0.00005 nA for *m/z* 47 (equivalent to 50 mV) (Figure 2).

Figure 3 shows the nonlinearity corrections corresponding to the dates of analyses shown in Figure 2. Due to the higher value of the baselines, measurements from the Thermo 253 require a significant linearity slope correction for Δ_{47} (Figure 3). For the Nu Perspective, there is secondary electron suppression on the *m/z* 47 cup, which means that there should not be a large nonlinearity correction for Δ_{47} (Figure 3).

5.2 | Use of published values

To report the figures of merit for carbonate-based standardization, we present the consistency standard values in Tables S3 and S4 (supporting information). Both are on the absolute reference frame and use the IUPAC parameter set (Brand et al[32]). The values in Table S3 (supporting information) were calculated using ETH-1, ETH-2, and ETH-3 as working standards (analysis A), with bulk and
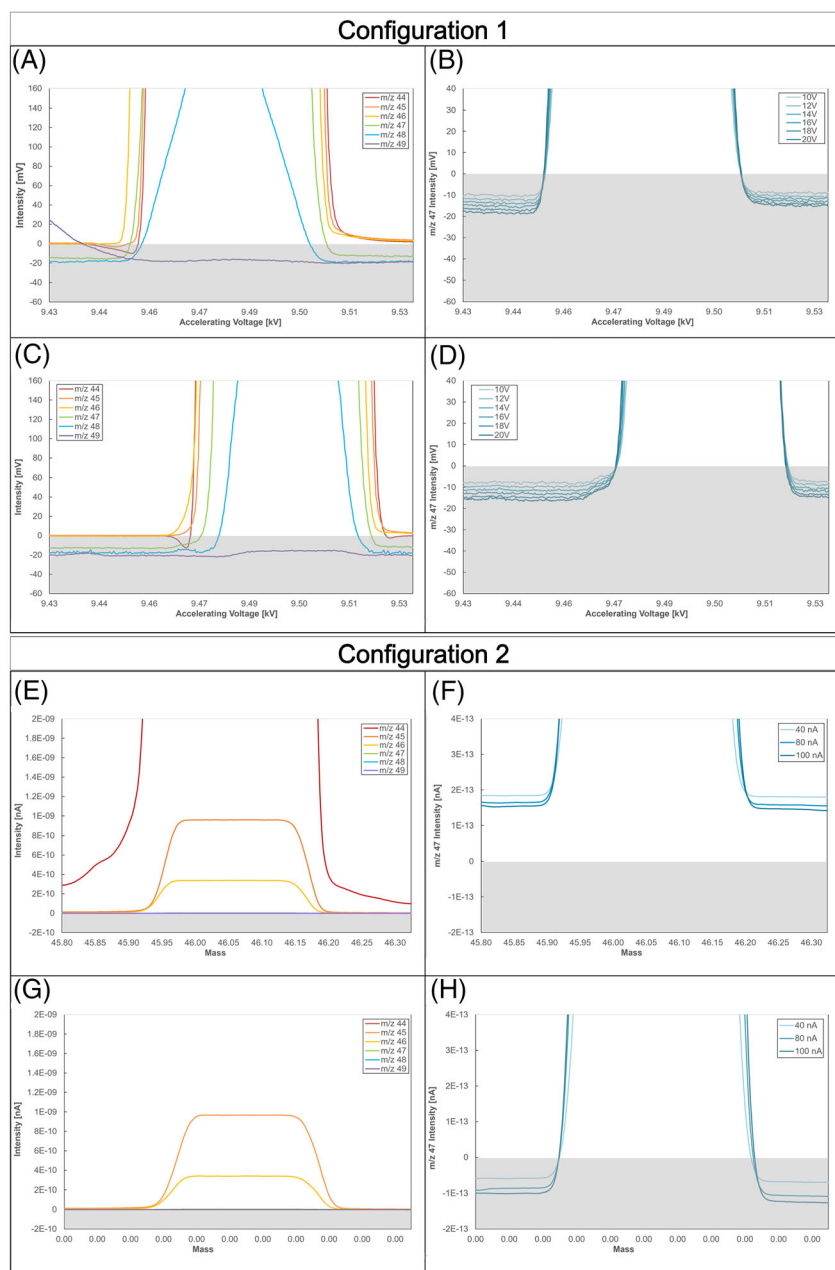
FIGURE 2 Pressure baseline variations over the course of 2014–2017 on configuration 1 (Thermo 253) and configuration 2 (Nu Perspective) in the Tripati Lab. Negative values are indicated by gray shading. On the left is an expanded view of the lowermost section of the *m/z* 44–49 peaks, with *m/z* 44 at 16 V. On the right, *m/z* 47 is shown, with *m/z* 44 signal corresponding to 10, 12, 14, 16, 18, and 20 V for configuration 1 and to 40, 80, and 100 nA for configuration 2. A and B, configuration 1 on October 4, 2014, C and D, configuration 1 on August 1, 2017, E and F, configuration 2 on October 12, 2015, and G and H, configuration 2 on January 18, 2016

clumped isotope values from Bernasconi et al.[17] For analysis B (Table S4 [supporting information]), the ETH standards are defined using the Bernasconi et al[17] values and the remaining working standards from analysis A. Standard values are also recalculated on the absolute reference frame using updated values from Bernasconi et al[21] (Table 2).

5.3 | Long-term accuracy and precision of clumped isotope data from multiple instruments

A comparison of standard values for analyses A and B is reported in Tables S3 and S4 (supporting information) so as to assess machine intercomparability and the impact of using different numbers of

working standards for reference frame calculations. We analyzed these results in Table S3 (supporting information) to examine whether there are differences in the accuracy and precision of clumped isotope data resulting from acid digestion, mass spectrometer, the number of working standards used, and the use of equilibrated gases for reference frame calculations. The standards chosen for comparisons in Table S3 (supporting information) were selected because they represent consistency standards (not working standards/anchors) for their respective configurations. They were also selected because they were replicated at least thrice (similar to samples) and because they span a broad compositional space. Other standards were not feasible because either they were working standards/anchors or at the time of the analysis they had not been sufficiently replicated. The use of a different set of standards as

FIGURE 3 Δ_{47} and Δ_{48} nonlinearity corrections corresponding to the dates of peak scans conducted in Figure 2, for configuration 1 (Thermo 253) and configuration 2 (Nu Perspective) in the Tripati Lab. Correction dates are as follows: A, configuration 1 Δ_{47} nonlinearity from October 4, 2014; B, configuration 1 Δ_{48} nonlinearity from October 4, 2014; C, configuration 1 Δ_{47} nonlinearity from August 1, 2017; D, configuration 1 Δ_{48} nonlinearity from August 1, 2017; E, configuration 2 Δ_{47} nonlinearity from October 12, 2015; F, configuration 2 Δ_{48} nonlinearity from October 12, 2015; G, configuration 2 Δ_{47} nonlinearity from January 18, 2016; and H, configuration 2 Δ_{48} nonlinearity from January 18, 2016
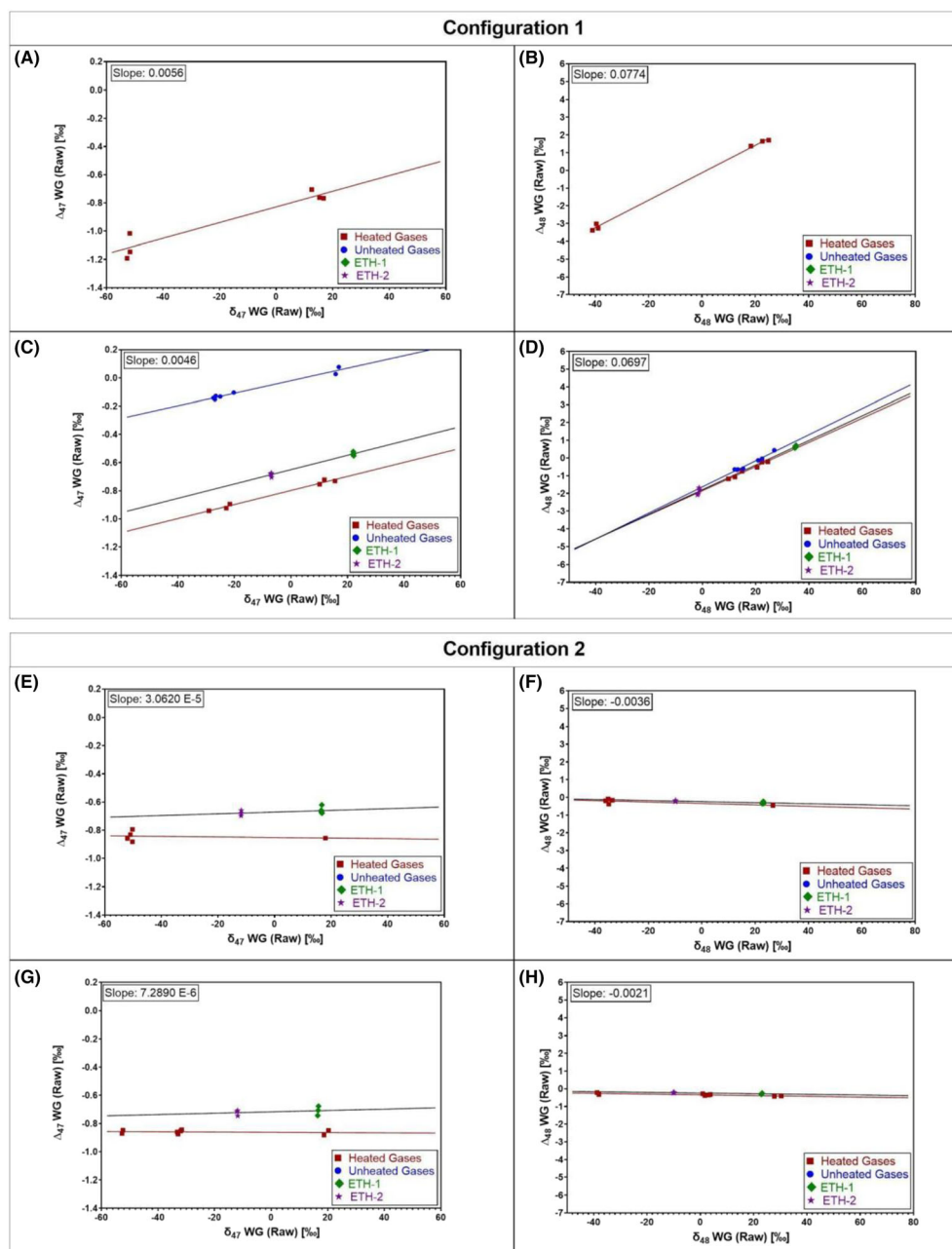
working standards/anchors yielded similar conclusions; this can be deduced in the comparison of analysis A (using three working standards/anchors) with analysis B (using eight anchors/working standards), as discussed in Section 5.4.

To assess accuracy, we compare standard Δ_{47} values with those from the Nu Perspective IS with a 90°C CAB (configuration 2). We assume this to be the most accurate of the configurations investigated for five reasons:

1. A relatively small linearity correction for Δ_{47} due to secondary electron suppression.
2. The use of equilibrated gas standards in addition to carbonate standards allows for values to be directly tied to the absolute

reference frame that is based on the Carbon Dioxide Equilibrium Scale (CDES)[16] and provides an additional check on the origin of drift.

3. Larger samples that should minimize the impact of any heterogeneity in measured materials are analyzed.
4. Larger samples facilitate long counting times (Table 1).
5. Stable and intense ion beams (80 nA or ~16 V for mass 44) with high signal-to-noise ratio.

Standard Δ_{47} values are presented in Tables S3 and S4 (supporting information); Figure 4 and Figure S1 (supporting information) provide a visual representation of data presented in these tables. To assess precision, we considered models for Δ_{47} that comprised standard
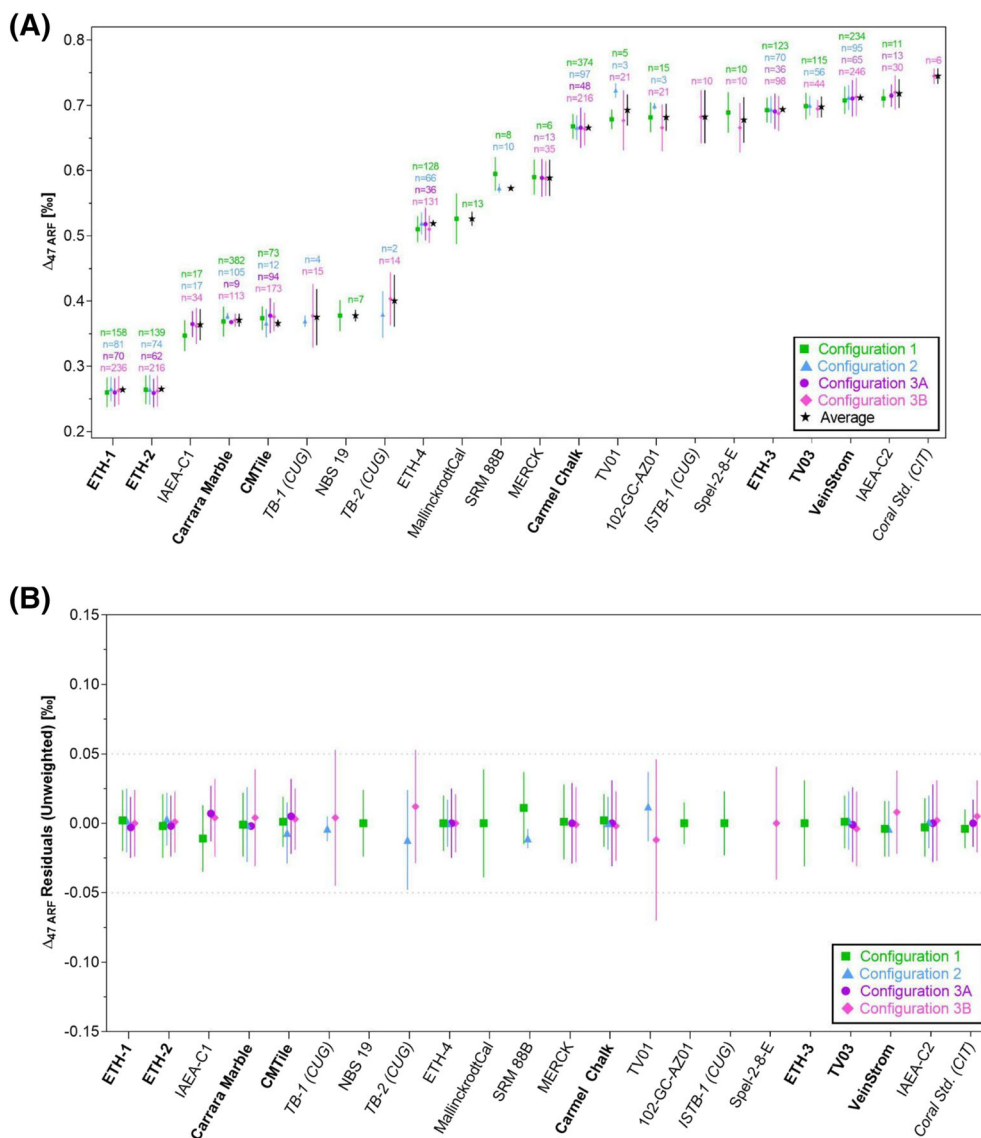
FIGURE 4 Mean standard Δ_{47} CDES (Carbon Dioxide Equilibrium Scale) and residuals for each instrumental configuration in the Tripati Lab determined from analysis B. A, Mean standard Δ_{47} and standard deviation (bold = working standards; n is indicated next to each standard). Also shown is the average Δ_{47} from all configurations (black star). B, Unweighted Δ_{47} residuals and standard deviations

alone, the additive effects of standard plus configuration, the additive effects of standard plus analysis, and the additive effects of standard plus configuration and analysis.

5.4 | Comparison of mean values across analyses A and B (using three vs. eight working standards) for the Nu Perspective IS with CAB

A comparison of analyses A and B allows us to investigate the effect of using a larger number of working standards with the CAB system paired with the Nu Perspective IS (configuration 2), the instrumental configuration that we assume is the most robust. Both analyses use ETH-1, ETH-2, and ETH-3 as working standards. In addition, we use the values of five consistency standards determined during analysis A (Carmel Chalk, Carrara Marble, CMTile, TV03, and VeinStrom) to define the values for the remaining consistency standards in B.

The mean Δ_{47} values are within error of each other for all instrumental configurations, and the number of standards used for

drift and transfer function corrections makes a negligible difference in both the mean and the reproducibility of clumped data (Table 3; Tables S3–S5 [supporting information]).

We find no evidence of a statistically significant difference between the two analyses for either standard. This is shown via a comparison of the "difference", "SE," "t-ratio," and "P-value" between these analyses, where these terms are defined as below throughout the remainder of the text:

- "Difference" refers to the difference between the estimated marginal mean Δ_{47} values for a standard from a particular configuration (configuration 2, in this particular case) and the estimated marginal mean Δ_{47} value of that standard in configuration 2.
- "SE" refers to the standard error around the pairwise estimate for the given configurations.
- "t-ratio" refers to the ratio comparing the departure of a given estimate from its hypothesized value divided by the standard error around the estimate.

TABLE 3 Compilation of mean Δ_{47}, Δ_{47} standard error, 95% confidence intervals, n, and absolute difference in Δ_{47} values for consistency standard comparisons discussed in Sections 5.4–5.7.

Corresponding section	Standard	Analysis and configuration	Δ_{47} ARF (‰)	95% CI	n	Absolute difference in Δ_{47} ARF (‰)
Section 5.4: Comparison of mean values across analyses A and B (using three vs. eight working standards) for Nu Perspective IS with CAB	ETH-4	Analysis A, configuration 2	0.518 ± 0.017	0.583–0.542	69	0.001
		Analysis B, configuration 2	0.519 ± 0.017	0.540–0.544	66	
	SRM 88B	Analysis A, configuration 2	0.577 ± 0.005	0.583–0.542	10	0.004
		Analysis B, configuration 2	0.573 ± 0.007	0.540–0.544	10	
		Difference = −0.00142, SE = 0.0008, t-ratio = −1.624, P = 1				
Section 5.5.1: No evidence for differences in accuracy and precision between acid digestion systems, acid temperatures, and small vs. large samples	Carmel Chalk	Analysis A, configuration 2	0.665 ± 0.018	0.660–0.670	95	0.007
		Analysis A, configuration 3A	0.658 ± 0.022	0.659–0.670	33	
	ETH-4	Analysis A, configuration 2	0.518 ± 0.017	0.508–0.518	69	0.005
		Analysis A, configuration 3A	0.513 ± 0.022	0.507–0.517	23	
		Difference = 0.0025, SE = 0.002, t-ratio = 0.206, P = 0.9				
	Carmel Chalk	Analysis A, configuration 2	0.665 ± 0.018	0.660–0.670	95	0.001
		Analysis A, configuration 3B	0.664 ± 0.024	0.663–0.673	129	
	ETH-4	Analysis A, configuration 2	0.518 ± 0.017	0.508–0.518	69	0.002
		Analysis A, configuration 3B	0.516 ± 0.021	0.511–0.521	82	
		Difference = 0.0012, SE = 0.002, t-ratio = 0.729, P = 1				
Section 5.5.2: No evidence for difference in accuracy using the Nu Perspective vs. Thermo 253	Carmel Chalk	Analysis A, configuration 1	0.662 ± 0.021	0.654–0.664	68	0.003
		Analysis A, configuration 2	0.665 ± 0.018	0.660–0.670	95	
	ETH-4	Analysis A, configuration 1	0.499 ± 0.021	0.502–0.512	69	0.019
		Analysis A, configuration 2	0.518 ± 0.017	0.508–0.518	69	
		Difference = −0.007, SE = 0.002, t-ratio = −3.889, P = 0.005				
		Adjusted: difference = −0.007, SE = 0.0018, t-ratio = −3.889, P = 0.37				
	Carmel Chalk	Analysis A, configuration 1	0.662 ± 0.021	0.654–0.664	68	0.002
		Analysis A, configuration 3B	0.664 ± 0.024	0.663–0.673	129	
	ETH-4	Analysis A, configuration 1	0.499 ± 0.021	0.502–0.512	69	0.017
		Analysis A, configuration 3B	0.516 ± 0.021	0.511–0.521	82	
	TVO3	Analysis A, configuration 1	0.687 ± 0.026	0.695–0.707	35	0.083
		Analysis A, configuration 3B	0.770 ± 0.030	0.704–0.716	15	
		Difference = −0.006, SE = 0.002, t-ratio = −3.43, P = 0.02				
		Adjusted: difference = −0.006, SE = 0.002, t-ratio = −3.43, P = 0.7				
Section 5.6.1: No evidence for differences in accuracy and precision between acid digestion systems and acid temperatures on the Nu	ETH-4	Analysis B, configuration 2	0.519 ± 0.017	0.508–0.518	66	0.001
		Analysis B, configuration 3A	0.518 ± 0.025	0.508–0.517	36	
		Difference = 0.001, SE = 0.002, t-ratio = 0.583, P = 1				
	ETH-4	Analysis B, configuration 2	0.519 ± 0.017	0.508–0.518	66	0.009
		Analysis B, configuration 3B	0.510 ± 0.021	0.508–0.518	131	
		Difference = 0.0009, SE = 0.002, t-ratio = 0.578, P = 1				

(Continues)

TABLE 3 (Continued)

Corresponding section	Standard	Analysis and configuration	Δ_{47} ARF (‰)	95% CI	n	Absolute difference in Δ_{47} ARF (‰)
Section 5.6.2: Slight difference in accuracy using the Nu Perspective vs. Thermo 253	ETH-4	Analysis B, configuration 1	0.510 ± 0.020	0.507–0.516	128	0.009
		Analysis B, configuration 2	0.519 ± 0.017	0.508–0.518	66	
	SRM-88B	Analysis B, configuration 1	0.595 ± 0.026	0.577–0.595	8	0.022
		Analysis B, configuration 2	0.573 ± 0.007	0.578–0.596	10	
		Difference = −0.004, SE = 0.001, t-ratio = −3.55, P = 0.003				
	ETH-4	Analysis B, configuration 1	0.510 ± 0.020	0.507–0.516	128	0
		Analysis B, configuration 3B	0.510 ± 0.021	0.508–0.517	131	
	IAEA-C1	Analysis B, configuration 1	0.347 ± 0.024	0.352–0.365	17	0.015
		Analysis B, configuration 3B	0.362 ± 0.028	0.353–0.366	34	
		Difference = −0.003, SE = 0.001, t-ratio = −2.924, P = 0.02				
Section 5.7: Comparison of mean values across carbonate-based standardization vs. gas-and-carbonate-based standardization	**Analysis A**					
	See data for Section 5.5.1 for comparison between configuration 2 and configuration 3A/3B					
	See data for Section 5.5.2 for comparison between configuration 1 and configuration 3B, in conjunction with configuration 1 vs. 3A					
	Carmel Chalk	Analysis A, configuration 1	0.662 ± 0.021	0.654–0.664	68	0.004
		Analysis A, configuration 3A	0.658 ± 0.022	0.659–0.670	33	
	ETH-4	Analysis A, configuration 1	0.499 ± 0.021	0.502–0.512	69	0.014
		Analysis A, configuration 3A	0.513 ± 0.022	0.507–0.517	23	
		Difference = −0.004, SE = 0.002, t-ratio = −2.168, P = 0.4				
	Analysis B					
	See data for Section 5.6.1 for comparison between configuration 2 and. 3A/3B					
	See data for Section 5.6.2 for comparison between configuration 1 and configuration 3B, in conjunction with configuration 1 vs. 3A					
	Carmel Chalk	Analysis B, configuration 1	0.510 ± 0.020	0.507–0.516	128	0.008
		Analysis B, configuration 3A	0.518 ± 0.025	0.508–0.517	36	
	ETH-4	Analysis B, configuration 1	0.347 ± 0.024	0.352–0.365	17	0.018
		Analysis B, configuration 3A	0.365 ± 0.020	0.353–0.367	17	
		Difference = −0.004, SE = 0.002, t-ratio = −0.258, P = 1				

Notes. Statistical parameters mentioned include "difference," "SE," "t-ratio," and "P-value" and are defined as follows.

"Difference": difference between the estimated marginal mean Δ_{47} values for a standard from a particular configuration and the estimated marginal mean Δ_{47} value of that standard in configuration 2; "SE": standard error around the pairwise estimate for the given configurations; "T-ratio": ratio comparing the departure of a given estimate from its hypothesized value divided by the standard error around the estimate; "P-value": probability that the null hypothesis (of there being no difference between the results of two analyses or configurations) is true and is interpreted by a comparison with the t-ratio.

Abbreviation: CI, confidence interval.

- "*P*-value" refers to the probability that the null hypothesis of there being no difference between the results of two analyses or configurations is true and is interpreted by a comparison with the aforementioned "t-ratio". A *P*-value <0.05 indicates that there is sufficient evidence to reject the null hypothesis and accept the alternative hypothesis that two analyses or configurations show statistically significant difference in their values at the α = 0.05 level.

Comparing analyses A and B, we derive a difference = -0.00142, SE = 0.0008, t-ratio = -1.624, and $P = 0.1$ (analysis A 95% CI: 0.5384–0.5423; analysis B 95% CI: 0.5399–0.5437) (Table 3). For full summary statistics and confidence intervals produced by *emmeans*[47] for each standard across all instruments and configurations, see Table S6 (supporting information).

5.5 | Comparison of mean values across machine configurations in analysis A

For analysis A, we examine the impact that acid digestion system, mass spectrometer model, and the use of a mixed gas-and-carbonate standardization scheme have on the reproducibility of consistency standards.

5.5.1 | No evidence for differences in accuracy and precision between acid digestion systems, acid temperatures, and small versus large samples

We first assess the effect of acid digestion procedure, acid temperature, and sample size by comparing the results from the 90°C CAB system (configuration 2, AFF = 0.082[15]) with large samples with
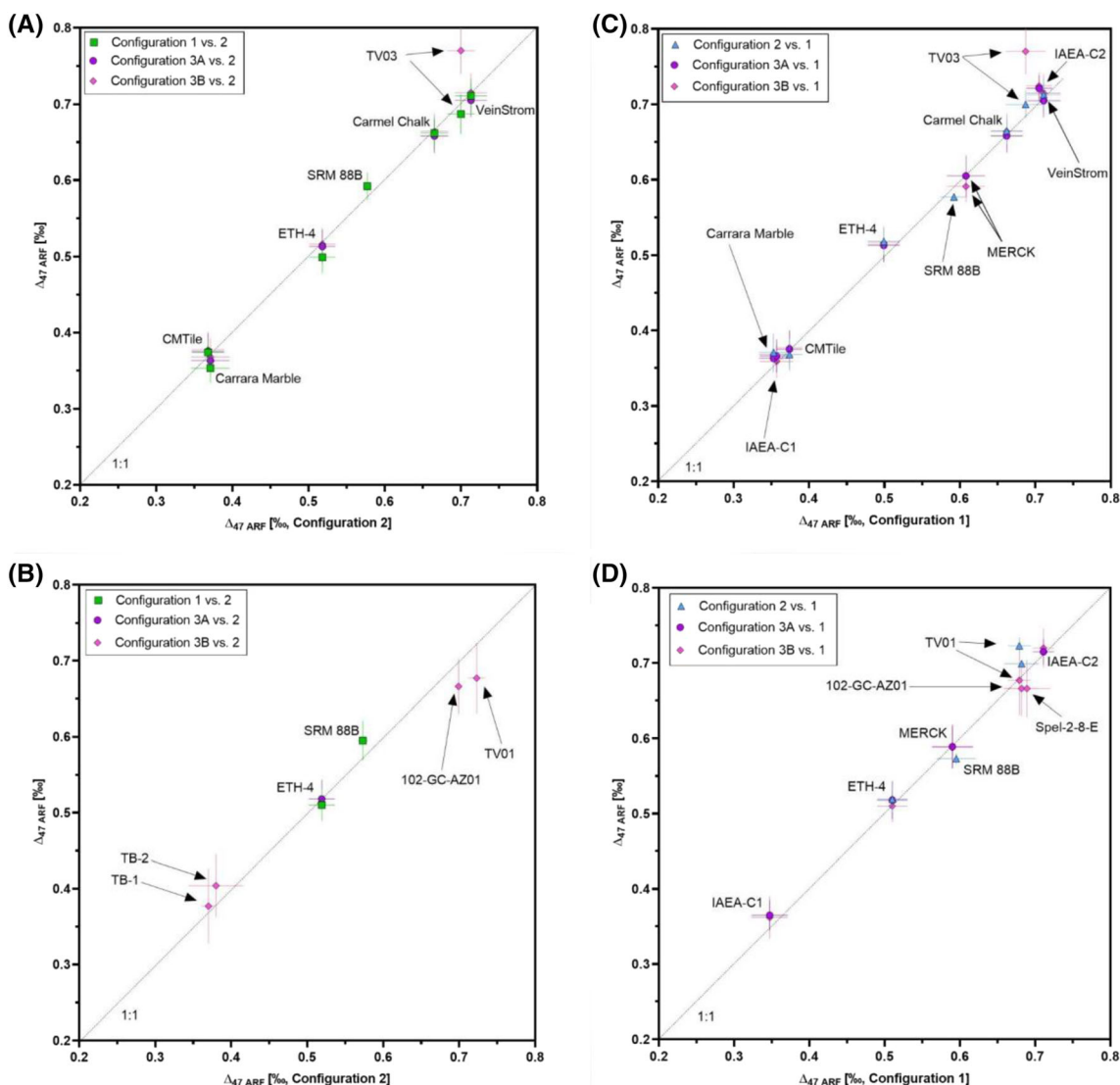

FIGURE 5 Comparison of mean Δ_{47} values (‰; CDES [Carbon Dioxide Equilibrium Scale]) for consistency standards determined on four different instrumental configurations in the Tripati Lab, with standard groupings labeled. Also shown is a 1:1 line. Left panels have configuration 2 as the *x*-axis; right panels have configuration 1 as the *x*-axis. A and C, analysis A. B and D, analysis B. All but one standard (TV03) yielded statistically indistinguishable values in analysis A; all standards yielded statistically indistinguishable values in analysis B

those from the 70°C Nu Carb system (configuration 3A, AFF = 0.062[15]) with small samples. These configurations utilize the same mass spectrometer model (Nu Perspective) but differ in the use of gas-and-carbonate-based standardization versus carbonate-based standardization for the nonlinearity correction (Table 1). The comparison demonstrates that acid digestion systems, acid temperatures, and sample size negligibly affect the reproducibility of clumped data. The low values for absolute difference from "true" Δ_{47} values in these cases demonstrate that both acid digestion systems yield similarly accurate and precise data with sufficient replication (Figure 5; Table S3 and Figure S1 [supporting information]). After adjustment for multiple comparisons, we find no evidence of a difference between configurations 2 and 3A in analysis A (difference = 0.0025, SE = 0.002, t-ratio = 0.206, P = 0.9) across all measured standards (Table 3).

Similar results are obtained when we compare configuration 2 with configuration 3B (the first Nu Perspective + 90°C CAB and the second Nu Perspective + 70°C Nu Carb used in this study, respectively). Once again, we find no evidence of a statistically significant difference between configurations (difference = 0.0012, SE = 0.002, t-ratio = 0.729, P = 1) (Table 3). This suggests that both machine configurations yield similarly accurate and precise data (Table S3 [supporting information]).

5.5.2 | No evidence for difference in accuracy using the Nu Perspective versus the Thermo 253 with the three standard correction

To determine the effect of mass spectrometer model, we compare results for consistency standards between configuration 1 (Thermo 253) and configuration 2 (Nu Perspective). We also conduct a comparison of consistency standards between configurations 1 and 3B (the Thermo 253 and a second Nu Perspective used in this study, respectively). In these comparisons, we find no difference in accuracy of precision between the Nu Perspective and Thermo 253.

We observe a slight but significant difference in Δ_{47} values between configuration 1 (Thermo 253) and configuration 2 (Nu Perspective) (difference = −0.007, SE = 0.002, t-ratio = −3.889, P = 0.005) (Figure 5; Table 3). However, when standards are examined on a pairwise basis, this difference is eliminated by adjustment for multiple comparisons (difference = −0.007, SE = 0.0018, t-ratio = −3.889, P = 0.37 for all standards) (Table 3).

We observe a significant difference in Δ_{47} values between configurations 1 and 3B (the Thermo 253 and Nu Perspective) (difference = −0.006, SE = 0.002, t-ratio = −3.43, P = 0.02) at the instrument level (see standards ETH-4 and TV03 in Figure 5). The difference in results for TV03 between configurations 1 and 3B may be because this standard is a heterogeneous travertine tile with several different aliquots measured that were prepared separately, potentially with different mean Δ_{47} values. However, when standards are examined on a pairwise basis, this difference is eliminated by

adjustment for multiple comparisons (difference = −0.006, SE = 0.002, t-ratio = −3.43, P = 0.7 for all standards) (Table 3).

5.6 | Comparison of mean values across machine configurations in analysis B

5.6.1 | No evidence for differences in accuracy and precision between acid digestion systems and acid temperatures on the Nu

A comparison of configuration 2 (Nu Perspective + 90°C CAB) and configuration 3A (Nu Perspective + 70°C Nu Carb) (Table S4 [supporting information]) yields results that are similar to those from analysis A.

We find no evidence of a statistically significant difference between the configurations (difference = 0.001, SE = 0.002, t-ratio = 0.583, P = 1) (Figure 5; Table 3). The results are similar to those obtained from analysis A, with low deviation from "true" Δ_{47} values and low variability in standard deviations, suggesting that acid digestion system, acid temperature, and sample size do not affect reproducibility.

The same comparison was carried out with configuration 2 and configuration 3B (a second Nu Perspective + 70°C Nu Carb). The results are similar to those obtained from the earlier comparison and to those from analysis A. Once again, the low absolute difference from "true" Δ_{47} values demonstrates that both machine configurations yield similarly accurate and precise data (Figures 4 and 5; Table S4 [supporting information]). We find no evidence of a statistically significant difference (difference = 0.0009, SE = 0.002, t-ratio = 0.578, P = 1) (Table 3).

5.6.2 | Slight difference in accuracy using the Nu Perspective versus the Thermo 253

A comparison of configuration 1 (Thermo 253 + CAB) with configuration 2 (Nu Perspective + CAB) and configuration 3B (Nu Perspective + Nu Carb) shows there is a slight but statistically significant difference between mass spectrometer models (configuration 1 vs. 2: difference = −0.004, SE = 0.001, t-ratio = −3.556, P = 0.003; configuration 1 vs. 3B: difference: −0.003, SE = 0.001, t-ratio = −2.924, P = 0.02) (Table 3; Figure 5).

5.7 | Comparison of mean values across carbonate-based standardization versus gas-and-carbonate-based standardization

The Nu Carb systems used in configurations 3A and 3B utilize carbonates for the nonlinearity correction, whereas the CAB systems used in configurations 1 and 2 use carbonates in conjunction with equilibrated gases to construct the nonlinearity correction. A

comparison of values across digestion systems indicates that there may not be a direct benefit in incorporating equilibrated gases in the nonlinearity correction (Figure 5; Tables S3 and S4 [supporting information]).

For analysis A, we measure the effect of utilizing equilibrated gases by first comparing results from configuration 2 (gas-and-carbonate-based standardization) with those from configurations 3A and 3B (carbonate-based standardization) (Figure 5; Table S3 [supporting information]). We find no evidence that any of these differences are statistically significant (configuration 2 vs. 3A: difference = 0.003, SE = 0.002, t-ratio = 1.215, P = 0.9; configuration 2 vs. 3B: difference = −0.001, SE = 0.002, t-ratio = 0.729, P = 1) (Table 3).

We also compare the results of analysis A from configuration 1 (gas-and-carbonate-based standardization) with those from configurations 3A and 3B (carbonate-based standardization) (Figure 5; Table S3 [supporting information]). There is no improved accuracy or precision in the use of a carbonate-based versus gas-and-carbonate-based standardization, given that there are no statistically significant differences between them (configuration 1 vs. 3A: difference = −0.004, SE = 0.002, t-ratio = −2.168, P = 0.4; as earlier the difference between configuration 1 and 3B [difference = −0.006, SE = 0.002, t-ratio = −3.434, P = 0.02] is no longer significant at the

standard level after adjustment for multiple comparisons [difference = −0.006, SE = 0.002, t-ratio = −3.4344, P = 0.7 for all standards]) (Table 3).

For analysis B, we follow the same approach and first compare data from configuration 2 (gas-and-carbonate-based standardization) with those from configurations 3A and 3B (carbonate-based standardization) (Figure 5; Table S4 [supporting information]). These data show no significant differences in accuracy or precision by use of a mixed gas-and-carbonate-based standardization approach (configuration 2 vs. 3A: difference = 0.001, SE = 0.002, t-ratio = 0.583, P = 1; configuration 2 vs. 3B: difference = 0.0009, SE = 0.002, t-ratio = 0.578, P = 1) (Table 3).

We also compare the results of analysis B from configuration 1 (carbonate + equilibrated gas standardization) with those from configurations 3A and 3B (carbonate-based standardization) (Table 3; Figure 5; Table S4 [supporting information]). No standards show a significant difference across configurations (configuration 1 vs. 3A: difference = −0.0004, SE = 0.002, t-ratio = −0.258, P = 1; configuration 1 vs. 3B: difference: −0.0006, SE = 0.001, t-ratio = −0.466, P = 1).

Thus, for all consistency standards measured across these configurations in analyses A and B, the absolute differences in Δ_{47} values are minimal, consistently below 0.02‰. Considering analyses
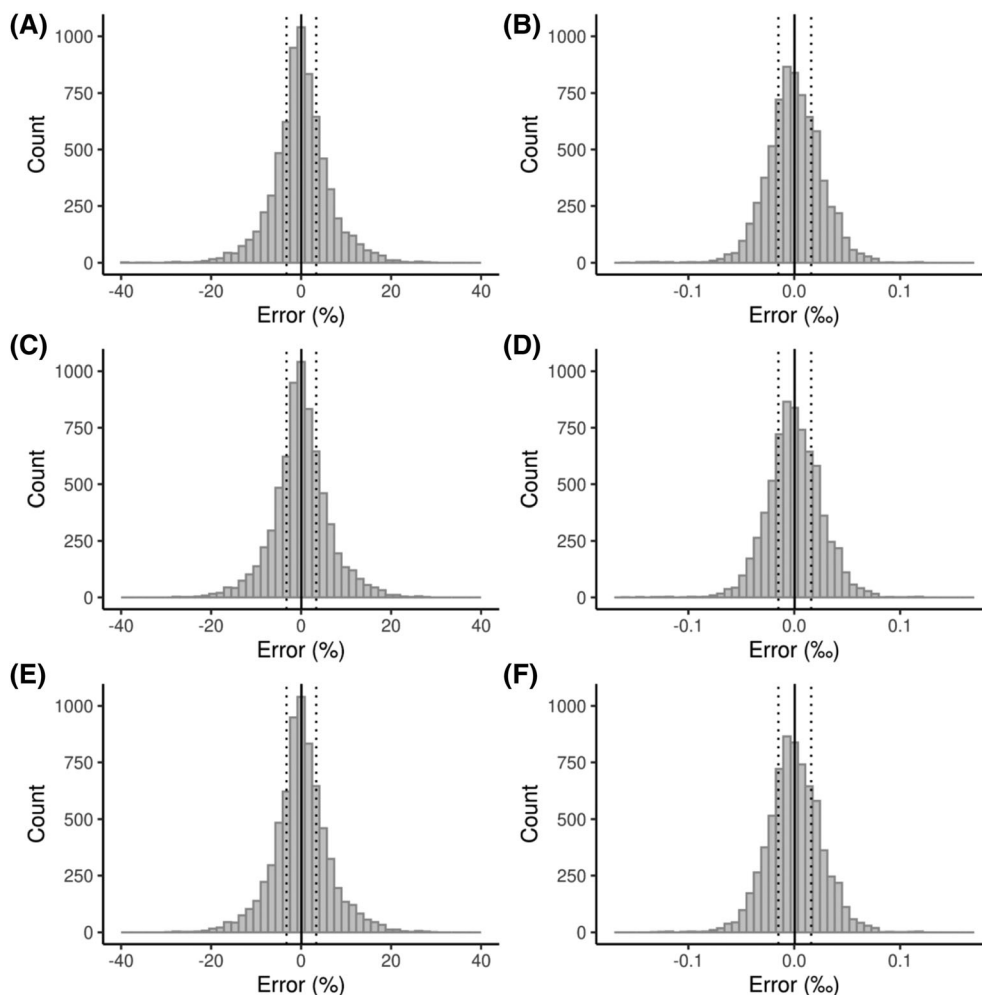

FIGURE 6 Histograms of percentage error (A, C, E) and absolute error (B, D, F) in Δ_{47} measurements. A, Before removal of outliers in percentage error. B, Before removal of outliers in permil. C and D, After removal of outliers based on percentage error in percentage error (C) and absolute error (D). E and F, After removal of outliers based on absolute error in percentage error (E) and absolute error (F) (see Section 5.3 and Tables S5 and S7 [supporting information]). Solid vertical lines indicate the mean percentage error, and dotted vertical lines indicate the first and third quarters. Median errors for all subplots are 0% and 0‰

across all standards and configurations, there is no evidence of a difference between analyses A and B (difference = −0.0004, SE = 0.001, pooled df = 6992, t-ratio = −0.350, P = 0.7). We see no improved or reduced accuracy or precision due to the use of carbonate-based standardization versus implementation of a mixed gas-and-carbonate-based standardization scheme using these methods.

5.8 | Comparison of long-term precision

Long-term precision was determined by comparing analyses from *emmeans* (i.e., standard alone, standard plus configuration, standard plus analysis, and standard plus configuration and analysis). The results for these investigations reveal that the differences in precision between acid digestion systems and mass spectrometers are less than 2% (absolute value Δ_{47} = 0.02‰). We also find there is no major improvement in accuracy through the use of carbonate standards relative to mixed gas-and-carbonate standardization for the nonlinearity correction, or if three or eight standards are used for drift corrections and the construction of transfer functions, assuming that multiple standards are run each day. The mean percentage error across all Δ_{47} measurements in this study is 0.03% (median = 0%), and it ranges from −38.7% to 29.2% and is normally distributed. The mean absolute error is 0.0002‰, and it ranges from −0.1620‰ to 0.1190‰. However, the 9 most extreme errors are true outliers if considering percentage error, and the 10 most extreme errors are true outliers if considering absolute error (Rosner's test, k = 10; see Table S7 [supporting information] for full output, Figure 6 for histograms of errors, and Table S5 [supporting information] for errors by standard, configuration, and analysis). After the outliers are removed, the errors have a mean of 0.06% (0.0003‰) and a median of 0%, with a range from −27.6% to 27.8% (−0.162‰ to 0.119‰) if considering the percentage error. If testing for outliers is based on absolute error, the final errors have a mean of 0.075% (0.0004‰) with a median of 0% and a range of −28.6% to 29.2% (−0.129‰ to 0.119‰). In both cases, the errors are normally distributed.

Figure 7 and Figure S2 (supporting information) show standard residuals as a function of analysis date for each instrumental configuration, for analyses B and A, respectively. We note that, as in Meckler et al,[43] the use of a moving average correction does help to characterize instrument drift and reduce systematic variation. However, as in their study,[43] we also occasionally observe some run-specific variability during certain correction intervals in addition to scatter that is random (Figure 7; Figure S2 [supporting information]). This points to the importance of running standards frequently during a given correction interval, with a mix of standards, and choosing moving average windows and lengthy correction intervals (when possible), to characterize systematic drift. We also note that most standards are indistinguishable from a normal distribution, for both analyses A and B (Figures 5 and 7; Figure S2 [supporting information]). The vast majority of standards fall within 3? of their average Δ47 value, consistent with long-term

reproducibility for all instrumentation and standardization techniques used and intercomparability. Figure 4 and Figure S2 (supporting information) show Δ_{47} residuals and their standard deviation for each machine configuration in both analyses. In each of these figures, with the exception of data obtained for TV03 on configuration 1 in analysis A, all configurations and standards yield very similar results for Δ_{47} residual and Δ_{47} residual standard deviation. A normality test for the residuals, using those from configuration 2 as "true" values, shows that the distribution of residuals is indistinguishable from a normal distribution. Figure S3 (supporting information) shows that all the instrument configurations
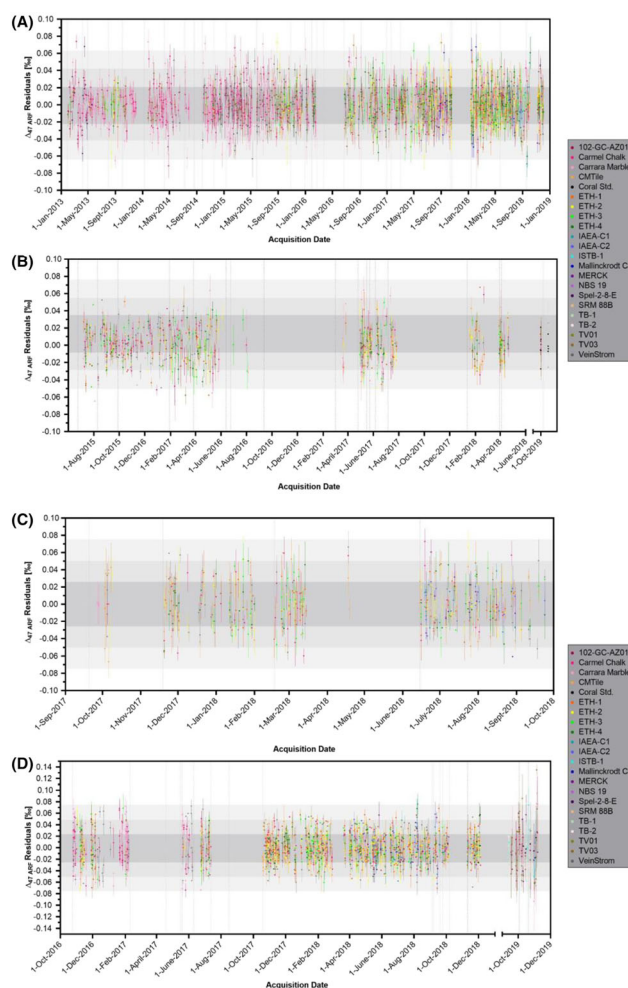

FIGURE 7 Individual standard Δ_{47} residuals for each measurement for each instrumental configuration measured in the Tripati Lab over multiple years for analysis B, after Meckler et al.[43] Each symbol represents a residual calculated from a single analysis relative to mean values from Table S4 (supporting information), and the error bar represents 1 standard error. Vertical dotted gray lines indicate the boundaries of each correction interval. Horizontal dark-gray band indicates a standard deviation of 1σ from the mean, intermediate gray indicates a standard deviation of 2σ from the mean, and light-gray band indicates a standard deviation of 3σ from the mean. A, Configuration 1, B, configuration 2, C, configuration 3A, and D, configuration 3B (note different *y*-axis range in this panel)

FIGURE 8 Comparison of Δ_{47} values (‰; I-CDES [InterCarb-Carbon Dioxide Equilibrium Scale]) for standards. Mean values for seven standards from the Tripati Lab (filled red symbols) and individual lab instrumental configurations (other colored symbols compared to mean values reported from Bernasconi et al[21] (x-axis). Also shown are mean values for individual laboratories (open symbols) from Bernasconi et al
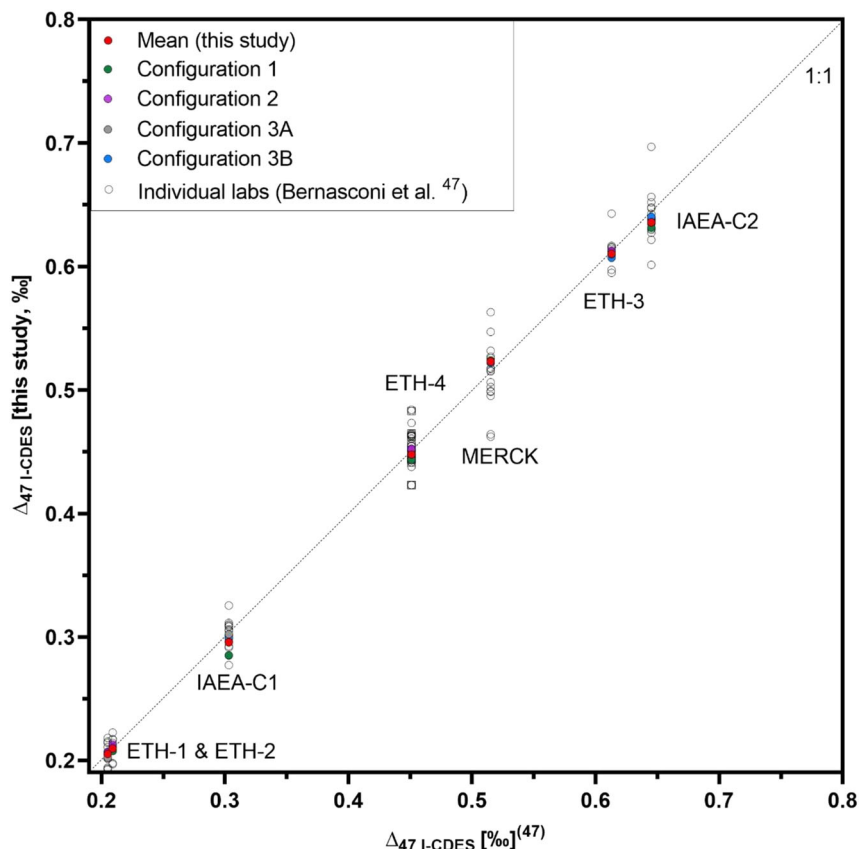
5.9 | Comparison with published standard values

Data correction procedures and values for the ETH standards used in this study were directly influenced by the multiple iterations of the InterCarb study.[21] Here, we discuss results of this study compared with the most recent InterCarb exercise.[21] Figure 8 compares values for the seven InterCarb carbonate standards measured in this study (the ETH standards, IAEA-C1, IAEA-C2, and MERCK) (Table 1; Table S8 [supporting information]), projected onto the InterCarb-CDES frame relative to published values.[21] An unpaired t-test reveals that there is no statistically significant difference between the mean values determined from this study and their updated mean values from Bernasconi et al.[21]

If individual instrument means are compared with the multi-laboratory mean value from Bernasconi et al,[21] there is strong agreement; the P-values for all seven standards are far greater than the 0.05 threshold for significance, with the exception of ETH-3 and ETH-4 on configurations 1 and 3B ($P = 0.015$) and IAEA-1 on configuration 1 ($P = 0.0002$). We note that the reported interlaboratory variance for isotope ratios determined for the four standards that were calculated using carbonate-based standardization (ETH-4, IAEA-1, IAEA-2, and MERCK) in Bernasconi et al[21] ranges from 0.05‰ to 0.10‰, and for the four standards that were

calculated using gas standardization (ETH-1 to ETH-4), the variance ranges from 0.05‰ to 0.09‰. A comparison of our determined values for ISTB-1 with published values from Chang et al[42] indicates an offset of 0.04‰, which is within the range of variance between laboratories reported by Bernasconi et al.[21]

Within our laboratory, among the four instrument configurations examined for this study, the determined Δ_{47} values for each of the seven standards for analysis A (Table S3 [supporting information]) agree to within 0.02‰ and for analysis B agree to within 0.01‰ (Table S4 [supporting information]), with one exception that is to within 0.02‰. Values calculated using emmeans agree to within 0.009‰ (Table S8 [supporting information]). Thus, the methods utilized in this study, despite representing four configurations using two different generations of instrumentation and two acid digestion systems, yield accurate values, as determined through comparison with InterCarb results from 25 machine configurations with three mass spectrometer models and multiple acid digestion systems. In addition, the results from multiple configurations in our laboratory are highly reproducible (Figure 8).

6 | RECOMMENDATIONS

These results show that it is possible to obtain comparable Δ_{47} data for different mass spectrometers, using methods described in this study and references therein. We recommend a high frequency of standard analyses, with three to five standards run in

approach shot noise limits after 3–20 replicates, depending on the beam intensity and integration time.

a day for the instrumental configurations described here, as this allows for the adequate characterization of the behavior of each instrument that can occur on short time frames. Typically, for all instrumental configurations, during most standardization windows, a 20-point moving average was sufficient to characterize instrument drift (with 10 points on either side of a given working standard).

Related to this, a significant step in obtaining reproducible results on one instrument through time, and comparable results between different types of instrumentation, is the determination of standardization windows (i.e., correction intervals). Here these are defined based on either major instrument maintenance-related events (e.g., source retuning) or phenomenological changes in raw data. A previous study from our laboratory[18] showed that the use of 1 month standardization windows on these instrumental configurations yields comparable results between instruments.

We recommend using at least two standards (gas and/or carbonates) with a stochastic distribution to construct the nonlinearity correction, as previously suggested.[1,16,17] Our analyses show that the use of carbonates only or of a mixed gas-and-carbonate standardization does not make a statistically significant difference in results. This implies that it is feasible to eliminate the use of equilibrated gases. However, we note that for diagnostic purposes, we continue to run equilibrated gases for our CAB systems, as this allows us to assess whether systematic drift over the course of a correction interval results from the carbonate digestion and purification system or the mass spectrometer.

For transfer functions, the use of three standards that span a range of δ_{47} and Δ_{47} space, such as ETH-1, ETH-2, and ETH-3, yields similar results to using eight standards, if multiple standards are measured each day.[18,21,44] We also recommend using multiple ancillary standards as a consistency check. If a sufficient number of standards is used, similarly accurate and precise data should be achievable on different types of instrumentation irrespective of (a) the size of baselines, (b) the slopes of the nonlinearity corrections, (c) the number of working standards used (as long as n > 3 and appropriately constrains the nonlinearity correction and Empirical Transfer Function), or (d) whether a carbonate-based or gas-and-carbonate-based standardization procedure is used. We found that with a range of instruments and digestion systems standards yield similar mean values and that a long-term reproducibility of 0.02‰, or a precision of 98.2%, was achievable. These results validate the use of the InterCarb recommendations, specifically that accurate and precise results may be achieved using carbonate-based standardization over multiyear timescales in a laboratory.

7 | CONCLUSIONS

This study shares our understanding of practices that contribute to the intercomparability of Δ_{47} measurements from different acid digestion systems and mass spectrometer models based on data for four instrumental configurations in the Tripati Lab. We report the long-term accuracy and reproducibility of consistency standards measured on each configuration, and using either a combination of equilibrated gas standards and heated carbonate standards to determine the nonlinearity correction or carbonate standards on their own. We test the recent InterCarb proposal for carbonate-based standardization. We examine the use of three or eight carbonate standards for drift corrections and the construction of transfer functions, using a 20-point moving average, and report the isotopic compositions of 21 different standards.

We show that when instrument behavior is well characterized, negligible differences in reproducibility stem from the acid digestion system used, the mass spectrometer model used, the use of three or eight carbonate standards (with at least two having a stochastic distribution that vary in δ_{47}) or the use of stochastic carbonate standards or equilibrated gases in the nonlinearity correction, and two carbonate standards (or gases) that vary in Δ_{47} with a compositional range similar to our study to constrain stretching/scale compression to construct an empirical transfer function. The CAB system and small-sample Nu Carb system yield similar results. As indicated by the P-values in Table 3, all four configurations yield similar values. Table S9 (supporting information) presents standard values compared with each other for each configuration and each analysis, with 503 comparisons shown. All machine configurations yield robust, comparable results for both analyses A and B, with P-values consistently well over the threshold for significance (no P-values <0.05; 18 comparisons with $0.3 < P < 0.5$ and the remainder >0.5). Any apparent variability may result from random noise, low replication at the time of the analysis, and/or different aliquots of a heterogeneous standard. These results indicate that, depending on instrument configuration, counting times, sample size, and sample heterogeneity, it should be possible to obtain close to shot noise on all instrument configurations.

We report values for 21 different standards on the absolute reference frame using the IUPAC parameter set,[32] including standards from UCLA, ETH Zürich, Caltech, and CUG; 4628 standard measurements between 2013 and 2019 are provided for four instrument configurations, including three mass spectrometers and two acid digestion systems. Tables S3 and S4 (supporting information) show standard numbers for analysis A that used three carbonate standards and analysis B that used eight carbonate standards for corrections; for these analyses, 1350 measurements of InterCarb standards were used for standardization across all instrument configurations for analysis A and 1900 measurements across all instrument configurations for analysis B. Thus, these values provide robust constraints on long-term reproducibility for these types of configurations. Furthermore, the demonstration of intercomparability between different instrumental configurations in concert with these standard values strengthens the assertion that intercomparable data sets can be produced between a diversity of laboratories using different generations of instrumentation. The standard values reported here also facilitate a restandardization of published data going back to the first carbonate clumped isotope

measurements in 2006, on the absolute reference frame, so these data can be compared with recently generated data and be intercomparable.

DATA AVAILABILITY STATEMENT

All data that support the findings of this study are openly available at https://github.com/Tripati-Lab/Upadhyay-et-al.

ORCID

Deepshikha Upadhyay https://orcid.org/0000-0002-3489-4618
Jamie Lucarelli https://orcid.org/0000-0002-9104-2518

REFERENCES

1. Ghosh P, Adkins J, Affek H, et al. ^{13}C–^{18}O bonds in carbonate minerals: A new kind of paleothermometer. *Geochim Cosmochim Acta*. 2006;70(6):1439-1456. https://doi.org/10.1016/j.gca.2005.11.014

2. Eiler JM. "Clumped-isotope" geochemistry—The study of naturally-occurring, multiply-substituted isotopologues. *Earth Planet Sci Lett*. 2007;262(3-4):309-327. https://doi.org/10.1016/j.epsl.2007.08.020

3. Tripati AK, Hill PS, Eagle RA, et al. Beyond temperature: Clumped isotope signatures in dissolved inorganic carbon species and the influence of solution chemistry on carbonate mineral composition. *Geochim Cosmochim Acta*. 2015;166:344-371. https://doi.org/10.1016/j.gca.2015.06.021

4. Douglas PMJ, Stolper DA, Eiler JM, et al. Methane clumped isotopes: Progress and potential for a new isotopic tracer. *Org Geochem*. 2017;113:262-282. https://doi.org/10.1016/j.orggeochem.2017.07.016

5. Wang DT, Gruen DS, Lollar BS, et al. Nonequilibrium clumped isotope signals in microbial methane. *Science*. 2015;348(6233):428-431. https://doi.org/10.1126/science.aaa4326

6. Schauble EA, Ghosh P, Eiler JM. Preferential formation of 13C–18O bonds in carbonate minerals, estimated using first-principles lattice dynamics. *Geochim Cosmochim Acta*. 2006;70(10):2510-2529. https://doi.org/10.1016/j.gca.2006.02.011

7. Hill PS, Tripati AK, Schauble EA. Theoretical constraints on the effects of pH, salinity, and temperature on clumped isotope signatures of dissolved inorganic carbon species and precipitating carbonate minerals. *Geochim Cosmochim Acta*. 2014;125:610-652. https://doi.org/10.1016/j.gca.2013.06.018

8. Tripati AK, Sahany S, Pittman D, et al. Modern and glacial tropical snowlines controlled by sea surface temperature and atmospheric mixing. *Nat Geosci*. 2014;7(3):205-209. https://doi.org/10.1038/ngeo2082

9. Suarez MB, Passey BH. Assessment of the clumped isotope composition of fossil bone carbonate as a recorder of subsurface temperatures. *Geochim Cosmochim Acta*. 2014;140:142-159. https://doi.org/10.1016/j.gca.2014.05.026

10. Eagle RA, Tütken T, Martin TS, et al. Dinosaur body temperatures determined from isotopic (13C-18O) ordering in fossil biominerals. *Science*. 2011;333(6041):443-445. https://doi.org/10.1126/science.1206196

11. Kimball J, Eagle R, Dunbar R. Carbonate "clumped" isotope signatures in aragonitic scleractinian and calcitic gorgonian deep-sea corals. *Biogeosciences*. 2016;13(23):6487-6505. https://doi.org/10.5194/bg-13-6487-2016

12. Guo W, Eiler JM. Temperatures of aqueous alteration and evidence for methane generation on the parent bodies of the CM chondrites. *Geochim Cosmochim Acta*. 2007;71(22):5565-5575. https://doi.org/10.1016/j.gca.2007.07.029

13. Guo W. Carbonate Clumped Isotope Thermometry: Applications to Carbonaceous Chondrites and Effects of Kinetic Isotope Fractionation. Dissertation. California Institute of Technology; 2009.

14. Halevy I, Fischer WW, Eiler JM. Carbonates in the Martian meteorite Allan Hills 84001 formed at 18 ±4 °C in a near-surface aqueous environment. *Proc Natl Acad Sci*. 2011;108(41):16895-16899. https://doi.org/10.1073/pnas.1109444108

15. Saenger C, Affek HP, Felis T, Thiagarajan N, Lough JM, Holcomb M. Carbonate clumped isotope variability in shallow water corals: Temperature dependence and growth-related vital effects. *Geochim Cosmochim Acta*. 2012;99:224-242. https://doi.org/10.1016/j.gca.2012.09.035

16. Dennis KJ, Affek HP, Passey BH, Schrag DP, Eiler JM. Defining an absolute reference frame for 'clumped' isotope studies of CO_2. *Geochim Cosmochim Acta*. 2011;75(22):7117-7131. https://doi.org/10.1016/j.gca.2011.09.025

17. Bernasconi SM, Müller IA, Bergmann KD, et al. Reducing uncertainties in carbonate clumped isotope analysis through consistent carbonate-based standardization. *Geochem Geophys Geosyst*. 2018;19(9):2895-2914. https://doi.org/10.1029/2017gc007385

18. Defliese WF, Tripati A. Analytical effects on clumped isotope thermometry: Comparison of a common sample set analyzed using multiple instruments, types of standards, and standardization windows. *Rapid Commun Mass Spectrom*. 2020;34(8):e8666. https://doi.org/10.1002/rcm.8666

19. Petersen SV, Defliese WF, Saenger C, et al. Effects of improved ^{17}O correction on Interlaboratory agreement in clumped isotope calibrations, estimates of mineral-specific offsets, and temperature dependence of acid digestion fractionation. *Geochem Geophys Geosyst*. 2019;20(7):3495-3519. https://doi.org/10.1029/2018gc008127

20. Fernandez A, Müller IA, Rodríguez-Sanz L, van Dijk J, Looser N, Bernasconi SM. A reassessment of the precision of carbonate clumped isotope measurements: Implications for calibrations and paleoclimate reconstructions. *Gechem Geophys Geosyst*. 2017;18(12):4375-4386. https://doi.org/10.1002/2017GC007106

21. Bernasconi SM, Daëron M, Bergmann KD, et al. A community effort to improve inter-laboratory standardization of the carbonate clumped isotope thermometer using carbonate anchors. *Geochem Geophys Geosyst*. 2021;22(5). https://doi.org/10.1002/essoar.10504430.4

22. Passey BH, Levin NE, Cerling TE, Brown FH, Eiler JM. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proc Natl Acad Sci*. 2010;107(25):11245-11249. https://doi.org/10.1073/pnas.1001824107

23. Eiler JM, Schauble E. $^{18}O^{13}C^{16}O$ in Earth's atmosphere. *Geochim Cosmochim Acta*. 2004;68(23):4767-4777. https://doi.org/10.1016/j.gca.2004.05.035

24. Guo W, Daëron M, Niles P, et al. ^{13}C-^{18}O bonds in dissolved inorganic carbon: Implications for carbonate clumped isotope thermometry. *Geochim Cosmochim Acta*. 2008;72(12):A336-A338. https://doi.org/10.1016/j.gca.2008.05.010

25. Tripati AK, Eagle RA, Thiagarajan N, et al. ^{13}C–^{18}O isotope signatures and 'clumped isotope' thermometry in foraminifera and coccoliths.

Geochim Cosmochim Acta. 2010;74(20):5697-5717. https://doi.org/10.1016/j.gca.2010.07.006

26. Huntington KW, Eiler JM, Affek HP, et al. Methods and limitations of 'clumped' CO_2 isotope (Δ_{47}) analysis by gas-source isotope ratio mass spectrometry. *J Mass Spectrom*. 2009;44(9):1318-1329. https://doi.org/10.1002/jms.1614

27. Affek HP, Eiler JM. Abundance of mass 47 CO_2 in urban air, car exhaust, and human breath. *Geochim Cosmochim Acta*. 2006;70(1):1-12. https://doi.org/10.1016/j.gca.2005.08.021

28. Müller IA, Fernandez A, Radke J, et al. Carbonate clumped isotope analyses with the long-integration dual-inlet (LIDI) workflow: Scratching at the lower sample weight boundaries. *Rapid Commun Mass Spectrom*. 2017;31(12):1057-1066. https://doi.org/10.1002/rcm.7878

29. Daëron M, Blamart D, Peral M, Affek HP. Absolute isotopic abundance ratios and the accuracy of Δ_{47} measurements. *Chem Geol*. 2016;442:83-96. https://doi.org/10.1016/j.chemgeo.2016.08.014

30. John CM, Bowen D. Community software for challenging isotope analysis: First applications of 'Easotope' to clumped isotopes. *Rapid Commun Mass Spectrom*. 2016;30(21):2285-2300. https://doi.org/10.1002/rcm.7720

31. Schauer AJ, Kelson J, Saenger C, Huntington KW. Choice of ^{17}O correction affects clumped isotope (Δ_{47}) values of CO_2 measured with mass spectrometry. *Rapid Commun Mass Spectrom*. 2016;30(24):2607-2616. https://doi.org/10.1002/rcm.7743

32. Brand WA, Assonov SS, Coplen TB. Correction for the ^{17}O interference in $\delta(^{13}C)$ measurements when analyzing CO_2 with stable isotope mass spectrometry (IUPAC technical report). *Pure Appl Chem*. 2010;82(8):1719-1733. https://doi.org/10.1351/pac-rep-09-01-05

33. Bernasconi SM, Hu B, Wacker U, Fiebig J, Breitenbach SFM, Rutz T. Background effects on faraday collectors in gas-source mass spectrometry and implications for clumped isotope measurements. *Rapid Commun Mass Spectrom*. 2013;27(5):603-612. https://doi.org/10.1002/rcm.6490

34. Swart PK, Burns SJ, Leder JJ. Fractionation of the stable isotopes of oxygen and carbon in carbon dioxide during the reaction of calcite with phosphoric acid as a function of temperature and technique. *Chem Geol: Isot Geosci*. 1991;86(2):89-96. https://doi.org/10.1016/0168-9622(91)90055-2

35. He B, Olack GA, Colman AS. Pressure baseline correction and high-precision CO_2 clumped-isotope ($\Delta47$) measurements in bellows and micro-volume modes. *Rapid Commun Mass Spectrom*. 2012;26(24):2837-2853. https://doi.org/10.1002/rcm.6436

36. Wacker U, Fiebig J, Schoene BR. Clumped isotope analysis of carbonates: Comparison of two different acid digestion techniques. *Rapid Commun Mass Spectrom*. 2013;27(14):1631-1642. https://doi.org/10.1002/rcm.6609

37. Defliese WF, Hren MT, Lohmann KC. Compositional and temperature effects of phosphoric acid fractionation on $\Delta47$ analysis and implications for discrepant calibrations. *Chem Geol*. 2015;396:51-60. https://doi.org/10.1016/j.chemgeo.2014.12.018

38. Murray ST. *The Application of Clumped Isotopes to Dolomites*. Ann Arbor, MI: ProQuest Dissertations Publishing; 2016.

39. Müller IA, Violay MES, Storck J-C, et al. Clumped isotope fractionation during phosphoric acid digestion of carbonates at 70°C.

Chem Geol. 2017;449:1-14. https://doi.org/10.1016/j.chemgeo.2016.11.030

40. Henkes GA, Passey BH, Wanamaker AD Jr, Grossman EL, Ambrose WG, Carroll ML. Carbonate clumped isotope compositions of modern marine mollusk and brachiopod shells. *Geochim Cosmochim Acta*. 2013;106:307-325. https://doi.org/10.1016/j.gca.2012.12.020

41. Wang Z, Schauble EA, Eiler JM. Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. *Geochim Cosmochim Acta*. 2004;68(23):4779-4797. https://doi.org/10.1016/j.gca.2004.05.039

42. Chang B, Defliese WF, Li C, Huang J, Tripati A, Algeo TJ. Effects of different constants and standards on the reproducibility of carbonate clumped isotope (Δ_{47}) measurements: Insights from a long-term dataset. *Rapid Commun Mass Spectrom*. 2020;34(8):e8678. https://doi.org/10.1002/rcm.8678

43. Meckler AN, Ziegler M, Millán MI, Breitenbach SFM, Bernasconi SM. Long-term performance of the Kiel carbonate device with a new correction scheme for clumped isotope measurements. *Rapid Commun Mass Spectrom*. 2014;28(15):1705-1715. https://doi.org/10.1002/rcm.6949

44. Kocken IJ, Müller IA, Ziegler M. Optimizing the use of carbonate standards to minimize uncertainties in clumped isotope data. *Geochem Geophys Geosyst*. 2019;20(11):5565-5577. https://doi.org/10.1029/2019GC008545

45. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. https://www.r-project.org/. Accessed October 13, 2020.

46. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and nonlinear mixed effects models. R package version 3.1–152. 2021. https://cran.r-project.org/package=nlme. Accessed May 27, 2021.

47. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.0. https://CRAN.R-project.org/package=emmeans Accessed May 27, 2021.

48. Kuhn M. caret: Classification and Regression Training. R package version 6.0–88. 2021. https://cran.r-project.org/package=caret. Accessed May 27, 2021.

49. Millard SP. *EnvStats: An R Package for Environmental Statistics*. 2nded. New York, NY: Springer-Verlag; 2013.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.