## Supplementary Materials included at end of pdf.

Earth and Planetary Science Letters 382 (2013) 47-57

Contents lists available at ScienceDirect

# Earth and Planetary Science Letters

www.elsevier.com/locate/epsl

# A revised calibration of the clumped isotope thermometer

## Shikma Zaarur\*, Hagit P. Affek, Mark T. Brandon

Department of Geology and Geophysics, Yale University, New Haven, CT, United States

#### ARTICLE INFO

Article history: Received 5 December 2012 Received in revised form 3 July 2013 Accepted 16 July 2013 Available online xxxx Editor: G. Henderson

Keywords: clumped isotopes carbonates thermometer calibration

#### ABSTRACT

A growing number of materials and environmental settings are studied using the carbonate clumped isotope ( $\Delta_{47}$ ) thermometer. The method has been applied in both biogenic and non-biogenic carbonate systems, in marine and terrestrial settings, over a wide range of geological timescales. The current  $\Delta_{47}$  temperature calibration gives good temperature estimates for most biogenic materials, however, systematic biases are commonly observed at low temperatures.

In this study we report additional calibration data, that covers a wider temperature range, at more evenly distributed temperatures, and are measured at higher analytical precision than the original calibration. Combining these data with the original calibration yields a  $\Delta_{47}$ -T relationship that is similar to the original calibration, though slightly less steep:  $\Delta_{47} = (0.0526 \pm 0.0025) \times 10^6/T^2 + (0.0520 \pm 0.0284)$ . This revised calibration is in better agreement with biogenic carbonates, including those grown at low temperatures. The difference between the original and revised calibrations is significant for carbonates forming below 16 °C or above 49 °C ( $\Delta_{47}$  values of 0.68‰ and 0.56‰). Additionally, we include a comprehensive analysis of the sources of error associated with  $\Delta_{47}$  measurements and estimated temperatures and recommend measurement strategies for obtaining the desired precision.

As an illustration, we apply the revised calibration and uncertainty analysis to 3 previously published studies. At low temperatures, the revised calibration results in significant differences from the original calibration and hence affects the interpretation of the environmental signal recorded. In light of our  $\Delta_{47}$  errors analysis, in cases where the temperature signals are small, we find that replicate analyses are critical to identify a temperature signal.

© 2013 Elsevier B.V. All rights reserved.

#### 1. Introduction

Carbonate clumped isotope thermometry is a new proxy for estimating paleotemperatures. This technique is based on the natural abundance of <sup>13</sup>C-<sup>18</sup>O bonds in the carbonate lattice, relative to that expected for a random distribution of isotopes among all isotopologues, quantified by the parameter  $\Delta_{47}$  (Affek, 2012; Eiler, 2007; Wang et al., 2004). As such, the thermometer is based on the thermodynamically controlled preference of two heavy isotopes to bind with each other, and it is independent of the absolute abundance of <sup>13</sup>C and <sup>18</sup>O in the carbonate mineral. Carbonate clumped isotope thermometry is therefore a powerful approach for determining the growth temperature of CaCO<sub>3</sub> minerals.

Ghosh et al. (2006a) were the first to present a calibration of the  $\Delta_{47}$ -*T* CaCO<sub>3</sub> relationship. Their calibration is based on  $\Delta_{47}$  measurements of 7 calcite samples that were formed by slow laboratory precipitation at controlled temperatures between 1 °C and 50 °C. This experimental approach followed the method used by Kim and O'Neil (1997) for defining the  $\delta^{18}$ O-*T* relationship in calcite. Biogenic carbonate materials, particularly of marine organisms, grown at known temperatures, generally agree with the Ghosh et al. (2006a) calibration (as reviewed by Eiler, 2007, 2011 and Tripati et al., 2010). Other studies, however, have shown disagreement with that calibration in both synthetic materials (Dennis and Schrag, 2010) and biogenic materials (Eagle et al., 2013; Henkes et al., 2013; Saenger et al., 2012). The source of these discrepancies is still unresolved.

The Ghosh et al. (2006a) calibration is based on analyses that were done in the early days of clumped isotope measurements. Since its publication, the importance of long acquisition times and replicate analysis had been recognized as essential for the precision required for the measurement of the low abundance  ${}^{13}C{}^{18}O{}^{16}O$  isotopologue (46 ppm of all CO<sub>2</sub> molecules; Eiler and Schauble, 2004).

In this study, we examine the sources of errors in clumped isotope measurements to improve the estimate of uncertainties for the derived temperatures. We then re-examine the clumped isotope thermometer calibration by independently repeating the carbonate precipitation experiments using the same method as Ghosh et al. (2006a). Our samples span a larger temperature range and are measured in triplicates; thus they have higher analytical







<sup>\*</sup> Corresponding author. Tel.: +1 203 4323761.

*E-mail addresses:* shikma.zaarur@yale.edu (S. Zaarur), hagit.affek@yale.edu (H.P. Affek), mark.brandon@yale.edu (M.T. Brandon).

<sup>0012-821</sup>X/\$ – see front matter  $\,\,\odot$  2013 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.epsl.2013.07.026

precision. We further examine our findings to suggest an optimized protocol for lower sample uncertainties.

Our  $\Delta_{47}$ -*T* relationship is similar to the original calibration but has a slightly lower slope and a better agreement with low temperature biogenic carbonates. The agreement of the biogenic carbonates with the revised calibration line strengthens and validates the applicability of the  $\Delta_{47}$  thermometer. It further implies that this method of precipitation, even if not reflecting true equilibrium, is relevant for most biogenic carbonates that form between 0 °C and 40 °C.

#### 2. Materials and methods

#### 2.1. CaCO<sub>3</sub> precipitation experiments

We follow the precipitation method and setup described by Kim and O'Neil (1997) for  $\delta^{18}$ O and that was used by Ghosh et al. (2006a) in the first carbonate clumped isotopes thermometer calibration. Saturated Ca(HCO<sub>3</sub>)<sub>2</sub> solutions were prepared by bubbling 100% CO<sub>2</sub> for  $\sim$ 1 h through 1 L of deionized water. Reagent grade CaCO<sub>3</sub> (Mallinckrodt Chemical Works) was added to the solution while CO<sub>2</sub> bubbling continued for another hour, to increase carbonate solubility. The solution was then filtered (Whatman #40, 8 µm, filter paper) to remove non-dissolved calcium carbonate particles. Slow bubbling of humidified  $N_{2(\mathrm{g})}$  (roughly 20 bubbles per 30 seconds) deep in the solution was used to remove  $CO_2$  and induce calcium carbonate precipitation. The humidity of the N<sub>2(g)</sub> flow was adjusted to saturation at the experiment temperature in order to minimize water evaporation and enable reliable  $\delta^{18}O$ measurements in addition to  $\Delta_{47}$ . This precipitation method resulted in CaCO<sub>3</sub> formation deep within the solution. Samples were precipitated at 5, 8, 15, 25, 35, 50 and 65 °C, in a temperature controlled reactor (New Brunswick Scientific, Excella E24 incubator shaker series), which has an observed precision of  $\sim \pm 0.5$  °C. The precipitation continued between 4 days and a few weeks, depending on the precipitation temperature, and was stopped when sufficient CaCO<sub>3</sub> accumulated for analysis. The precipitated CaCO<sub>3</sub> was collected using a rubber policeman, filtered through a glass microfiber filter (Whatman 934-AH, 0.3 µm), and dried under vacuum at room temperature. Mineralogy was determined by x-ray diffraction (in the XRD laboratory at Yale University).

#### 2.2. Isotopic analysis

CaCO<sub>3</sub> (3–4 mg) was digested overnight in 105% H<sub>3</sub>PO<sub>4</sub> ( $\rho$  = 1.95 gr/cm<sup>3</sup>) at 25 °C. CO<sub>2</sub> was extracted cryogenically on a vacuum line and cleaned by passing it through a GC column (Supelco Q-Plot, 30 m × 0.53 mm) at –20 °C (following Affek and Eiler, 2006; Huntington et al., 2009; and Zaarur et al., 2011). Measurements were performed using a Thermo MAT253 gas source isotope ratio mass spectrometer (in the Earth Systems Center for Stable Isotopic Studies at Yale University), modified to simultaneously measure masses 44–49 in a dual inlet mode. Each measurement consisted of 90 cycles of sample-standard comparison, with a signal integration time of 20 s for each measurement. Samples were measured in triplicates, repeating the whole extraction procedure on separate aliquots of the powdered CaCO<sub>3</sub>.

 $\Delta_{47}$  is defined as the excess of the mass 47 signal in CO<sub>2</sub> over what is expected based on random distribution of <sup>13</sup>C and <sup>18</sup>O among all CO<sub>2</sub> isotopologues (Eiler, 2007; Wang et al., 2004). Standardization is hence performed using a set of CO<sub>2</sub> gases that are heated at 1000 °C for ~2 h to obtain random distribution (Huntington et al., 2009; Wang et al., 2004). To compare our data to previous studies, we report our values standardized to the original reference frame (used by Ghosh et al., 2006a). Our system is normalized and traceable to that original reference frame by the

measurements of several carbonates with  $\Delta_{47}$  values that were pre-determined in that original system; this normalization was later verified by inter-laboratory comparison as part of the development of the absolute reference frame (Dennis et al., 2011). To allow future inter-laboratory use of the calibration, we provide the data also in the absolute reference frame (Dennis et al., 2011, Section 4.2.5). Unless otherwise noted, values are reported in the original reference frame (Ghosh et al., 2006a). We refer the interested reader to Supplement SI1 for more details about the measurements and associated reference frames.

Both  $\delta^{18}$ O and  $\delta^{13}$ C values, measured together with the  $\Delta_{47}$ , are reported using the VPDB reference frame as defined by a pre-calibrated Oztech CO<sub>2</sub> tank used as a reference working gas with values of -15.80% and -3.64% for  $\delta^{18}O$  and  $\delta^{13}C$ , respectively. These values are verified using NBS-19 with measured  $\delta^{18}$ O and  $\delta^{13}$ C values of  $-2.17 \pm 0.04\%$  and  $+2.11 \pm 0.13\%$  (1 SD, n = 12), respectively (Kluge and Affek, 2012), which are comparable to the IAEA nominal values of -2.2% and +1.95%. Oxygen isotopic fractionation factors associated with the acid digestion reaction at 25 °C are  $^{18}\alpha_{acid}=1.01030$  and 1.01063 for calcite and aragonite, respectively (Kim et al., 2007a, 2007b). The temperature dependence of calcite-water oxygen isotopic fractionation  $(^{18}\alpha_{carbonate-water})$  is compared to the temperature dependence relationships derived by Friedman and O'Neil (1977) and by Kim and O'Neil (1997), modified to account for the above acid digestion fractionation (Eqs. (1), (2), respectively). The aragonite-water oxygen isotope fractionation is compared to Kim et al. (2007b) (Eq. (3)):

$1000 \ln \alpha_{(\text{calcite-water})} = 2.78 (10^6 T^{-2}) - 2.89$	(1)
--	-----

$1000 \ln \alpha_{\text{(calcite-water)}} =$	$= 18.03(10^{5}T^{-1})$	) — 32.17	(2)	
--	-------------------------	-----------	-----	--

 $1000 \ln \alpha_{(\text{aragonite-water})} = 17.88 (10^3 T^{-1}) - 31.14$ (3)

#### 3. Results

#### 3.1. Synthetic carbonate precipitation

 $\Delta_{47}$  values range between 0.741‰ and 0.513‰ for precipitation temperatures of 5°C to 65°C, respectively (Table 1). Depending on precipitation temperatures, first particles were observed within 1 to 5 days for high and low temperatures, respectively. As was observed in other precipitation experiments (e.g. Wray and Daniels, 1957; Zhou and Zheng, 2001), our precipitated carbonates are mixed polymorphs. Most samples are predominantly calcite with traces of aragonite and vaterite, but one sample (carb #37, precipitated at T = 50°C) is a mixture of calcite and aragonite and another (carb #43, precipitated at T = 65°C) is primarily aragonite. Note that although Ghosh et al. (2006a) reports the original calibration samples to be calcite, renewed inspection of the XRD spectra of these samples reveals traces of aragonite as well.

A positive correlation is observed between  $\Delta_{47}$  and  $\delta^{18}$ O values. This correlation is likely due to the temperature dependence of both isotopic systems and not to a kinetic affect. It strengthens the assumption of precipitation close to isotopic equilibrium in both systems. The fractionation derived from carbonate and water  $\delta^{18}$ O values approximately agrees with the temperature dependence defined by Kim and O'Neil (1997) and Friedman and O'Neil (1977) for calcite and by Kim et al. (2007b) for aragonite (Fig. 1).

#### 3.2. Long and short term standard measurements

 $\Delta_{47}$  measurements are influenced by two independent sources of error, Poisson-distributed shot noise and additional non-Poisson errors (see Eq. (21) in SI2). Internal precision is the combination

Sample	Mineralogy	$\Delta_{47}$	$\Delta_{47}$ Abs. Ref.	Growth temp.	$\delta^{13}$ C (VPDB)	$\delta^{18}O(\text{carb})$ (VPDB)	$\delta^{18}O(H_2O)$ (VSMOW)	$^{18}\alpha^{c}$	$\delta^{18}O(H_2O)^d$ (VSMOW)
		(‰)	Fram.	(°C)	(‰)	(‰)	(‰) (measured)		(‰) (calculated)
Carb 42 <sup>a</sup>	Calcite	$0.736 \pm 0.019$	0.801	$5\pm0.5$	$-32.19 \pm 0.02$	$-4.00\pm0.02$	-7.15	1.0342	$-6.21 \pm 0.14$
Carb 42 <sup>a</sup>	Calcite	$0.741 \pm 0.019$	0.806	$5\pm0.5$	$-32.16 \pm 0.02$	$-3.93 \pm 0.02$	-7.15	1.0342	$-6.14\pm0.14$
Carb 42 <sup>a</sup>	Calcite	$0.745 \pm 0.019$	0.811	$5\pm0.5$	$-32.17 \pm 0.02$	$-3.82 \pm 0.02$	-7.15	1.0344	$-6.02 \pm 0.14$
Carb 19 <sup>a</sup>	Calcite+vaterite	$0.735 \pm 0.019$	0.800	$8\pm2$	$-30.84 \pm 0.02$	$-3.49\pm0.02$	-5.64	1.0331	$-5.01 \pm 0.47$
Carb 19 <sup>a</sup>	Calcite+vaterite	$0.707 \pm 0.019$	0.771	$8\pm2$	$-31.05 \pm 0.02$	$-3.39 \pm 0.02$	-5.64	1.0332	$-4.92 \pm 0.47$
Carb 19 <sup>a</sup>	Calcite+vaterite	$0.677 \pm 0.019$	0.739	$8\pm2$	$-30.91 \pm 0.02$	$-3.44\pm0.02$	-5.64	1.0332	$-4.96 \pm 0.47$
Carb 36 <sup>a</sup>	Calcite	$0.650 \pm 0.019$	0.710	$15 \pm 0.5$	$-30.35 \pm 0.02$	$-6.45\pm0.02$	-7.59	1.0321	$-6.42 \pm 0.13$
Carb 36 <sup>a</sup>	Calcite	$0.660 \pm 0.019$	0.721	$15 \pm 0.5$	$-30.71 \pm 0.02$	$-6.42 \pm 0.02$	-7.59	1.0321	$-6.39 \pm 0.13$
Carb 36 <sup>a</sup>	Calcite	$0.719 \pm 0.019$	0.783	$15 \pm 0.5$	$-29.72 \pm 0.02$	$-7.09 \pm 0.02$	-7.59	1.0314	$-7.06 \pm 0.13$
Carb 23 <sup>a</sup>	Calcite	$0.640 \pm 0.019$	0.700	$25\pm2$	$-33.14\pm0.02$	$-7.67 \pm 0.02$	-5.70	1.0289	$-5.55 \pm 0.42$
Carb 23 <sup>a</sup>	Calcite	$0.652 \pm 0.019$	0.713	$25\pm2$	$-33.03\pm0.02$	$-7.73 \pm 0.02$	-5.70	1.0288	$-5.62 \pm 0.42$
Carb 23 <sup>a</sup>	Calcite	$0.644 \pm 0.019$	0.703	$25\pm2$	$-33.20\pm0.02$	$-7.62 \pm 0.02$	-5.70	1.0289	$-5.50 \pm 0.42$
Carb 39 <sup>a</sup>	Calcite	$0.616 \pm 0.019$	0.674	$35\pm0.5$	$-30.71 \pm 0.02$	$-9.64 \pm 0.02$	-6.01	1.0271	$-5.58 \pm 0.11$
Carb 39 <sup>a</sup>	Calcite	$0.612 \pm 0.019$	0.670	$35\pm0.5$	$-30.60 \pm 0.02$	$-9.78 \pm 0.02$	-6.01	1.0270	$-5.72 \pm 0.11$
Carb 39ª	Calcite	$0.610 \pm 0.019$	0.668	$35\pm0.5$	$-30.68 \pm 0.02$	$-9.76 \pm 0.02$	-6.01	1.0270	$-5.69 \pm 0.11$
Carb 37 <sup>a</sup>	Calcite+Aragonite	$0.564 \pm 0.019$	0.619	$50\pm0.5$	$-35.09 \pm 0.02$	$-11.66 \pm 0.02$	-6.42	1.0255	$-4.91 \pm 0.11$
Carb 37ª	Calcite+Aragonite	$0.548 \pm 0.019$	0.602	$50\pm0.5$	$-34.92 \pm 0.02$	$-11.97 \pm 0.02$	-6.42	1.0251	$-5.22 \pm 0.11$
Carb 37 <sup>a</sup>	Calcite+Aragonite	$0.560 \pm 0.019$	0.615	$50\pm0.5$	$-35.04\pm0.02$	$-11.65 \pm 0.02$	-6.42	1.0255	$-4.90 \pm 0.11$
Carb 43 <sup>a</sup>	Aragonite+calcite	$0.529 \pm 0.019$	0.583	$65\pm0.5$	$-34.86 \pm 0.02$	$-15.50 \pm 0.02$	-7.29	1.0224	$-6.90 \pm 0.10$
Carb 43 <sup>a</sup>	Aragonite+calcite	$0.507 \pm 0.019$	0.559	$65\pm0.5$	$-34.62\pm0.02$	$-15.59 \pm 0.02$	-7.29	1.0223	$-7.00 \pm 0.10$
Carb 43 <sup>a</sup>	Aragonite+calcite	$0.502\pm0.019$	0.553	$65\pm0.5$	$-34.27 \pm 0.02$	$-15.49\pm0.02$	-7.29	1.0224	$-6.89\pm0.10$
HA 3 <sup>b</sup>	Calcite	$0.77\pm0.022$	0.826	$1\pm0.2$	$-25.47\pm0.00$	$-5.26\pm0.02$	-7.6	1.0333	$-8.40\pm0.07$
HA 1 <sup>b</sup>	Calcite	$0.65 \pm 0.021$	0.701	$23 \pm 1$	$-17.53 \pm 0.01$	$-10.49 \pm 0.01$	-7.47	1.0278	$-8.79 \pm 0.21$
HA 2 <sup>b</sup>	Calcite+Aragonite	$0.71 \pm 0.023$	0.764	$23 \pm 1$	$-24.81 \pm 0.01$	$-10.31 \pm 0.01$	-7.54	1.0280	$-8.61 \pm 0.21$
HA 7 <sup>b</sup>	Calcite	$0.62 \pm 0.023$	0.670	$23 \pm 1$	$-23.74 \pm 0.01$	$-11.22 \pm 0.03$	-7.88	1.0274	$-9.52 \pm 0.23$
HA 9 <sup>b</sup>	Calcite+Aragonite	$0.6 \pm 0.024$	0.649	$33 \pm 2$	$-21.59 \pm 0.01$	$-11.37 \pm 0.01$	-7.37	1.0267	$-7.70 \pm 0.39$
HA 12 <sup>b</sup>	Calcite+Aragonite	$0.55 \pm 0.024$	0.598	$50 \pm 2$	$-21.58 \pm 0.02$	$-15.62 \pm 0.04$	-8.09	1.0231	$-8.90 \pm 0.38$
HA 4 <sup>b</sup>	Calcite+Aragonite	$0.55\pm0.022$	0.598	$50\pm2$	$-26.38 \pm 0.00$	$-14.70\pm0.02$	-7.45	1.0234	$-7.97\pm0.36$

Table 1 Data of CaCO<sub>3</sub> laboratory precipitation experiments from this and the Ghosh et al. (2006a) studies.

<sup>a</sup> Data from this study.

<sup>b</sup> Data from this study. <sup>b</sup> Data from Ghosh et al. (2006a). <sup>c</sup> <sup>18</sup> $\alpha = (\delta^{18}O_{(carbonate)} + 1000)/(\delta^{18}O_{(water-measured)} + 1000).$ <sup>d</sup>  $\delta^{18}O_{(water-calculated)}$  is calculated using the measured  $\delta^{18}O_{(carbonate)}$  and <sup>18</sup> $\alpha$  values derived from Eqs. (1)–(3) using the temperature measured during precipitation experiment. <sup>\*</sup> The first polymorph in the mineralogy column is the major phase in the sample.



**Fig. 1.** Data points show carbonate-water oxygen isotopic fractionation factors ( ${}^{18}\alpha$ ) determined for synthetic carbonates in this study and in that of Ghosh et al. (2006a). The lines show accepted  $\delta^{18}$ O-*T* fractionation lines (Eqs. (1), (2) and (3) in the text).

of shot noise error (Poisson error), which is a mass spectrometric source of uncertainty and is the limiting case for measurement error, and of potential instrument instability at a time scale of a few hours. The Poisson error,  $SE_p(\Delta_{47})$ , can be determined directly from mass spectrometric parameters using the approach derived in SI1. To summarize,  $SE_p(\Delta_{47}) = SE_0(\Delta_{47})/t^{1/2}$ , where *t* is the sample count time, and  $SE_0(\Delta_{47})$  is the Poisson standard error for 1 s of count time.  $SE_0(\Delta_{47})$  is typically 0.357‰ for 1 s of signal counting. Note that this value is instrument dependent; see SI1 for the full calculation. Our measurement protocol of 1800 s count time, results in  $SE_p(\Delta_{47}) = 0.0084\%$ . Based on the Ghosh et al. (2006a) reported count time for each sample, the estimated  $SE_p(\Delta_{47})$  for their synthetic carbonate samples ranges between 0.0146‰ and 0.010‰.

The external precision of a measurement represents the combination of both internal errors, and other errors, "external" to the instrument. We attribute these errors to mass spectrometer instability at a time scale of days to months, as well as the uncertainties associated with sample preparation and possibly to factors that are specific to a particular type of sample (e.g., inhomogeneity).

We use long term (3 years) and short term (1–3 days) measurements of standard materials (Table 2) to assess the reproducibility of data and to test for instrumental drifts that would require applying standard corrections to our measured sample values. Long term standards include individual CO<sub>2</sub> extractions from Carrara Marble, individual preparation of cylinder CO<sub>2</sub> equilibrated with water at 25 °C (termed "25-CO<sub>2</sub>"), and aliquots of cylinder CO<sub>2</sub> gas (termed "cylinder CO<sub>2</sub>"). The long term standard measurements indicate that despite individual analyses being close to the shot noise limit, the variance among data points is significantly higher with  $SD/SE_p$ , of  $\sim$ 3–4.

Short term standard measurements are aliquots of cylinder CO<sub>2</sub> that were measured as individual samples over the course of a few days in 2 approaches: The first included measurements of multiple CO<sub>2</sub> samples that were prepared and analyzed individually. This was repeated independently 3 times and is referred to in Table 2 as "Cylinder  $CO_2$  expt. 1–3". The second was designed to eliminate the preparation step of individual aliquots by introducing large aliquots of cylinder gas into the bellow and measuring it as replicates. In this test, cylinder gas was measured with a close match between sample and reference gas bellow compression. This was repeated independently twice and is referred to in Table 2 as "Cylinder CO<sub>2</sub> batch 1–2". The  $SD/SE_p$  observed in the Cylinder CO<sub>2</sub> expt. (2.9, 1.3 and 2.1; Table 2) reflect uncertainty that is introduced in sample preparation.  $SD/SE_p$  are slightly lower in the cylinder CO<sub>2</sub> batch analysis (1.2 and 2.1; Table 2). The batch measurements rule out pressure imbalance between sample and standard as a major source of uncertainty. The results of this experiment suggest that when sample preparation is eliminated, homogeneous materials like a pure CO<sub>2</sub> gas may result in total precision closer to the shot noise limit. Even though the uncertainties without sample preparation are closer to shot noise, there is a contribution of non-shot-noise mass spectrometric component in inter-replicate variability.

We apply the Peirce elimination criterion (Peirce, 1852; Ross, 2003) to the measurements of each standard in the long term data sets to screen for data points that have exceptionally large deviations from the mean relative to the probability of such deviations given the observed distribution of the data. Using this criterion, we exclude  $\sim 2\%$  of the data: 3 measurements of Carrara Marble, 1 of 25-CO<sub>2</sub>, and 5 of the cylinder CO<sub>2</sub>. Comparisons of the original and 'Peirced' means in Table 2 show only minor changes – in the third decimal place of the mean values. No data in the short term experiments were identified as outliers.

Our 3-year data set shows occasional time periods during which the variance was higher than average (perhaps reflecting instrumental instability), however, no trends in the distribution of data are observed and this external inter-sample variance is characterized by random variation. The Durbin–Watson (DW) test was used to test for serial correlations as a function of time (see SI2 for details about the DW test). Table 2 shows the DW statistics. The values cluster around 2, indicting no serial correlations.

In addition to their use for monitoring instrument stability, standards are also used for establishing a reference frame for  $\Delta_{47}$  data (Dennis et al., 2011; Huntington et al., 2009) and have been applied for data correction in some cases (e.g., Affek et al., 2008; Passey et al., 2010; Tripati et al., 2010). We therefore routinely measure standard materials and use them to standardize our reference frame, for inter-laboratory comparisons, and to test for instrumental stability. The random nature of data distribution, however, suggests that data correction based on short time-scale standards measurements is likely to increase the uncertainty (due to the

#### Table 2

Statistical analysis of long term measurements of different standard materials and short term multiple measurements of cylinder CO<sub>2</sub>.

Material	n	Mean	SD	# Peirced	P. Mean <sup>a</sup>	P. SD <sup>a</sup>	DW	$SD/SE_p$
Carrara Marble	119	0.362	0.036	3	0.358	0.029	1.7	4.3
25-CO <sub>2</sub>	55	0.849	0.032	1	0.851	0.030	1.6	3.9
Cylinder CO <sub>2</sub>	229	0.876	0.030	5	0.872	0.023	1.9	3.5
Cylinder CO <sub>2</sub> expt. 1	19	0.890	0.024	1	0.886	0.017	2.2	2.9
Cylinder CO <sub>2</sub> expt. 2	16	0.887	0.011	0	-	-	2.4	1.3
Cylinder CO <sub>2</sub> expt. 3	14	0.872	0.017	0	-	-	2.2	2.1
Cylinder CO <sub>2</sub> batch 1	7	0.868	0.010	0	-	-	1.4	1.2
Cylinder CO <sub>2</sub> batch 2	7	0.866	0.017	0	-	-	1.6	2.1

# Peirced refers to the number of samples eliminated using the Peirce Criterion.

<sup>a</sup> Are the mean and standard deviation after application of the Peirce Criterion ('Peirced').

added uncertainty in the standards analyses) rather than correct for drifts.

Even though our data suggests that a standard correction is not required, this may not be the case for other instruments; long term mass spectrometer and standards behavior should be characterized for each laboratory. Additionally, regular standard measurements are required to identify trending data that may be corrected for. That can only be done in retrospect when assessing variance and trends within a long term data set.

To conclude, reducing  $\Delta_{47}$  measurement uncertainties to the often desirable values of ~0.010‰ (e.g., for paleoclimate studies) requires minimizing both Poisson and non-Poisson errors, through an optimal combination of count time and replicate analyses. Since the Poisson uncertainties are already close to the limit of measurement error, the total precision can be most effectively improved by reducing the non-Poisson errors. This can be done by identifying the source of errors in sample preparation and reducing it and through replicate analyses. A detailed discussion on strategies and considerations for optimizing the use of mass spectrometer time, while reaching the desired uncertainties can be found in SI3.

#### 3.3. Uncertainty in calibration samples

Standard materials are measured many times over a long period of time. Having large data sets allow for statistical characterization of the measurements. Unlike the standards, which are measured routinely, usually only a few (3–5) replicates of samples are measured. This small sample size does not allow for a robust statistical characterization of each sample. Furthermore, comparison of our 3 standard materials (Table 2) shows that the variance can vary among materials, such that a "general variance" behavior cannot be assumed for all materials.

The replicate analyses of a large data set of related samples can be used to account for potential differences among different sample types. This approach provides a more reliable estimate of the variance within a specific group of samples. The variance is calculated using the deviations of individual replicates from the mean value of the replicated sample (see SI5). In this way, the variance of a specific group of materials is characterized with the understanding that some samples within the group replicate better than others. This is also helpful in cases where certain samples are too small to be replicated, since a general uncertainty is estimated for all samples in the group. Applying this method to our synthetic carbonates data set results in analytical error of 0.017‰ per replicate.

None of the synthetic carbonates in the original clumped isotope calibration study (Ghosh et al., 2006a) were analyzed in replicates, and only the internal uncertainties are reported there. For these data we also use 0.017‰ as the non-Poisson error (see Section 4.2.1).

#### 4. Discussion

The focus and purpose of this study is to characterize the  $\Delta_{47}$ -T relationship, compare new data to the original calibration (Ghosh et al., 2006a), and test the implications of this revised calibration to geological questions. Comparing the new data to the original calibration through an error weighted regression requires an evaluation and standardization of the analytical uncertainties for each of the data sets. In the following sections, and in more details in the online supplementary material, we describe a statistical analysis of uncertainties in clumped isotope measurements, followed by the revised calibration and its implications.

#### 4.1. Oxygen isotope fractionation between carbonate and water

We compare the carbonate  $\delta^{18}$ O values obtained here to the carbonate  $\delta^{18}$ O thermometry calibrations of Kim and O'Neil (1997), Friedman and O'Neil (1977), and Kim et al. (2007b) by calculating the carbonate-water fractionation factors using Eqs. (1)-(3) (Fig. 1, Table 1).  $\delta^{18}$ O values obtained as part of the clumped isotopes precipitation experiments generally agree with these previous studies. The fractionation observed in the current study, however, is slightly higher than that in Kim and O'Neil (1997) and Kim et al. (2007b) (with a somewhat better fit to Friedman and O'Neil, 1977), whereas the synthetic carbonate samples of Ghosh et al. (2006a) are slightly lower. These deviations cannot be attributed to differences in precipitation mechanisms, as all these studies followed the same precipitation method. The deviations could result from some unrecognized water evaporation from the precipitation solution, in particular in the Ghosh et al. (2006a) experiments. In the current study, to the best of our ability, we minimized evaporation of the solution by capping the precipitation vessel with only minimal opening to allow the N<sub>2</sub> purging gas to escape, and by humidifying the incoming N<sub>2(g)</sub>. Alternatively, the slight inconsistencies between the three studies may be related to differences in carbonate precipitation rates. Dietzel et al. (2009) and Gabitov et al. (2012) observed lower fractionation in faster growing calcite minerals, which may indicate preferential uptake of <sup>16</sup>O into the fast growing mineral. The difference between the precipitation experiments may then be due to slight differences in supersaturation and mineral growth rates among the different studies. Whereas it is difficult to determine that precipitation is at equilibrium, the general consistency of these carbonate samples with the accepted  $\delta^{18}$ O–*T* calibrations, together with the consistency of the clumped isotope experiments with each other and with biogenic carbonates (see below), suggest that the calibration reflects precipitation close to isotopic equilibrium conditions. Namely, dissolved inorganic carbon (DIC) in the precipitating solution has undergone full oxygen isotope exchange with water. As this isotope exchange is also the rate-limiting step for clumped isotopes equilibrium (Affek, 2013), the  $\delta^{18}$ O consistency may be extended to assume clumped isotopes equilibrium. We therefore consider this  $\Delta_{47}$ -T relationship as nominal equilibrium.

#### 4.2. Clumped isotopes thermometer calibration

#### 4.2.1. Synthetic CaCO<sub>3</sub>

The original calibration data of the  $\Delta_{47}$  thermometer was measured at a relatively low precision, lower than is typically acceptable today. As a result, efforts to reduce the uncertainties in temperatures derived from clumped isotopes are limited by a relatively large calibration error. This may be overcome by reducing the calibration uncertainties through additional calibration data points that are analyzed at higher analytical precision. Our laboratory precipitation experiments were designed to provide such data, by following the same method of Ghosh et al. (2006a), while using a wider temperature range (5–65 °C), a more even temperature distribution, and measuring at higher analytical precision (through longer count time and triplicate analyses). The new data set has a  $\Delta_{47}$ -T relationship that is similar, but has a slightly lower slope, than the original thermometer calibration, with a slope and associated 1 SE uncertainty of  $0.0526 \pm 0.0025$  and an intercept of  $0.0520 \pm 0.0284$ . For comparison, the calibration using only the Ghosh et al. (2006a) data gives a slope and associated 1 SE of  $0.0597 \pm 0.0083$  and an intercept of  $-0.0239 \pm 0.0925$ . Note that these values are slightly different than those reported by Ghosh et al. (2006a) as they reflect an error weighted regression to the original data and no data averaging as was done in the original study.



**Fig. 2.** (a)  $\Delta_{47}$ –*T* calibration line (heavy black line) estimated using synthetic carbonate data from this study and from Ghosh et al. (2006a). Thick gray lines are theoretical predictions for calcite (solid) and aragonite (long dash), respectively (Guo et al., 2009). The dashed black curves show the 95% confidence envelope for the estimate of the calibration line. The solid black curves show the 95% confidence envelope for a temperature prediction for a  $\Delta_{47}$  measurement with an assumed  $SE(\Delta_{47}) = 0.025\%$ . (b) Graphic summary of 95% uncertainty for a temperature estimated for an "unknown" sample using the synthetic carbonate calibration. The uncertainty is shown as a function of temperature (horizontal axis) and standard error (contours) for the "unknown"  $\Delta_{47}$  measurement. The uncertainty for the calibration estimate alone (calibration error) is shown by the  $SE(\Delta_{47}) = 0\%$  contour. The circle-cross symbol marks the data centroid of the least-squares estimate. The centroid (~27 °C) marks the minimum in the calibration error, which is  $\pm 2^\circ$ C.

As the two sets of samples were precipitated using the same method and the data statistically overlap, we consider them as two statistical realizations of the same  $\Delta_{47}$ -T relationship that can be combined into one thermometer calibration. The two data sets are then combined using an error weighted regression (see SI2) to establish a revised calibration (Fig. 2):

$$\Delta_{47} = (0.0526 \pm 0.0025) 10^6 / T^2 + (0.0520 \pm 0.0284) (1 \text{ SE}) R^2 = 0.93$$
(4)

The weighted least-squares estimate was done in two steps. First, solving for a best-fit calibration using the Poisson standard errors alone,  $SE_p(\Delta_{47})$ , for weights. That estimate resulted in residuals that were much larger than the predicted Poisson errors. Second, using the residuals to estimate a separate non-Poisson standard error  $SE_r(\Delta_{47})$ , which was added to the weights. With this approach, each of i = 1 to n samples is weighted by an estimate of its total standard error,

$$SE^2(\Delta_{47,i}) \approx \frac{SE_0^2(\Delta_{47,i})}{t_i} + \frac{SE_r^2(\Delta_{47,i})}{m_i}$$
 (5)

where  $m_i$  is the number of replicate measurements averaged for the  $\Delta_{47,i}$  measurement, and  $t_i$  is the total sample count time for all replicate measurements combined. This approach indicates  $SE_r(\Delta_{47}) \approx 0.017\%$ , which is consistent with the non-Poisson errors observed for the time-series of the standards measurements (Table 2). For comparison, this estimate indicates a ratio  $SD/SE_p \approx$ 2.0 for a  $\Delta_{47}$  measurement with 1800 s of sample count time.

The calibration and prediction uncertainties were calculated using the inverse-regression method (see SI4). Fig. 2(b) shows the estimated 95% uncertainty for a temperature predicted from  $\Delta_{47}$  as a function of *T* (on the horizontal axis) and SE of the  $\Delta_{47}$  measurement for the "unknown" sample (represented as contours). The zero contour shows the precision of the calibration alone. The 95% uncertainty for the temperature calibration is  $\pm 2$  °C at the data centroid of 27 °C. For example, an "unknown" sample with a measurement precision of SE ~0.020‰ results in 95% uncertainty of  $\pm 11$  °C at the data centroid. The non-linearity of the thermometer, being scaled as  $1/T^2$ , results in a larger uncertainty at high temperatures.

Most of the calibration samples are calcitic, but one of them (#43) is predominately aragonite (Table 1). In comparing the bestfit result to our data with and without the aragonitic sample, we find that the calculation without the aragonitic sample is consistent with Eq. (4). The similarity in the results, together with the agreement of both calcitic and aragonitic biogenic carbonates with the calibration (Section 4.3.2 below), suggest that within our measurement uncertainty these two polymorphs behave the same with respect to  $\Delta_{47}$ , despite theoretical predictions for differences (Guo et al., 2009).

#### 4.2.2. Comparison with biogenic carbonates

Since the publication of the original  $\Delta_{47}$ -T calibration relationship, a number of studies have shown the applicability of this thermometer to biogenic carbonates, by measuring the  $\Delta_{47}$ values of modern biogenic carbonates grown at known temperatures (Came et al., 2007; Eagle et al., 2010; Ghosh et al., 2007; Grauel et al., 2013; Thiagarajan et al., 2011; Tripati et al., 2010). The general conformity of these biogenic materials to the thermometer calibration, irrespective of mineral type, leads to this relationship being considered as reflecting nominal isotopic equilibrium (Eiler, 2011; Tripati et al., 2010). This relies on the assumption that the <sup>13</sup>C-<sup>18</sup>O bond distribution is primarily determined within the solution phase (Guo et al., 2009). It has been noted, however, that at low temperatures biogenic materials deviate somewhat from the calibration line, showing lower than expected  $\Delta_{47}$  values (Eiler, 2011). The revised  $\Delta_{47}$ -T relationship proposed here (Eq. (4)) agrees better with the published biogenic carbonates at both moderate and low temperatures (Fig. 3).

A notable exception to the biogenic carbonate conformity are fast growing shallow water corals and mollusks measured in some laboratories. The discrepancy in corals was first noted by Ghosh et al. (2006a) who observed higher than expected  $\Delta_{47}$  values in winter growth bands of a coral from the Red Sea. Similar offsets were observed also in shallow water corals from other locations

Table 3

The quality of the fit of the different biogenic carbonate groups to the synthetic carbonate  $\Delta_{47}$ -*T* calibration line. The fit is assessed by a  $\chi^2$  test for the scatter of data points in each group with respect to the calibration curve.

Biogenic carbonate	n	Reduced $\chi^2$	Probability $(\chi^2)^{**}$
Brachiopods (Came et al., 2007)*	3	0.225	88%
Cocoliths (Tripati et al., 2010) <sup>*</sup>	2	0.022	98%
Deep Sea Coral (Thiagarajan et al., 2011)	25	0.902	60%
Foraminifera (Tripati et al., 2010)	34	0.478	100%
Mollusks (Came et al., 2007) <sup>*</sup>	3	0.072	98%
Otoliths (Ghosh et al., 2007)	12	0.704	75%
Surface Coral (Ghosh et al., 2006a) <sup>+</sup>	30	4.700	0%
Surface Coral (Saenger et al., 2012) <sup>+</sup>	21	15.972	0%
Teeth (Eagle et al., 2010)	28	0.652	92%

<sup>\*</sup> A reduced  $\chi^2$  value of 1 and probability of 50% reflect dispersion that is similar to the synthetic carbonate data base.

<sup>\*</sup> High uncertainty to the  $\chi^2$  test estimate due to the small size of the data sets.

<sup>+</sup> Examples of materials that do not fit the synthetic carbonates calibration.



**Fig. 3.** Comparison of observed temperature estimates and predicted temperatures calculated from  $\Delta_{47}$  for published biogenic carbonates. Error bars are at the 95% confidence level. The 1:1 line is included as a reference. Fast growing shallow water corals are given as an example of a biogenic carbonate that is not in agreement with the calibration line.

and are likely to result from kinetic or 'vital' effects that are related to the very high coral growth rate and to specific processes in coral calcification (Saenger et al., 2012). Some laboratories have found anomalously low  $\Delta_{47}$  values in mollusks (Eagle et al., 2013; Henkes et al., 2013). This discrepancy is still unresolved but it may be related to the different sample preparation methods used in these studies (Eagle et al., 2013).

We use the  $\chi^2$  test (see SI2) to test how well biogenic data with known temperatures compare with the calibration derived from synthetic carbonates. The objective is to check whether the biogenic data are consistent both in trend and variance with the synthetic calibration data. The  $\chi^2$  probabilities in Table 3 are close to the expected value of 50%, indicating that the biogenic data are consistent with the synthetic carbonates data. A notable exception is the foraminifera data (Tripati et al., 2010), which have low variance (high  $\chi^2$  probability in Table 3) relative to the synthetic carbonates and the other biogenic groups. The foraminifera data are otherwise consistent with the trend of the synthetic calibration line. It is possible that the pre-treatment cleaning specific to this foraminifera study eliminated the source of some of the non-Poisson error but the reason for the low variance remains unclear. There are other biogenic groups that have high  $\chi^2$  probabilities; these groups, however, are difficult to assess given their small sample sizes.

The agreement of most biogenic materials with the calibration relationship at the full temperature range provides further support for this revised carbonate  $\Delta_{47}-T$  relationship being a relevant thermometer calibration for these biogenic materials in spite of differences in the calcification mechanisms and irrespective of mineralogy.

#### 4.2.3. The calibration at low temperatures

Low temperature carbonates, for example North Atlantic foraminifera (Tripati et al., 2010), disagree with the original calibration and have  $\Delta_{47}$  values that are too low at the low temperature end. This can be explained by the synthetic carbonates grown at these low temperatures reflecting precipitation under equilibrium conditions, with deviations from equilibrium in the biogenic carbonates (Ghosh et al., 2007; Thiagarajan et al., 2011; Tripati et al., 2010), or vice versa. Theoretical ab initio calculations of  $\Delta_{47}$  temperature dependence, although suffering from considerable uncertainty due to the complexity of solid state calculations, suggest a slope that is lower than the empirical calibrations (Guo et al., 2009; Schauble et al., 2006), supporting potential non-equilibrium in the empirical synthetic carbonate calibration relationship. On the other hand, the general conformity to the calibration line of numerous biogenic carbonates of different types of organisms, calcification mechanisms, and mineral polymorphs supports the calibration as reflecting equilibrium.

The revised calibration line presented here significantly reduces this discrepancy, as the low temperature samples are now closer to the clumped isotope thermometer calibration. The new calibration data set includes additional low temperature samples that fill the gap between the 23 °C and 1 °C samples of the Ghosh et al. (2006a) data set, thus making the low temperature portion of the calibration more robust. The discrepancy between the original calibration and the low temperature biogenic carbonates is therefore likely due to the large uncertainty ( $\pm 0.021\%$ ) in the  $\Delta_{47}$  value of the original calibration data point HA3 that was precipitated at 1 °C and the lack of other low temperature portion of the calibration.

#### 4.2.4. $\Delta_{47}$ -T calibration including biogenic carbonates

Having established that biogenic carbonates (with the exception of shallow water hermatypic corals; Table 3) agree with the synthetic carbonates calibration, these biogenic data points can be integrated into the calibration calculation. As these materials show the same statistical behavior as the synthetic carbonates, combining them would not significantly change the calibration equation but will reduce the calibration uncertainty. The resulting combined calibration is (Fig. 4):

$$\Delta_{47} = (0.0539 \pm 0.0013)10^6 / T^2 + (0.0388 \pm 0.0155) (1 \text{ SE}) R^2 = 0.92$$
(6)



**Fig. 4.** (a)  $\Delta_{47}-T$  calibration line (heavy black line) estimated for a combination of synthetic and biogenic carbonates (data from Table 3 excluding shallow water corals). The 95% confidence curves are the same as in Fig. 2(a). (b) Graphic summary of 95% uncertainty for a temperature estimated for an "unknown" sample for the calibration estimate using both synthetic and biogenic carbonates (excluding shallow water corals). See Fig. 2(b) for explanation of the plot. Note that the minimum in the calibration error has shifted with the data centroid to a temperature of ~20°C. The minimum 95% uncertainty for the calibration estimate is now ~0.6°C.

The temperature calibration uncertainty at the 95% confidence level is  $\pm 0.6$  °C at the centroid of 20 °C. Note that this dataset gives the same estimate of 0.017% for *SE<sub>r</sub>*, which indicates that this is a reasonable estimate for non-Poisson standard error for a single replicate measurement in these data sets.

#### 4.2.5. Absolute reference frame

In order to facilitate comparison with the Ghosh et al. (2006a) and previously published biogenic carbonate data, we report our data above normalized to the same reference frame used in the Ghosh et al. (2006a) work. Recently, an absolute reference frame for clumped isotope measurements has been established, aimed at standardizing data among laboratories (Dennis et al., 2011). This reference frame accounts for mass spectrometric effects, making it also relevant for comparison with theoretical  $\Delta_{47}$  values. For fu-

ture use, we take advantage of the standards measured together with our synthetic carbonates to provide the revised calibration equation also in the absolute reference frame (Tables 1 and 4). The synthetic carbonate based calibration (Eq. (4)) converted to the absolute reference frame is:

$$\Delta_{47,abs} = (0.0555 \pm 0.0027) 10^6 / T^2 + (0.0780 \pm 0.0298) (1 \text{ SE}) R^2 = 0.93$$
(7)

Unfortunately, no long term standards are reported with the biogenic studies, preventing direct conversion to the new reference frame, without the addition of considerable uncertainty. However, as the biogenic data do not significantly change the calibration line, it is likely to be consistent also at the absolute reference frame. As additional modern biogenic carbonate data become available they can be integrated into the absolute reference frame calibration.

#### 4.2.6. Other laboratory precipitation experiments

Dennis and Schrag (2010) precipitated synthetic carbonates using a different experimental approach to both the preparation of the initial solution and CO<sub>2</sub> degassing. The slope of their  $\Delta_{47}$ -T relationship is significantly lower than the original calibration and our calibration. Most data in that study at both low and high temperatures are inconsistent with the calibration line (Fig. 3). The cause of this discrepancy has not been resolved to date, but it is likely the result of differences in experimental procedures. In our study, carbonate was precipitated by bubbling  $N_{2(g)}$  through a saturated HCO<sub>3</sub> solution. In the Dennis and Schrag (2010) study, carbonates were precipitated by titrating NaHCO<sub>3</sub> and CaCl<sub>2</sub> solutions, with CO<sub>2</sub> passively diffusing out of the final solution to the atmosphere. These different methods may result in a different amount of disequilibrium, but, as it is nearly impossible to prove equilibrium, it is also difficult to assess the level of disequilibrium associated with the two procedures. Alternatively, the discrepancy could be the result of differences in the clumped isotopes laboratory measurement procedures, similarly to those suggested to explain discrepancies in mollusks (Eagle et al., 2013). The measurements conducted in our study, followed the Ghosh et al. (2006a) sample preparation procedure of overnight carbonate digestion by phosphoric acid at 25 °C. Carbonate digestion in the Dennis and Schrag (2010) study was conducted using a common acid bath at 90°C. Differences in laboratory procedures may explain the discrepancies, for example due to incomplete understanding of the  $\Delta_{47}$  fractionation factor associated with acid digestion at different temperatures. Regardless of the reason, the inconsistency of the Dennis and Schrag (2010) synthetic carbonate data with the majority of biogenic carbonate data makes this relationship irrelevant as a thermometer calibration for these materials.

#### 4.3. Implications to previous studies

The revised calibration (Eq. (4)) agrees with the original calibration to within 1 °C between  $\Delta_{47}$  values of 0.59‰ and 0.65‰, corresponding to a temperature range of 39 °C to 25 °C. The difference between the calibrations becomes significant at low temperatures (e.g., an offset of >2 °C for materials growing at <16 °C, corresponding to  $\Delta_{47}$  values above 0.68‰), and is relevant to paleoclimate applications at the glacial–interglacial time scale and at high latitudes or high elevation during warmer climatic states.

In this section, we revisit a number of previously published studies that are characterized by  $\Delta_{47}$  values outside the 'safe' range, particularly at low temperatures, in order to illustrate the potential impact of our revised calibration. It is worth noting that the changes at low temperatures resulting from the revised calibration are not large enough to account for the offsets observed by Henkes et al. (2013) and Eagle et al. (2013) in mollusks.

#### Table 4

 $\Delta_{47}$  values of standards used for conversion from the original to the absolute reference frame. The Carrara Marble, NBS-19, CO<sub>2</sub>-H<sub>2</sub>O (25 °C) and Heated Gas values for the absolute reference frame values follow Dennis et al. (2011).

Material	Long term avg.	Absolute reference frame
Carrara Marble	0.358	0.395
CO <sub>2</sub> -H <sub>2</sub> O (25 °C)	0.851	0.925
Heated Gas	0.000	0.027
Cylinder CO <sub>2</sub> *	0.872	0.945
NBS 19 <sup>+</sup>	0.352	0.392

 $^+$  Heated Gas and NBS-19 measured with the Ghosh et al. (2006a) data are used to transfer these data to the absolute reference frame.

<sup>\*</sup> The first three materials are used to establish a value for the cylinder gas in the absolute reference frame, creating an additional in-house standard material.

Discrepancies occur also at high temperatures, where they are relevant for petrologic applications (such as Dennis and Schrag, 2010; Huntington et al., 2010; Swanson et al., 2012). The calibration uncertainty at high temperatures can be estimated by extrapolating the error envelop using the equations derived in SI4. The uncertainty at this temperature range is significantly higher than we discussed so far, as it is far from the centroid of the calibration. Furthermore, the calibration extrapolation required at this temperature range limits their precision irrespective of which calibration is used.

Low  $\Delta_{47}$  values reflecting higher temperatures than expected for certain environments (e.g., >40 °C for oceanic environments) have been used as an indicator for potential diagenetic alteration. The revised calibration suggests that this threshold should be shifted to higher  $\Delta_{47}$  values (0.589‰), reducing the range of  $\Delta_{47}$ values acceptable for well-preserved samples. Below we illustrate the effect of the revised calibration and uncertainties estimate for different cases, by examining 3 examples of previously published studies.

#### 4.3.1. Rate of Altiplano uplift (Ghosh et al., 2006b)

One of the first applications of carbonate clumped isotope thermometry to a geologic question is the study of the paleotopography of the Alitplano (Ghosh et al., 2006b). Paleosol carbonates of three age groups 11.4–10.3, 7.6–7.3 and 6.7–5.8 Ma from the Bolivian Altiplano were analyzed for  $\Delta_{47}$ . Temperatures were derived based on the Ghosh et al. (2006a)  $\Delta_{47}$ –T calibration, assuming it reflects the formation temperature of paleosol carbonates. More recent clumped isotopes paleosol studies (e.g., Passey et al., 2010; Peters et al., 2013; Quade et al., 2013) discuss seasonal formation of soil carbonates that complicates the interpretation of the paleosol  $\Delta_{47}$  signal. For our discussion here, we treat the Altiplano data solely from the calibration perspective.

Based on the  $\Delta_{47}$  data, Ghosh et al. (2006b) estimated an average uplift rate of  $0.94 \pm 0.17$  mm/year, for a total uplift of 3.4 km between 10.3 and 6.7 Ma. This is based on an estimated change in land surface temperature from  $28 \pm 1$  °C to  $13 \pm 3$  °C through this time interval. Using the revised calibration, the same  $\Delta_{47}$  values result in lower temperatures for the younger paleosols (11 °C) while the older ones do not significantly change (Table 5). We estimate the uncertainty of these data by assuming the analytical parameters that were customary at the time of these measurements (similar to Ghosh et al., 2006a synthetic carbonates discussed above), and then we treat all the samples within an age group as replicates (n = 8, 2, 4 for the different age groups), resulting in a temperature uncertainty of 2–3 °C (1 SE).

The revised temperatures do not significantly change the main conclusions of Ghosh et al. (2006b). The increase in estimated uncertainty makes the absolute temperatures less robust but as the temperature differences among age groups are large, the increased uncertainty does not significantly change the conclusions. The agreement of the total estimated uplift with the current land surface temperature of the Altiplano lends further support to the derived temperatures.

#### 4.3.2. Late Ordovician–Early Silurian glaciation (Finnegan et al., 2011)

Finnegan et al. (2011) used carbonate clumped isotopes to constrain sea surface temperatures through the Late Ordovician-Early Silurian glaciation event, a time window of  $\sim 20$  Ma around the 445 Ma Hirnantian glaciation peak. These temperatures were combined with  $\delta^{18}O_{carbonate}$  values to estimate the change in ice volume. The reconstructed baseline temperatures before and after the glaciation, range between 32 °C and 37 °C, and temperatures during the glaciation event range between 28-31 °C. Whereas at this temperature range the revised  $\Delta_{47}$ -T calibration does not significantly affect the reconstructed temperatures, the analytical errors are large, such that the difference between baseline and event temperatures may not be robust. Samples in this study were measured in 1 to 5 replicates, resulting in estimated  $\Delta_{47}$  uncertainties of 0.022-0.010% (1 SE), or a 95% temperature uncertainty of 4–7 °C. In light of the glaciation event being characterized by only a few data points, these large analytical uncertainties result in a temperature overlap between background and event and reduce the calculated glaciation  $\delta^{18}O_{water}$  signal. The large  $\delta^{18}O_{carbonate}$  signal of the event is sufficiently large to support the occurrence of a glaciation event irrespective of reconstructed temperatures, though the quantitative ice volume estimate is uncertain. In order to confidently identify a relatively small temperature signal, such as observed in this study, it is recommended to measure a larger number of samples during the event itself and analysis should be replicated to reduce the analytical uncertainty.

As a way to identify samples that are diagenetically altered, the authors considered a combination of high trace metals concentrations and low  $\Delta_{47}$  values as indicative of alterations.  $\Delta_{47}$  values lower than 0.589‰ were designated as the threshold for altered samples, corresponding to 39 °C or to 41 °C using the original and revised calibrations, respectively. In this case, the results using the two calibrations do not differ significantly from one another. However, temperature limits for life can be translated into practical limits of  $\Delta_{47}$  values in biogenic carbonates.

#### 4.3.3. Early Pliocene arctic temperatures (Csank et al., 2011)

Csank et al. (2011) estimated Early Pliocene arctic temperatures by measuring clumped isotopes in fresh water mollusk shells from Ellesmere Island, Canada, and compared them to estimates obtained from cellulose  $\delta^{18}$ O. Using the Ghosh et al. (2006a) cali-

Table 5

Paleosol data associated with the Altiplano uplift (Ghosh et al., 2006b) recalculated using the revised  $\Delta_{47}$ -T calibration.

Age Ma	$\Delta_{47}$ (%)	T <sup>a</sup> (°C)	Altitude <sup>a</sup> (km ASL)	Uplift rate <sup>a</sup> (mm/yr)	<i>T</i> <sup>b</sup> (°C)	Altitude <sup>b</sup> (km ASL)	Uplift rate <sup>b</sup> (mm/yr)
11.4-10.3	$0.631 \pm 0.009$	$28 \pm 2$	0.1		$28\pm2$	0.1	
7.6-7.3	$0.680 \pm 0.012$	$16 \pm 3$	2.7	0.9	$18 \pm 3$	2.4	0.9
6.7-5.8	$0.706 \pm 0.008$	$11 \pm 2$	3.9	1.3	$13 \pm 2$	3.5	1.2
0		10	4.0	0.0	10	4.0	0.1

<sup>a</sup> Estimate using the revised calibration developed in this study.

<sup>b</sup> Estimates using the Ghosh et al. (2006a) calibration.

#### Table 6

The temperatures derived from the  $\Delta_{47}$  measurements of Csank et al. (2011) using the Ghosh et al. (2006a) and the revised calibrations compared with the temperatures derived in that study from mosses cellulose.

	$\Delta_{47}$ (%))	T <sup>a</sup> (°C)	<i>T</i> <sup>b</sup> (°C)	Cellulose T (°C)	$\Delta T$
	$0.695 \pm 0.022$	$13.0\pm5$	14.6	14.2	1.6
	$0.716 \pm 0.022$	$8.4 \pm 5$	10.5	7.1	2.1
	$0.701 \pm 0.022$	$11.6 \pm 5$	13.4	19.2	1.8
	$0.73 \pm 0.022$	$5.5\pm4$	7.8	17.45	2.3
	$0.719 \pm 0.022$	$7.8\pm5$	9.9	10.5	2.1
	$0.744 \pm 0.022$	$2.6 \pm 4$	5.2	14.7	2.6
				15.2	
				15	
Avg.	$0.718\pm0.009$	8.1	10.2	14.2	2
SD	0.018	3.8	3.5	3.8	0

<sup>a</sup> Temperature estimates using the revised calibration developed in this study.

<sup>b</sup> Temperature estimates using the Ghosh et al. (2006a) calibration.

bration, the reconstructed temperature is  $10 \pm 3.5$  °C (1 SE, m = 6). The revised calibration (Table 6) reduces the temperature estimate to  $\sim 8 \,^{\circ}$ C, in better agreement with estimates of other studies in this region (as discussed by Csank et al., 2011), making the Early Pliocene temperatures  $\sim 10 \,^{\circ}$ C warmer than today. Although the difference between the calibrations is small, it increases the difference between temperatures derived from clumped isotopes and those derived by these authors using a combination of  $\delta^{18}$ O in moss cellulose and mollusk aragonite shells (8 °C versus 14 °C, which is significantly different at the 95% confidence level, based on a Student's *t*-test). This difference challenges the assumption that mollusk  $\Delta_{47}$  and moss cellulose  $\delta^{18}$ O reflect the same seasonal preference. The light requirements for moss cellulose formation may bias that proxy toward summer conditions whereas the mollusks may be limited by either temperature or food supply and may reflect a longer growing season.

#### 5. Conclusions

We present a  $\Delta_{47}$  data set of synthetic carbonates precipitated experimentally under controlled temperatures, principally repeating the Ghosh et al. (2006a) experiment at higher analytical precision, and at a wider and more evenly spaced temperature distribution. We revise the carbonate clumped isotope thermometer to a  $\Delta_{47}$ -T relationship that is similar but with a slightly lower slope than the original calibration. The calibration using only laboratory-precipitated carbonates, has its smallest uncertainty at 27 °C, equal to  $\pm 2$  °C at the 95% confidence level. When using both biogenic and laboratory carbonates, the uncertainty minimum shifts to 20 °C, and is  $\pm 0.6$  °C at the 95% level. The revised calibration is consistent with most biogenic carbonates, including those growing at low temperatures, implying that the systematic deviations previously observed at low temperatures were due to uncertainty in the calibration, rather than kinetic fractionations at low temperatures. The revised calibration would significantly affect the results in studies of low temperature settings (10-15°C and below) and at temperatures above 40-50 °C.

In long term measurements of standard materials, we see no trend, suggesting that random variations in our mass spectrometer behavior and in sample preparation are responsible for most of the variance in measured replicates. This result may not hold for other laboratories, depending on equipment and procedures. Longer count times will help reduce errors due to shot noise. However, there is minimal improvement in precision for sample count time greater than about 900 s. At this point, the total error is dominated by non-Poisson sources, which can only be reduced by replicate analyses. Replicate analysis is thus critical in studies in which the sought after temperature signal is small.

#### Acknowledgements

Shikma Zaarur and Hagit Affek designed and executed the calibration measurements and Mark Brandon contributed the statistical analysis. We thank the editor and reviewers for their insights and helpful comments; Tobias Kluge, Casey Saenger, Peter Douglas, and Brian Goldstein for fruitful discussion; technical support of Gerard Olack, Dominic Colosi, Glendon Husinger of the Earth System Center for Stable Isotope Studies. The work was supported by the National Science Foundation Grant NSF-EAR-0842482 to Hagit Affek.

#### Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.epsl.2013.07.026.

#### References

- Affek, H.P., 2012. Clumped isotope paleothermometry: Principles, applications, and challenges. In: Ivany, L.C., Huber, B. (Eds.), Reconstructing Earth's Deep-Time Climate – The State of the Art in 2012. In: Paleontological Society Papers. Paleontological Society Papers, pp. 101–114.
- Affek, H.P., 2013. Clumped isotopic equilibrium and the rate of isotope exchange between CO<sub>2</sub> and water. Am. J. Sci. 313, 309–325.
- Affek, H.P., Bar-Matthews, M., Ayalon, A., Matthews, A., Eiler, J.M., 2008. Glacial/interglacial temperature variations in Soreq cave speleothems as recorded by 'clumped isotope' thermometry. Geochim. Cosmochim. Acta 72, 5351–5360.
- Affek, H.P., Eiler, J.M., 2006. Abundance of mass 47 CO<sub>2</sub> in urban air, car exhaust, and human breath. Geochim. Cosmochim. Acta 70, 1–12.
- Came, R.E., Eiler, J.M., Veizer, J., Azmy, K., Brand, U., Weidman, C.R., 2007. Coupling of surface temperatures and atmospheric CO<sub>2</sub> concentrations during the Palaeozoic era. Nature 449, 198–202.
- Csank, A.Z., Tripati, A.K., Patterson, W.P., Eagle, R.A., Rybczynski, N., Ballantyne, A.P., Eiler, J.M., 2011. Estimates of Arctic land surface temperatures during the early Pliocene from two novel proxies. Earth Planet. Sci. Lett. 304, 291–299.
- Dennis, K.J., Affek, H.P., Passey, B.H., Schrag, D.P., Eiler, J.M., 2011. Defining an absolute reference frame for 'clumped' isotope studies of CO<sub>2</sub>. Geochim. Cosmochim. Acta 75, 7117–7131.
- Dennis, K.J., Schrag, D.P., 2010. Clumped isotope thermometry of carbonatites as an indicator of diagenetic alteration. Geochim. Cosmochim. Acta 74, 4110–4122.
- Dietzel, M., Tang, J.W., Leis, A., Kohler, S.J., 2009. Oxygen isotopic fractionation during inorganic calcite precipitation – Effects of temperature, precipitation rate and pH. Chem. Geol. 268, 107–115.
- Eagle, R.A., Eiler, J.M., Tripati, A.K., Ries, J.B., Freitas, P.A., Hiebenthal, C., Wanamaker Jr., A.D., Taviani, M., Elliot, M., Marenssi, S., Nakamura, K., Ramirez, P., 2013. The influence of temperature and seawater carbonate saturation state on <sup>13</sup>C<sup>-18</sup>O bond ordering in bivalve m. Biogeosci. Discuss. 10, 157–194.
- Eagle, R.A., Schauble, E.A., Tripati, A.K., Tutken, T., Hulbert, R.C., Eiler, J.M., 2010. Body temperatures of modern and extinct vertebrates from <sup>13</sup>C-<sup>18</sup>O bond abundances in bioapatite. Proc. Natl. Acad. Sci. USA 107, 10377–10382.
- Eiler, J.M., 2007. "Clumped-isotope" geochemistry The study of naturally-occurring, multiply-substituted isotopologues. Earth Planet. Sci. Lett. 262, 309–327.
- Eiler, J.M., 2011. Paleoclimate reconstruction using carbonate clumped isotope thermometry. Quat. Sci. Rev. 30, 3575–3588.
- Eiler, J.M., Schauble, E., 2004. <sup>18</sup>O<sup>13</sup>C<sup>16</sup>O in Earth's atmosphere. Geochim. Cosmochim. Acta 68, 4767–4777.
- Finnegan, S., Bergmann, K., Eiler, J.M., Jones, D.S., Fike, D.A., Eisenman, I., Hughes, N.C., Tripati, A.K., Fischer, W.W., 2011. The magnitude and duration of Late Ordovician–Early Silurian glaciation. Science 331, 903–906.
- Friedman, I., O'Neil, J.R., 1977. Compilation of stable isotope fractionation factors of geological interest. In: Data of Geochemistry. Sixth ed. In: Geological Survey Professional Paper, vol. 440-KK.
- Gabitov, R.I., Watson, E.B., Sadekov, A., 2012. Oxygen isotope fractionation between calcite and fluid as a function of growth rate and temperature: An in situ study. Chem. Geol. 306, 92–102.
- Ghosh, P., Adkins, J., Affek, H., Balta, B., Guo, W.F., Schauble, E.A., Schrag, D., Eller, J.M., 2006a. <sup>13</sup>C<sup>18</sup>O bonds in carbonate minerals: A new kind of paleothermometer. Geochim. Cosmochim. Acta 70, 1439–1456.
- Ghosh, P., Eiler, J., Campana, S.E., Feeney, R.F., 2007. Calibration of the carbonate 'clumped isotope' paleothermometer for otoliths. Geochim. Cosmochim. Acta 71, 2736–2744.
- Ghosh, P., Garzione, C.N., Eiler, J.M., 2006b. Rapid uplift of the Altiplano revealed through <sup>13</sup>C<sup>18</sup>O bonds in paleosol carbonates. Science 311, 511–515.
- Grauel, A.-L., Schmid, T.W., Hu, B., Bergami, C., Capotondi, L., Zhou, L., Bernasconi, S.M., 2013. Calibration and application of the 'clumped isotope' thermometer to

foraminifera for high-resolution climate reconstructions. Geochim. Cosmochim. Acta 108, 125–140.

- Guo, W., Mosenfelder, J.L., Goddard III, W.A., Eiler, J.M., 2009. Isotopic fractionations associated with phosphoric acid digestion of carbonate minerals: Insights from first-principles theoretical modeling and clumped isotope measurements. Geochim. Cosmochim. Acta 73, 7203–7225.
- Henkes, G.A., Passey, B.H., Wanamaker Jr., A.D., Grossman, E.L., Ambrose Jr., W.G., Carroll, M.L., 2013. Carbonate clumped isotope compositions of modern marine mollusk and brachiopod shells. Geochim. Cosmochim. Acta 106, 307–325.
- Huntington, K.W., Eiler, J.M., Affek, H.P., Guo, W., Bonifacie, M., Yeung, L.Y., Thiagarajan, N., Passey, B., Tripati, A., Daeron, M., Came, R., 2009. Methods and limitations of 'clumped' CO<sub>2</sub> isotope  $\Delta_{47}$  analysis by gas-source isotope ratio mass spectrometry. J. Mass Spectrom. 44, 1318–1329.
- Huntington, K.W., Wernicke, B.P., Eiler, J.M., 2010. Influence of climate change and uplift on Colorado Plateau paleotemperatures from carbonate clumped isotope thermometry. Tectonics 29.
- Kim, S.T., Mucci, A., Taylor, B.E., 2007a. Phosphoric acid fractionation factors for calcite and aragonite between 25 and 75 °C: Revisited. Chem. Geol. 246, 135–146.
- Kim, S.T., O'Neil, J.R., 1997. Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. Geochim. Cosmochim. Acta 61, 3461–3475.
- Kim, S.T., O'Neil, J.R., Hillaire-Marcel, C., Mucci, A., 2007b. Oxygen isotope fractionation between synthetic aragonite and water: Influence of temperature and Mg<sup>2+</sup> concentration. Geochim. Cosmochim. Acta 71, 4704–4715.
- Kluge, T., Affek, H.P., 2012. Quantifying kinetic fractionation in Bunker Cave speleothems using  $\Delta_{47}$ . Quat. Sci. Rev. 49, 82–94.
- Passey, B.H., Levin, N.E., Cerling, T.E., Brown, F.H., Eiler, J.M., 2010. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. Proc. Natl. Acad. Sci. USA 107, 11245–11249.
- Peirce, B., 1852. Criterion for the rejection of doubtful observations. Astron. J. 45, 161–163.
- Peters, N.A., Huntington, K.W., Hoke, G.D., 2013. Hot or not? Impact of seasonally variable soil carbonate formation on paleotemperature and O-isotope records

from clumped isotope thermometry. Earth Planet. Sci. Lett. 361, 208-218.

- Quade, J., Eiler, J., Daeron, M., Achyuthan, H., 2013. The clumped isotope geothermometer in soil and paleosol carbonate. Geochim. Cosmochim. Acta 105, 92–107.
- Ross, S.M., 2003. Peirce's criterion for the elimination of suspect experimental data. J. Eng. Technol. 20, 38–41.
- Saenger, C., Affek, H.P., Felis, T., Thiagarajan, N., Lough, J.M., Holcomb, M., 2012. Carbonate clumped isotope variability in shallow water corals: Temperature dependence and growth-related vital effects. Geochim. Cosmochim. Acta 99, 224–242.
- Schauble, E.A., Ghosh, P., Eiler, J.M., 2006. Preferential formation of <sup>13</sup>C<sup>18</sup>O bonds in carbonate minerals, estimated using first-principles lattice dynamics. Geochim. Cosmochim. Acta 70, 2510–2529.
- Swanson, E.M., Wernicke, B.P., Eiler, J.M., Losh, S., 2012. Temperatures and fluids on faults based on carbonate clumped-isotope thermometry. Am. J. Sci. 312, 1–21.
- Thiagarajan, N., Adkins, J., Eiler, J., 2011. Carbonate clumped isotope thermometry of deep-sea corals and implications for vital effects. Geochim. Cosmochim. Acta 75, 4416–4425.
- Tripati, A.K., Eagle, R.A., Thiagarajan, N., Gagnon, A.C., Bauch, H., Halloran, P.R., Eiler, J.M., 2010. <sup>13</sup>C<sup>18</sup>O isotope signatures and 'clumped isotope' thermometry in foraminifera and coccoliths. Geochim. Cosmochim. Acta 74, 5697–5717.
- Wang, Z.G., Schauble, E.A., Eiler, J.M., 2004. Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. Geochim. Cosmochim. Acta 68, 4779–4797.
- Wray, J.L., Daniels, F., 1957. Precipitation of calcite and aragonite. J. Am. Chem. Soc. 79, 2031–2034.
- Zaarur, S., Olack, G., Affek, H.P., 2011. Paleo-environmental implication of clumped isotopes in land snail shells. Geochim. Cosmochim. Acta 75, 6859–6869.
- Zhou, G.-T., Zheng, Y.-F., 2001. Chemical synthesis of CaCO<sub>3</sub> minerals at low temperatures and implication for mechanism of polymorphic transition. J. Mineral. Geochem. 176, 323–343.

## **Online supplementary information**

## SI-1. Estimation of shot-noise error for $\Delta_{47}$ measurements

## Introduction

Huntington et al., (2009) discussed the role of shot noise as a source of error for clumpedisotope measurements, but they did not provide a specific estimate of the Poisson standard error, which represents the lower limit for precision of a  $\Delta_{47}$  measurement. Our analysis here is similar to that of Merritt and Hayes (1994), but it accounts for issues specific to clumped isotope measurements. The most significant problem is estimating the influence of shot noise on the measurement of the stochastic concentration  $R_{47}^*$ . The overall analysis below is cumbersome, but the result is a simple approximation for the Poisson standard error for  $\Delta_{47}$ . We also include a Matlab script, D47Calc.m, that illustrates the calculation of  $\Delta_{47}$  using the procedure outlined below.

The carbonate clumped-isotope thermometer is based on exchange reactions that produce anomalous concentrations of  $CO_2$  isotopologues in a carbonate mineral (Ghosh et al., 2006). The concentration anomaly is defined in terms of "per mil" units by

$$\Delta_i = 1000 \left(\frac{R_i}{R_i^*} - 1\right). \tag{1}$$

The mass ratio is defined by  $R_i = N_i/N_{44}$ , where  $N_i$  is the abundance (as measured in mass spectrometer counts) for CO<sub>2</sub> isotopologues with a cardinal mass number of *i*. The ratio is relative to  $N_{44}$ , the abundance of the mass 44 isotopologue, which is the most common CO<sub>2</sub> isotopologue.  $R_i^*$  represents the isotopologue mass ratios for the CO<sub>2</sub> sample with the same isotope concentration but randomly distributed between all isotopologues.

Carbonate minerals are digested in phosphoric acid and the analysis is done on the resulting CO<sub>2</sub> gas (see Methods section 2.2. in paper). The gas is analyzed for  $N_{44}$ ,  $N_{45}$ ,  $N_{46}$ , and  $N_{47}$  counts. Measurements on the mass spectrometer are always relative to a reference gas, so

$$R_{i,SA} = R_{i,RG} \frac{N_{i,SA}}{N_{44,SA}} / \frac{N_{i,RG}}{N_{44,RG}},$$
(2)

where the subscripts "SA" and "RG" indicate counts measured for the sample  $CO_2$  gas and the reference  $CO_2$  gas, respectively. For our current protocol, a single sample is measured with 90 "20 s" cycles. Each cycle involves 20 s of count time on the sample gas, followed by 20 s on the reference gas. Mass numbers 44-49 are measured simultaneously during each count interval.

The variable  $R_{i,RG}$  in (2) represents the mass ratios of the reference gas relative to isotopic standards. The C and O isotopic composition of the reference gas is known. For example, the reference gas used in our study here has  ${}^{13}\delta C = -3.58\%$  and  ${}^{18}\delta O = 25.00\%$ relative to VPDB and VSMOW, respectively. The  $\Delta_i$  values, however, are usually not known. The mass ratios  $R_{45,RG}$ ,  $R_{46,RG}$ , and  $R_{47,RG}$  are estimated by assuming that reference gas has a stochastic concentration, where  $\Delta_i = 0$ . This assumption is rectified by the heated-gas correction discussed below. For a stochastic distribution, the relationships between the C and O stable isotopic composition and the CO<sub>2</sub> isotopologue composition are given by (Affek and Eiler, 2006),

$$R_{45}^* = R_{13} + 2R_{17}, (3a)$$

$$R_{46}^* = 2R_{18} + 2R_{13}R_{17} + R_{17}^2, \text{ and}$$
(3b)

$$R_{47}^* = 2R_{13}R_{18} + 2R_{17}R_{18} + R_{13}R_{17}^2, (3c)$$

where the asterisk indicates a stochastic concentration. The variables  $R_{13}$ ,  $R_{17}$ , and  $R_{18}$  are mass ratios for C and O isotopic concentrations, which are calculated by

$$R_{13} = R_{13,\text{VPDB}} \left( \delta^{13} \text{C} / 1000 + 1 \right), \tag{4a}$$

$$R_{18} = R_{18,\text{VSMOW}} (\delta^{18} \text{O}/1000 + 1), \text{ and}$$
 (4b)

$$R_{17} = K R_{18}^{\lambda} , \qquad (4c)$$

where  $R_{13,VPDB} = 0.0112372$ ,  $R_{17,VSMOW} = 0.00037995$ ,  $R_{18,VSMOW} = 0.0020052$ ,  $K = \frac{R_{17,VSMOW}}{R_{18,VSMOW}^{\lambda}}$ , and

 $\lambda = 0.5164$  (Gonfiantini et al., 1995). Note that (4c) is an empirical relationship that is calibrated for mass-dependent fractionation in terrestrial environments (Brand et al., 2010).

The carbonate anion isotopologue  ${}^{13}C^{18}O_2^{16}O_2^{2-}$  is multiply substituted in that it has two rare isotopes:  ${}^{13}C$ , and  ${}^{18}O$ . It accounts for ~96% of the mass 47 concentration measured in the CO<sub>2</sub> gas generated from a carbonate sample (Schauble et al., 2006). The isotope exchange reaction that produces this anion,

$${}^{12}\mathrm{C}{}^{18}\mathrm{O}{}^{16}\mathrm{O}{}^{2-}_2 + {}^{13}\mathrm{C}{}^{16}\mathrm{O}{}^{2-}_3 \Leftrightarrow {}^{13}\mathrm{C}{}^{18}\mathrm{O}{}^{16}\mathrm{O}{}^{2-}_2 + {}^{12}\mathrm{C}{}^{16}\mathrm{O}{}^{2-}_3,$$

has become the model for the carbonate clumped-isotope thermometer. If this were the only exchange reaction, then the  $\Delta$  value for the reaction would be

$$\Delta_{[3866]} = 1000 \left[ \left( \frac{R_{47}}{R_{47}^*} - 1 \right) - \left( \frac{R_{45}}{R_{45}^*} - 1 \right) - \left( \frac{R_{46}}{R_{46}^*} - 1 \right) \right], \tag{5}$$

where the subscript [3866] is a shorthand notation for this exchange reaction (Schauble et al., 2006).

There are two problems that limit the use of (5). The first is that there are other exchange reactions that contribute to the  $R_{45}$ ,  $R_{46}$  and  $R_{47}$  concentrations. The current generation of "clumped-isotopes" mass spectrometers cannot resolve the specific CO<sub>2</sub> isotopologues for these reactions (this situation is likely to change in the future as discussed by Eiler et al., 2013).

The second and more fundamental problem is that the current generation of mass spectrometers cannot make independent measurements of  $R_{45}$ ,  $R_{46}$  and  $R_{45}^*$ ,  $R_{46}^*$ , as needed to calculate  $\Delta_{45}$  and  $\Delta_{46}$ . In fact, the measured values for  $R_{45}$  and  $R_{46}$  are used to estimate  $\delta^{13}$ C and  $\delta^{18}$ O, as required for estimation of  $R_{47}^*$  (see below). In conclusion, the carbonate clumped-isotope thermometer is guided by theory, which predicts that  $\Delta_{47}$  should have a strong linear relationship with  $T^{-2}$ , but the calibration itself is entirely empirical at present. The fit of the calibration data to the linear  $T^{-2}$  equation provides a useful test of this predicted relationship.

Our statistical analysis requires a clear definition of the relationship between the massspectrometer measurements and the estimated  $\Delta_{47}$  value,

$$\Delta_{47,SA/ABS} = a_0 + a_1 \Big( \Delta'_{47,SA/RG} - c_0 - c_1 \delta_{47,SA/RG} \Big), \tag{6a}$$

where

$$\Delta'_{47,SA/RG} = 1000 \left( \frac{R_{47,SA}}{f(R_{45,SA}^*, R_{46,SA}^*)} - 1 \right)$$
and (6b)

$$\delta_{47,SA/RG} = 1000 \left( \frac{R_{47,SA}}{R_{47,RG}^*} - 1 \right).$$
(6c)

The prime in  $\Delta'$  indicates a raw delta value that has yet to be corrected for mass spectrometry effects. In (6c),  $\delta_{47,SA/RG}$  indicates the isotopic composition of the sample CO<sub>2</sub> gas relative to that for the stochastic concentration of the reference CO<sub>2</sub> gas. Affek and Eiler (2006) proposed defining  $\delta_{47}$  relative to the international standards VPDB and VSMOW, but the RG reference, as shown in (6c), has become the standard practice (Huntington et al., 2009; Dennis et al., 2011), which is appropriate given that  $\delta_{47}$  is only used for internal standardization. In equation (6b), the stochastic ratio,  $R_{47}^*$ , is represented by the function  $R_{47}^* = f(R_{45}^*, R_{46}^*)$  to highlight that this calculation is dependent on measured values for  $R_{45,SA}$  and  $R_{46,SA}$ , as discussed below. The subscript "ABS" refers to the absolute reference frame of Dennis et al. (2011). The variables  $c_0$ ,  $c_1$ ,  $a_0$ , and  $a_1$  are laboratory-specific constants that correct for mass-spectrometer effects, as described below.

### **Compositional Correction using the Heated-Gas Line**

The heated-gas line (HGL) is a linear function that corrects for the influence of isotopic composition on the measured  $\Delta'_{47,SA/RG}$ . This effect is estimated using a suite of CO<sub>2</sub>-gas samples that vary in isotopic composition but otherwise are equilibrated to  $\Delta_{47} = 0$  as a result of heating at 1000°C for >2 hours (Huntington, et al., 2009). The parameters  $c_0$  and  $c_1$  are determined by least squares using the model equation

$$\Delta'_{47,HG/RG} = c_0 + c_1 \delta_{47,HG/RG}, \tag{7}$$

where the subscript "HG" indicates measurements from the heated-gas samples.

The HGL correction is given by

$$\Delta'_{47,SA/ZV} = \Delta'_{47,SA/RG} - \left(c_0 + c_1 \delta_{47,SA/RG}\right),\tag{8}$$

where  $\Delta'_{47,SA/RG}$  and  $\delta_{47,SA/RG}$  are sample measurements. The subscript "ZV" indicates that the measurement, as now represented by  $\Delta'_{47,SA/ZV}$ , is now referenced to a "zero-value" CO<sub>2</sub> gas that has  $\Delta_{47} = 0$  but otherwise has the same isotopic composition as the sample. Note that the ZV reference was the first standardization used in carbonate clumped-isotope thermometry (e.g., Ghosh et al., 2006), but it is now recognized as incomplete (Dennis et al., 2011). For this reason, we include a prime with the delta symbol. The least-squares estimate for the heated-gas line is typically based on 10 to 30 measurements and the quality of the fit is usually excellent ( $R^2 > 0.95$ ). The errors for this correction are therefore small, and can be safely ignored in our error analysis for  $\Delta_{47}$ .

The first clumped-isotope measurements at Yale included a correction to account for scaling differences between mass spectrometers at Caltech and at Yale. This correction is equivalent to multiplying the right-hand side of (8) by  $-0.87/c_0$  This correction is now included in the absolute reference of Dennis et al., (2011) (see below). Nonetheless, studies that pre-date Dennis et al. (2011) may have used this older standardization scheme.

## Absolute reference frame using equilibrated gases

Dennis et al., (2011) define an additional correction that ensures that inter-laboratory measurements of  $\Delta_{47}$  are tied to a common "absolute" reference. The correction is defined on a lab-by-lab basis using a set of CO<sub>2</sub> gases equilibrated at different temperatures and then measured and corrected to the ZV reference using equation (8). The parameters  $a_0$  and  $a_1$  for the correction are determined by least squares using the model equation,

$$\Delta_{47,\text{EG/ABS}} = a_0 + a_1 \Delta'_{47,\text{EG/ZV}}, \tag{9}$$

where the subscript "EG" refers to measurements from the equilibrated-gas samples. This equation relates  $\Delta'_{47 \text{ EG/ZV}}$ , which are the measured values for the equilibrated gases relative to the

ZV reference, and  $\Delta_{47,EG/ABS}$ , which are the predicted values for the equilibrated gases. The predicted value is determined using the lab-controlled equilibration temperatures and the theoretical equilibrium relationship by Wang et al., (2004) for  $\Delta_{47}$  of CO<sub>2</sub> gas as a function of temperature. The fit for equation (9) is usually excellent ( $R^2 > 0.95$ ). As a result, the errors associated with this correction are very small and can be safely ignored in our error analysis for  $\Delta_{47}$ .

One might think that the HGL and ABS corrections could be combined into a single correction. The reason that this is not done is that the HGL calibration is done frequently to account for daily variations in the mass spectrometer, whereas the ABS calibration is done infrequently, usually once or twice a year. As such, it is more convenient to apply these corrections in two steps.

## Stochastic Concentration for R<sub>47</sub>

In equation (6), the calculated value for  $R_{47,SA}^*$  is represented by  $f(R_{45,SA}^*, R_{46,SA}^*)$ . The numerical calculation of this function is well known (e.g., Brand, 2010), but the details are important for clumped isotopes, so they are outlined here.

The first step is to convert  $R_{45}^*$  and  $R_{46}^*$  into  $R_{13}$ ,  $R_{17}$  and  $R_{18}$  for the C and O isotopes. As noted above,  $R_{45}^*$  and  $R_{46}^*$  are approximated by measured values for  $R_{45}$  and  $R_{46}$ , which implies that  $\Delta_{45} = \Delta_{46} = 0$ . This assumption is supported by Wang et al. (2004, Table 4), who show that  $\Delta_{45}$  and  $\Delta_{46}$  are typically very small, by two orders of magnitude relative to  $\Delta_{47}$ . Equations (3a, 3b, and 4c) are combined to give an equation for  $R_{18}$  as a function of  $R_{45}^*$  and  $R_{46}^*$  (Brand, 2010),

$$-3K^{2}R_{18}^{2\lambda} + 2KR_{45}^{*}R_{18}^{\lambda} + 2R_{18} - R_{46}^{*} = 0.$$
<sup>(10)</sup>

 $R_{18}$  is calculated using a root-finding algorithm, such as the Solve function in Excel or fzero in Matlab. The root-finding search should start with an initial guess of  $R_{18} = 0$ . For the search, (10) is modified to: real $\left(-3K^2R_{18}^{2\lambda} + 2KR_{45}^*R_{18}^{\lambda} + 2R_{18} - R_{46}^*\right)$ , where the function real(·) returns the

real value of (10). This function is called "IMREAL" in Excel and "real" in Matlab. This step is necessary to avoid errors with complex values should the algorithm choose to search  $R_{18}$  values less than zero.

Once a solution is found for  $R_{18}$ , the remaining values  $R_{13}$ ,  $R_{17}$ , and  $R_{47}^*$  are determined using (3c, 4a, and 4c). Plugging the values for  $R_{13}$ ,  $R_{17}$ , and  $R_{18}$  back into (3a,b) will confirm the assumption that  $R_{45}^* = R_{45}$  and  $R_{46}^* = R_{46}$ . In other words, the available measurements do not allow for estimation of the second and third terms in (5).

## **Predicted Shot-Noise Error**

Estimates of  $\Delta_{47}$  are related to 8 independent mass-spectrometer measurements: four count variables for the sample gas,  $N_{i,SA}$ , and four more for the reference gas,  $N_{i,RG}$ , where the subscript, i = 44, 45, 46, and 47, indicates the cardinal mass of the ions recorded by that count variable. The count rate  $\rho_i$  for the *i*th mass is given by

$$\rho_i = \frac{N_i}{t} = 6.24 \times 10^{18} \frac{E_i}{\Omega_i},\tag{11}$$

where  $E_i$  is the measured voltage on the detector,  $\Omega_i$  is the amplifier resistance in ohms, and *t* is the count time in seconds. The count rate is proportional to the current *I* (amperes) generated by the counts, where  $I = E/\Omega$  (Ohm's Law). The constant 6.24 x 10<sup>18</sup> defines the number of electrons per second in an ampere.

Mass	$E_i$ (volt)	$\Omega_{i}$ (ohm)	$\overline{ ho}_i(\mathrm{s})$
44	~16	$3 \times 10^8$	$3.328 \times 10^{11}$
45	~19	$3 \ge 10^{10}$	3.952 x 10 <sup>9</sup>
46	~22	$1 \ge 10^{11}$	1.373 x 10 <sup>9</sup>
47	~2.5	$1 \ge 10^{12}$	$1.560 \ge 10^7$

**Table S1. Mass-Spectrometer Parameters** 

Table S1 summarizes the parameters for the Yale mass spectrometer. The natural concentrations of CO<sub>2</sub> isotopologues have small ranges, which translates into small variations in count rates, generally less than 5%. As a result, it is common to use average count rates, shown as  $\bar{\rho}_i$  in Table S1, when estimating measurement errors (Merritt and Hayes, 1994).

In the absence of other sources of variation, the count variables will be Poisson distributed. This source of error is also called shot noise. The Poisson distribution has the unusual property that the expectation and variance for the distribution are equal, so that  $SE^2(N_i) = N_i$ .

The "propagation of errors" formula (Taylor, 1997), 
$$SE^2(y) \approx \sum \left(\frac{\partial y}{\partial x_i}SE(x_i)\right)^2$$
, is a

useful approximation for estimating how the error for an estimate of y relates to the errors for the measured values used to calculate y. Huntington et al. (2009) called attention to correlation between measured variables, but for the problem here, the count variables should have independent errors, so the correlation terms in the propagation formula are ignored.

Applying the propagation formula to (2) gives an approximation for the Poisson standard error of  $R_{i_i}$ 

$$\frac{SE_{p}^{2}(R_{i})}{R_{i}^{2}} \approx \frac{SE_{0}^{2}(R_{i})}{tR_{i}^{2}} \approx \frac{1}{N_{i,SA}} + \frac{1}{N_{44,SA}} + \frac{1}{N_{i,RG}} + \frac{1}{N_{44,RG}} \approx \frac{2}{t} \left(\frac{1}{\overline{\rho}_{i}} + \frac{1}{\overline{\rho}_{44}}\right), \quad (12)$$

where  $\overline{\rho}_{44}$  and  $\overline{\rho}_i$  are the average count rates (Table S1) and the subscript "p" indicates Poisson errors.  $SE_0$  represents the Poisson standard error for a one-second count interval. Note that  $\overline{\rho}_{45}$ ,  $\overline{\rho}_{46}$ , and  $\overline{\rho}_{47}$  are at least two orders of magnitude smaller than  $\overline{\rho}_{44}$ , which means that the contribution of  $1/\overline{\rho}_{44}$  can be safely ignored in (12). The result is a useful and accurate approximation (Merritt and Hayes, 1994):  $SE_0(R_i) \approx R_i \sqrt{2/\overline{\rho}_i}$ .

To simplify the propagation calculation, we recast (6) as a single equation in terms of mass ratios,

$$\Delta_{47,SA/ABS} = a_0 + a_1 \left( 1000 \ln\left(\frac{R_{47,SA}}{R_{47,SA}^*}\right) - c_0 - c_1 1000 \ln\left(\frac{R_{47,SA}}{R_{47,RG}}\right) \right).$$
(13)

The logarithmic formulation used in (13) is based on  $\Delta_{47} \approx 1000 \ln \left(\frac{R_{47}}{R_{47}^*}\right)$ , and

 $\delta_{_{47,SA/RG}} = 1000 \ln \left( \frac{R_{_{47,SA}}}{R_{_{47,RG}}} \right)$ , which are accurate approximations for these cases. Simplification of

(13) gives

$$\Delta_{47,SA/ABS} \approx a_0 - a_1 c_0 + 1000 a_1 \Big[ (1 - c_1) \ln(R_{47,SA}) - \ln(R_{47,SA}^*) + c_1 \ln(R_{47,RG}) \Big]$$
(14)

The identity  $\frac{d \ln x}{dx} = \frac{1}{x}$  is used to derive the partial derivatives for the propagation formula,

$$\frac{\partial \Delta_{47,SA/ABS}}{\partial R_{47,SA}} = 1000 \frac{a_1(1-c_1)}{R_{47,SA}},$$
(15a)

$$\frac{\partial \Delta_{47,SA/ABS}}{\partial R^*_{45,SA}} = 1000 \frac{-a_1}{R^*_{45,SA}} \frac{\partial \ln R^*_{47,SA}}{\partial \ln R^*_{45,SA}}, \text{ and}$$
(15b)

$$\frac{\partial \Delta_{47,SA/ABS}}{\partial R_{46,SA}^*} = 1000 \frac{-a_1}{R_{46,SA}^*} \frac{\partial \ln R_{47,SA}^*}{\partial \ln R_{46,SA}^*}.$$
 (15c)

Substitution of these derivatives into the "propagation of errors" formula gives an approximation for the Poisson standard error of  $\Delta_{\rm 47,SA/ABS}$ ,

$$SE_{p}^{2}(\Delta_{47,SA/ABS}) \approx \left(1000 a_{1}(1-c_{1}) \frac{SE_{p}(R_{47,SA})}{R_{47,SA}}\right)^{2} + \left(1000 a_{1}\left[\frac{\partial \ln R_{47,SA}^{*}}{\partial \ln R_{45,SA}^{*}}\right] \frac{SE_{p}(R_{45,SA}^{*})}{R_{45,SA}^{*}}\right)^{2} + \left(1000 a_{1}\left[\frac{\partial \ln R_{47,SA}^{*}}{\partial \ln R_{46,SA}^{*}}\right] \frac{SE_{p}(R_{46,SA}^{*})}{R_{46,SA}^{*}}\right)^{2}.$$
(16)

These equations can be simplified by using an approximation for the log derivatives (shown in square brackets in equation 16). The numerical algorithm above for  $f(R_{45}^*, R_{46}^*)$  was extended to calculate values for these derivatives for a range of compositions,  $\delta^{13}C_{VPDB} = -50$  to +10%, and  $\delta^{18}O_{VSMOW} = 0$  to +50%, which is representative of the natural range for these isotopes. Figure S1 shows that these derivatives are fairly constant over this compositional range.

Their average values are  $\left\langle \frac{\partial \ln R_{47}^*}{\partial \ln R_{45}^*} \right\rangle = 1.03161$  and  $\left\langle \frac{\partial \ln R_{47}^*}{\partial \ln R_{46}^*} \right\rangle = 0.98361$ . This calculation is

based on a reference gas with  $\delta^{13}C_{VPDB} = \delta^{18}O_{VSMOW} = 0\%$ . A working reference gas will generally have a different composition, but this difference would merely offset the data points parallel to the  $R_{47}^*$  axis of the plot. Thus, our conclusion is not affected by this detail.





By combining (11, 16) with the average values for the log derivatives, we get a simpler equation for the Poisson standard error for  $\Delta_{47,SA/ABS}$ ,

$$SE_{p}^{2}(\Delta_{47,SA/ABS}) \approx \frac{2 \times 10^{6} a_{1}^{2}}{t} \left( \left(1 - c_{1}\right)^{2} 6.41055 \times 10^{-8} + 2.72 \times 10^{-10} + 7.08 \times 10^{-10} \right).$$
(17)

The parameter  $c_1$  is always very small relative to one. Thus, the first term in (16, 17) dominates, and the other terms can be ignored.

Thus, the final version of the approximation for the Poisson standard error equation is

$$SE_{p}\left(\Delta_{47,\text{SA/ABS}}\right) \approx \frac{SE_{0}\left(\Delta_{47,\text{SA/ABS}}\right)}{\sqrt{t}} \approx \frac{a_{1}\left(1-c_{1}\right)}{\sqrt{t}} 0.358\% \,.$$

$$\tag{18}$$

For an example, consider our study, which uses  $a_1 = 1.0584$  and  $c_1 = 0.0041$  (this value varied between 0.0063 and 0.0030 over the measurements in this studies) and has little effect on  $SE_{0(\Delta 47,SA/ZV)}$ . For the ABS reference,  $SE_0(\Delta_{47,SA/ABS}) \approx 0.377\%$ . For the ZV reference, we set  $a_1$ = 1, which gives  $SE_0(\Delta'_{47,SA/ZV}) \approx 0.357\%$ .

## SI-2. Clumped-isotope thermometer calibration line calculation

We seek a functional relationship between  $\Delta_{47}$  and temperature *T*. Theoretical and empirical studies indicate that at low temperatures, like those used in our experiments, the equilibrium isotopic fractionation is well approximated by a linear function of  $T^{-2}$ . The model function for our calibration is

$$\Delta_{47} = b_0 + b_1 (1000/T)^2 + e \tag{19}$$

where  $b_0$  and  $b_1$  are parameters, *e* is the error in the measurement of  $\Delta_{47}$ , and *T* is in Kelvin. This relationship is tested and calibrated using measurements of  $\Delta_{47,i}$  made at prescribed temperatures  $T_i$ , where i = 1 to *n*. The entries for  $\Delta_{47,i}$  may be individual measurements or the mean values when more than one replicate is measured. The variable  $m_i$  represents the number of replicates for the *i*th measurement.

In our study, we use "sample" to refer to material to be analyzed, and "replicate" to indicate a measurement based on an aliquot of the sample. A "sample measurement" refers to the average value for all replicate measurements. A "known sample" refers to a carbonate that has been formed at a known temperature, and an "unknown sample", to a carbonate with an unknown temperature. The objective is to use  $\Delta_{47}$  to predict the temperature of carbonate precipitation for an unknown sample. Note that all uncertainties are cited at ±1 standard error (SE), unless noted otherwise.

The least-squares method is used to solve for  $\hat{b}_0$ ,  $\hat{b}_1$ , and  $\hat{e}_i$ , which are sample-based estimates for the parameters in (18). Note that  $\hat{e}_i$  are commonly called the residuals. The caret is used to distinguish between the population-based parameter (no caret) and its sample-based estimate (caret). Note that we include some familiar exceptions, such as  $\sigma$  and *s* for the standard deviation, where the older statistics convention is used, with the population and sample variables indicated by Greek and English letters, respectively.

There has been a long and complicated debate about how to treat least-squares problems when both variables have errors (Carroll et al., 2006; Buonaccorsi, 2009). A popular approach is the York method, which provides an explicit account of errors for both variables (York et al., 2004), but that method is difficult to extend to the problem considered here. Our calibration experiment is designed around a set of samples prepared at specified temperatures. This design corresponds to the Berkson error model (Berkson, 1950; Buonaccorsi, 2009), which means that  $\Delta_{47}$  should be treated as the dependent (or response) variable and  $(1000/T)^2$  as the independent (or predictor) variable. This conclusion is supported by another argument, perhaps more compelling, based on reliability ratios (Carroll et al., 2006). The reliability ratios for  $\Delta_{47}$  and  $(1000/T)^2$  are

$$\lambda_{\Delta 47} = \frac{SD^2(\Delta_{47})}{SE^2(\Delta_{47}) + SD^2(\Delta_{47})} \text{ and } \lambda_{T^{-2}} = \frac{SD^2((1000/T)^2)}{SE^2((1000/T)^2) + SD^2((1000/T)^2)},$$

respectively, where *SD* indicates the spread of the data (as represented by its standard deviation), *SE*, the size of the errors (as represented by the standard error). A value of  $\lambda$  close to 1 would indicate a variable that is insensitive to its error. For  $\Delta_{47}$ ,  $SE(\Delta_{47,i}) \approx 0.0188\%$  and  $SD(\Delta_{47,i}) \approx$ 0.0772‰, which gives  $\lambda_{\Delta 47} \approx 0.94$ .

The standard error for temperature is estimated to be ~0.5 K. Propagation of errors indicates that  $SE((1000/T)^2) \approx \left| \frac{\partial (1000/T)^2}{\partial T} \right| SE(T) = 2 \times 10^6 T^{-3} SE(T)$ . Thus, at the centroid of the data, located at 300K and 0.636‰, we estimate that  $SE((1000/T)^2) \approx 0.037$ . For  $(1000/T)^2$ ,  $SD((1000/T)^2) \approx 1.424$  and  $\lambda_{T^{-2}} \approx 0.99$ . These arguments support our decision, as shown in (18) to designate  $(1000/T)^2$  as the independent variable. The errors  $e_i$  in our least-squares analysis (19) are attributed solely to the  $\Delta_{47,i}$  measurements.

The calibration method used here is called "inverse regression" (Draper and Smith, 1998), where the model equation (19) is used in reverse to make predictions of T from measurements of  $\Delta_{47}$ . The prediction equation is

$$\hat{T}_0 = 1000 \left(\frac{\Delta_{47,0} - \hat{b}_0}{\hat{b}_1}\right)^{-1/2} , \qquad (20)$$

where  $\Delta_{47,0}$  is the measurement for a sample with an unknown temperature, and  $\hat{T}_0$  is the predicted temperature in Kelvin for the sample. Lavagnini and Magno (2006) provide an excellent review of this topic as applied to mass-spectrometry data.

We use weighted least squares to allow for the more general case where a data set might have heterogeneous analytical errors. There is usually little difference in parameter estimates produced by unweighted and weighted least squares. The important advantage of weighted least squares is that it produces more precise uncertainty estimates (Lavagnini and Magno, 2006). Our primary calibration data set is heterogeneous, in that it includes 7 data points from Ghosh et al., (2006) with sample count times ranging from 400 to 880 s, and 21 new data points, measured at Yale with sample count times of 1800 s. A weighted analysis is needed to merge these different data.

The weighted least squares method requires that the  $\Delta_{47,i}$  measurements include estimates of the standard errors  $SE(\Delta_{47,i})$ . In S1, we estimated the Poisson shot-noise, which is the limiting case for precision, in that  $SE(\Delta_{47,i}) \ge SE_p(\Delta_{47,i})$ . The measurements are likely influenced by other errors, such as sample preparation or non-Poisson variations in the mass spectrometer. We assume that these errors are random and can be characterized as  $SE_r(\Delta_{47,i})$ , which is the non-Poisson standard error for a single replicate measurement. Non-Poisson errors are independent of count time and can be reduced by replicate measurements.

These concepts indicate the following equation for the standard error of a  $\Delta_{47,i}$  measurement,

$$SE^{2}\left(\Delta_{47,i}\right) \approx \frac{SE_{0}^{2}\left(\Delta_{47,i}\right)}{t_{i}} + \frac{SE_{r}^{2}\left(\Delta_{47,i}\right)}{m_{i}},$$
(21)

where  $m_i$  is the number of replicate measurements averaged for the sample measurement  $\Delta_{47,i}$ , and  $t_i$  is the total sample count time for all replicate measurements combined.  $SE_0(\Delta_{47,i})$  is defined above (18) as the Poisson standard error for a measurement with a one-second count interval. For our analysis here, we use  $SE_0(\Delta'_{47,SA/ZV}) \approx 0.357\%$ , as estimated for the ZV reference. The non-Poisson standard error  $SE_r(\Delta'_{47,SA/ZV})$  is estimated using the residuals determined by the calibration estimate (see section 4.2.3 in the main text).

# Weighted Least-Squares

The weights for the least-squares calculation are defined by

$$W_{i} = \frac{n SE(\Delta_{47,i})^{-2}}{\sum SE(\Delta_{47,i})^{-2}} = \frac{SE(\Delta_{47,i})^{-2}}{\sigma_{e}^{-2}},$$
(22)

where the weights are normalized so that  $\sum W_i = n$ .

The variables used for the weighted least-squares solution are:

$$\overline{x} = n^{-1} \sum W_i (1000/T_i)^2$$
(23a)

$$\overline{y} = n^{-1} \sum W_i \Delta_{47,i} \tag{23b}$$

$$S_{xx} = \sum W_i \left( \left( 1000/T_i \right)^2 - \overline{x} \right)^2$$
(23c)

$$S_{yy} = \sum W_i \left( \Delta_{47,i} - \overline{y} \right)^2 \tag{23d}$$

$$S_{xy} = \sum W_i \Big( (1000/T_i)^2 - \overline{x} \Big) \Big( \Delta_{47,i} - \overline{y} \Big)$$
(23e)

$$\sigma_e^2 = \left(\frac{\sum SE(\Delta_{47,i})^{-2}}{n}\right)^{-1}$$
(23f)

The estimates for the parameters are

$$\hat{b}_0 = \overline{y} - \frac{S_{xy}}{S_{xx}} \overline{x}$$
, and (24a)

$$\hat{b}_1 = \frac{S_{xy}}{S_{xx}}$$
 (24b)

If our assumption is correct, that there are no other significant sources of error other than the specified  $SE(\Delta_{47,i})$ , then  $\sigma_e$  can be used to estimate uncertainties for the parameters and for temperatures predicted using the calibration equation.

$$SE(b_0)^2 = \sigma_e^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)$$
(25a)

$$SE(b_1)^2 = \frac{\sigma_e^2}{S_{xx}}$$
(25b)

For a given  $(1000/T_i)^2$ , the estimator for  $\Delta_{47}$  is

$$\widehat{\Delta_{47}} = \hat{b}_0 + \hat{b}_1 (1000/T_i)^2$$
(26a)

The standard error for  $\Delta_{47}$  is given by

$$SE\left(\widehat{\Delta_{47}}\right)^2 = \sigma_e^2 \left(\frac{1}{n} + \frac{\left(x - \overline{x}\right)^2}{S_{xx}}\right)$$
(26b)

# Assessing the Quality of the Least-Squares Fit

 $R^2$  Statistic. The quality of the fit is judged using the  $R^2$  statistic, the  $\chi^2$  test, and the Durbin-Watson test. The  $R^2$  statistic is estimated using the conventional  $R^2$  formula,

$$R^{2} = 1 - \frac{\sum \hat{e}_{i}^{2}}{\sum \Delta_{47,i}^{2} - (1/n) \left(\sum \Delta_{47,i}\right)^{2}}$$
(27)

where  $\hat{e}_i$  are the unweighted residuals for the best-fit solution,

$$\hat{e}_i = \Delta_{47,i} - \hat{b}_0 - \hat{b}_1 (1000/T_i)^2 .$$
<sup>(28)</sup>

Willett and Singer (1988) recommended this approach in their review of using  $R^2$  for weight least squares.

**Chi-Squared Test.** The  $\chi^2$  test is used to check one of our assumptions, that  $E(\hat{e}_i^2) = SE^2(\Delta_{47,i})$ , which means that the expectation for the residuals are equal to their associated standard errors. The sample statistic for the  $\chi^2$  test is

$$X^{2} = \sum \frac{\hat{e}_{i}^{2}}{SE(\Delta_{47,i})^{2}} = \sum \frac{W_{i}\hat{e}_{i}^{2}}{\sigma_{e}^{2}} .$$
<sup>(29)</sup>

If our assumption is correct, then  $T^{-2}$  should be  $\chi^2$  distributed with n-2 degrees of freedom. The test is based on  $P(\chi^2 > X^2)$ , which is the probability for the  $\chi^2$  distribution to have a value greater than the observed  $X^2$ . The expected value is  $E(X^2) = n-2$ , which corresponds to the middle of the  $\chi^2$  distribution (e.g.,  $P(\chi^2) \approx 50\%$ ). A small  $\chi^2$  probability (e.g., less than ~5%) indicates that the residuals are significantly larger than the specified standard errors. A large  $\chi^2$  probability (e.g., greater than ~95%) indicates residuals are significantly smaller than the specified standard errors. It is useful to define "reduced chi-square"  $X_R^2$ , which is equal to  $X^2$  normalized by its expected value.

**D47Calibration Program**. Our calibration estimates were calculated using the Matlab script, D47Calibration.m, which is included as a file in the supplementary materials. That program uses an Excel spreadsheet as the source for the input data. The input file for our study is called "Table for Calibration Calculation.xlsx". The program reads as input all of the values within the second to last row of this file, with each row corresponding to a single sample. The first row is assumed to have header information, and is thus ignored. Column A contains sample names as represented by ASCII character strings. Column B contains the known temperatures in Celsius. Column C is ignored but is used to show the transformed temperatures  $(1000/T(K))^2$ . Column D contains the measured  $\Delta_{47}$  (‰) for each sample. Column E contains the number of replicates used to produce (by averaging) the sample  $\Delta_{47}$  value. Column F is used to report the total sample count time for

all replicates combined. This column is ignored by the program but is useful for calculating the next column. Column G contains the estimated Poisson standard error, which is calculated using  $SE_p(\Delta_{47}) = SE_0(\Delta_{47})/\sqrt{t}$ , where  $SE_0(\Delta_{47})$  is determined from (18) and t is the total sample count time from column F. Column H contains ASCII strings indicating the group name for each sample. Column I contains ASCII strings that control the shape and color of the plot symbol for each sample (see Matlab for details). Column J contains an integer flag indicating the status for each sample during a calibration calculation: 0 = sample ignored, 1 = sample used for calibration estimate, and 2 = sample used for comparison to best-fit calibration (as discussed below). Note that the input data can be entered at the replicate level, with a replicate measurement in each row, or at a sample level, with each row consisting of an average of related replicate measurements. The number of replicate measurements, m, as reported in column E, is used to account for these different strategies. The final estimates should be the same.

The D47Calibration program has additional settings, located towards the top of the program. The variable excelDataFile is set to an ASCII string representing the filename for the Excel input file. The variable fitOption has settings: 1 = weighted least-squares estimate of the calibration line and the non-Poisson standard error  $SE_r(\Delta_{47})$ , 2 = weighted least-squares estimate of the calibration line using the Poisson standard errors for weights  $(SE_r(\Delta_{47})=0)$ , and 3 = conventional least-squares estimate of the calibration line (i.e., no weighting). The variable SED47Nominal is set to a typical value for the total standard error for  $\Delta_{47}$  for an unknown sample. This value is used for the plotted figures to define the confidence curves for the predicted temperature for an unknown sample. The variable gamma determines the probability interval bounded by the confidence curves in those figures. For example, gamma = 0.95 would give confidence curves that bounded the 95% confidence interval around the predicted temperature.

As an example, we describe the calibration estimate for the combination of known synthetic carbonate samples from Yale and Ghosh et al. (2006). We start with the weighted least-squares fit using Poisson standard errors alone (fitOption = 2), which gives a best-fit result with  $X_R^2 = 3.99$  and  $P(\chi^2 > X^2) \approx 0$ . The large  $X_R^2$  value means that the variance of the residuals is 4 times the variance of the predicted Poisson errors. This situation violates the

assumptions associated with the least squares method. It causes two problems: 1) the calibration estimate may be incorrect because the data are not correctly weighted, and 2) the uncertainty estimates for an unknown sample may be incorrect as well.

We interpret the large  $X_R^2$  as indicating the presence of significant non-Poisson errors. We use an iterative reweighting scheme (fitOption = 1) to estimate the standard error  $SE_r(\Delta_{47})$  for this non-Poisson variation. Iterative reweighting is widely used to estimate additional error terms in least squares (e.g., Holland and Welsch, 1977). The calculation involves an iterative search for a value of  $SE_r(\Delta_{47})$  that results in  $X^2$  equal to its expected value, n-2 (i.e.  $X_R^2 = 1$ ). Equation (21) is used, at each iteration, to calculate a new set of standard errors  $SE(\Delta_{47,i})$  based on the current combination of Poisson and non-Poisson standard errors. Those standard errors are used to define new weights and a new best-fit solution.

This method yielded the following final solution, which we take to be the best estimate for the known synthetic carbonates. The  $R^2$  statistic is 0.93. The estimated value for  $SE_r(\Delta'_{47,SA/ZV}) = 0.0168\%$ . For comparison, those samples with 1800 s count times have  $SE_p(\Delta'_{47,SA/ZV}) = 0.00841\%$  and a total standard error  $SE(\Delta'_{47,SA/ZV}) = 0.0188\%$ .

**Durbin-Watson Test.** Serial correlation or correlated residuals refer to the situation where the sense of offset for the residuals appears to be systematically related to the order of the data, such as the order along the x axis. This result would mean that the residuals are not independently distributed, which is a key assumption for the least-squares method. It may also indicate that the model function is incorrect (see Draper and Smith, 1998, case 3 on p. 55). Figure S2 shows a plot of the standardized residuals for our preferred least-squares solution. Standardization involved dividing the residuals by their standard errors (i.e., the same ones used for the weights). The expectation is that the standardized residuals should be normally distributed with a mean of zero and a standard deviation of 1. In fact, this conclusion has already been tested and supported by the  $\chi^2$  test above. Relevant to our present concern, the plot does show what might be an increase in variance to the right along the  $T^{-2}$  axis. The question is if this pattern is significant or random.



Figure S2. Standardized residuals of  $\Delta_{47}$ -*T* calibration using laboratory precipitated carbonates from this and the Ghosh et al., (2006) studies.

The Durbin-Watson test provides a test for first-order correlation of the residuals along the x axis (Draper and Smith, 1998, p. 68). For weighted least squares, the test statistic is calculated by (Kalina, 2004),

$$\hat{d} = \sum_{j=2}^{n} \left( \sqrt{W_j} \, \hat{e}_j - \sqrt{W_{j-1}} \, \hat{e}_{j-1} \right)^2 / \sum_{j=2}^{n} \left( \sqrt{W_j} \, \hat{e}_j \right)^2 \,, \tag{30}$$

where the subscript *j* indicates that the weights  $W_j$  and residuals  $\hat{e}_j$  for the samples as ordered in terms of increasing  $T^{-2}$  values for the samples. For no serial correlation, the expected value E(d) = 2. The Matlab function "dwtest" is used to calculate the probability for this test.

We applied this test to our iteratively reweighted solution. The Durbin-Watson probability is 75%, which suggests that the distribution of the residuals along the  $T^{-2}$  axis is random.

# Using $\chi^2$ to Compare the Calibration with Other Datasets

The  $\chi^2$  test can be used to compare our calibration with other calibration datasets. The approach is to use the best-fit result for our calibration to calculate residuals for the other data sets,  $\hat{e}_i = \Delta_{47,i} - \hat{b}_0 - \hat{b}_1 T_i^{-2}$ , and then estimate

$$X^{2} = \sum \frac{\hat{e}_{i}^{2}}{SE(\Delta_{47,i})^{2}} = \sum \frac{W_{i}\hat{e}_{i}^{2}}{\sigma_{e}^{2}}$$
(31)

where  $SE^2(\Delta_{47,i})$  is determined from (21). We apply this approach to several data sets of biogenic carbonates from the literature and see that most of them are consistent with the primary calibration, as indicated by a favorable  $\chi^2$  probability (see Table 3 in the main text). In such case, we can include these data in the calibration, as shown in in Figure 4 in the main text.

## **SI-3.** Optimizing Precision

Our calibration analysis shows that  $\Delta_{47}$  measurements are influenced by independent sources of Poisson and non-Poisson errors. The Poisson errors can be reduced by longer count times and the non-Poisson errors, by more replicates. We seek a strategy that will optimize precision relative to machine time.

In the past, we have focused entirely on reducing Poisson errors. For example, at the start, we used "8 s" cycles, meaning that the sample gas was counted for 8 s and then the reference gas for another 8 s. With each cycle, there is about 72 s of downtime, related to instrument activities other than counting. Thus, the total average time of each cycle is 88 s, with downtime accounting for 82% of the cycle. The usual practice is to measure for 90 cycles, which means a total time of 2.2 hours (7,920 s) to measure one sample.

For this study, we used a more efficient protocol with "20 s" cycles, with each cycle taking about 112 s. The fraction of downtime is lower, 64%, and the total time is 2.8 hour (10,080 s). The underlying assumption with this new protocol is that the non-Poisson source of error does not change significantly when averaged over a 20-s interval as opposed to an 8-s interval. This assumption holds if the non-Poisson error is due to sample heterogeneity or to long-term variations in the mass spectrometer.

The old and new protocols produced sample counts of 720 s and 1800 s, respectively. As a result, the new protocol gives measurements with smaller Poisson errors. Equation (21) indicates that total standard error  $SE(\Delta_{47})$  is 0.022‰ versus 0.019‰ for the old and new protocols, respectively (where  $SE_0(\Delta_{47}) = 0.357\%$  and  $SE_r(\Delta_{47}) = 0.017\%$ ). The precision is improved by only 8%, despite the fact that the new protocol has 2.5 times more counts. The reason is that the "8 s" protocol already had a low Poisson error, relative to the non-Poisson error. The "20 s" protocol produced an even smaller Poisson error, but the larger non-Poisson error continued to dominate the total error.

How can we optimize the tradeoff between longer count times and more replicates? The estimated standard error for  $\Delta_{47}$  is recast from (21) to give

$$SE(\Delta_{47})^{2} = \frac{1}{m} \left( \frac{SE_{0}(\Delta_{47})^{2}}{ct_{c}} + SE_{r}(\Delta_{47})^{2} \right),$$
(32)

where *m* is the number of replicate measurements, *c* is the number of cycles, and  $t_c$  is the count time per cycle for the sample gas.

Figure S4 provides a graphical illustration of (32), where the count time per replicate is equal to  $ct_c$ . For example, the "20 s" protocol has 1800 s of count time for each replicate sample, giving SE( $\Delta_{47}$ ) = 0.019‰ for a single replicate. In comparison, two replicates measured with the "8 s" protocol (720 s count time per replicate) would give SE( $\Delta_{47}$ ) = 0.015‰. The precision is improved by 23% and the total count time (1440 s) is 25% lower than that using the "20 s" protocol with one replicate. However, the greater downtime associated with the "8 s" protocol replicates means that the total mass spectrometer time will be greater, by 36%. This situation could be improved by trying other combinations of count-time intervals and number of cycles.



Figure S3. Estimated total standard error,  $SE(\Delta_{47})$ , as a function of count time (s) per replicate and the number of replicates.

The target is to resolve temperature variations at a level better than 2-3 °C, which would require a  $\Delta_{47}$  precision better than 0.010‰. Based on the optimization presented here, this target may be reached by different combinations of count time per cycle, number of cycles, and number of replicates. For example, a strategy of 5 replicates of the "8 s" protocol or of 4 replicates of the "20 s" protocol are both possible, with the total mass spectrometer time (including downtime) being only slightly shorter for the "8 s" case.

The alternative approach is to try to reduce the source of the non-Poisson errors. Possible strategies include a more streamlined procedure, possibly through automation (e.g., Passey et al., 2010), or pretreatment of the carbonate samples to remove organic matter, which could be degrading the mass spectrometer measurements.

# SI-4. Derivation of Confidence Interval for "Inverse Regression" Prediction

Draper and Smith (1998, p. 83-84) describe a method for estimating the confidence interval for a prediction from an "inverse regression" calibration. Their derivation is for a restricted case where the error in the *y* variable for the sample with the "unknown temperature" is scaled to errors in the *y* variable for the calibration. Our derivation here is long, but it provides access to all of the variables in the formula so it is well suited for any case where calibration and prediction errors are needed.



Figure S4. Lower and upper confidence curves for best-fit line (after Figure 3.2 in Draper and Smith, 1998).

Consider an "unknown" sample, where a calibration equation,

$$\hat{X}_0 = (Y_0 - \hat{b}_0)/\hat{b}_1.$$
(33)

is used to predict  $\hat{X}_0$  from measurement  $Y_0$ . The confidence interval is defined by the  $X_L$  and  $X_U$  coordinates for points where the line  $Y = Y_0$  intersects with the upper and lower confidence

curves. These curves represent the endpoints for the  $(1-\alpha)$  confidence intervals for the estimate  $\hat{Y} = \hat{b}_0 + \hat{b}_1 X$ , at a specified *X*.

The two curves are defined by

$$\left(\left\{\Delta Y_{L}, \Delta Y_{U}\right\}\right)^{2} = t_{e}^{2} SE\left(\left\{Y_{L}, Y_{U}\right\}\right)^{2} + t_{0}^{2} SE\left(Y_{0}\right)^{2}$$
(34)

where  $Y_0$  is the input value for the prediction (33),  $Y_L$  and  $Y_U$  are locations on the best-fit line, and  $\Delta Y_L$  and  $\Delta Y_U$  are offsets from the best-fit line to the lower and upper confidence curves, as illustrated in Figure S3. The values of  $\Delta Y_L$  and  $\Delta Y_U$  on the left hand side (left-hand side) of (34) are defined by the "propagation of errors" method on the right hand side (RHS), where the first term represents the uncertainty associated with the least-squares estimates of  $Y_L$  and  $Y_U$  at locations  $X_L$  and  $X_U$ , and the second term represents the uncertainty for the  $Y_0$  measurements used as input for the prediction of  $X_0$ . The variables  $t_e$  and  $t_0$  indicate the score from the tdistribution for a probability P and degrees of freedom v. In particular,  $t_e = t(P = 1 - \alpha/2, v = n - 2)$  and  $t_0 = t(P = 1 - \alpha/2, v = \infty)$ . Note that the confidence interval will generally be asymmetric around  $X_0$ . Setting  $\alpha = 2.5\%$  gives a 95% confidence interval, and  $\alpha = 16\%$  gives a 68% confidence interval, which can be used to estimate the  $SE(X_0)$ .

Next, substitute into (34) the parameters  $\hat{b}_0$  and  $\hat{b}_1$  relationships from the least-square estimate for the calibration line,

$$\left(Y_0 - \hat{b}_0 - \hat{b}_1 \{X_L, X_U\}\right)^2 = t_e^2 s_e^2 \left(\frac{1}{n} + \frac{\left(\{X_L, X_U\} - \overline{x}\right)^2}{S_{xx}}\right) + t_0^2 SE(Y_0)^2.$$
(35a)

Substitute  $Y_0 - \hat{b}_0 = \hat{b}_1 \hat{X}_0$  into the left-hand side of (35a) and then recast the equation so that all terms are on the left-hand side,

$$\hat{b}_{1}^{2}\left(\left\{X_{L}, X_{U}\right\} - \hat{X}_{0}\right)^{2} - t_{e}^{2}s_{e}^{2}\left(\frac{1}{n} + \frac{\left(\left\{X_{L}, X_{U}\right\} - \overline{x}\right)^{2}}{S_{xx}}\right) - t_{0}^{2}SE(Y_{0})^{2} = 0.$$
(35b)

Expand the terms, which gives

$$\hat{b}_{1}^{2} \left\{ X_{L}, X_{U} \right\}^{2} - 2\hat{b}_{1}^{2} \left\{ X_{L}, X_{U} \right\} \hat{X}_{0} + \hat{b}_{1}^{2} \hat{X}_{0}^{2} - \frac{t_{e}^{2} s_{e}^{2}}{S_{xx}} \left\{ X_{L}, X_{U} \right\}^{2} + 2\frac{t_{e}^{2} s_{e}^{2}}{S_{xx}} \left\{ X_{L}, X_{U} \right\} \overline{x} - \frac{t_{e}^{2} s_{e}^{2}}{S_{xx}} \overline{x}^{2} - \frac{t_{e}^{2} s_{e}^{2}}{n} - t_{0}^{2} SE(Y_{0})^{2} = 0.$$
(35c)

Collect terms by powers of  $\{X_L, X_U\}$ ,

$$\begin{pmatrix} \hat{b}_{1}^{2} - \frac{t_{e}^{2} S_{e}^{2}}{S_{xx}} \end{pmatrix} \{ X_{L}, X_{U} \}^{2} - 2 \begin{pmatrix} \hat{b}_{1}^{2} \hat{X}_{0} - \frac{t_{e}^{2} S_{e}^{2}}{S_{xx}} \overline{X} \end{pmatrix} \{ X_{L}, X_{U} \} + \begin{pmatrix} \hat{b}_{1}^{2} \hat{X}_{0}^{2} - t_{e}^{2} S_{e}^{2} \begin{pmatrix} \overline{X}^{2} \\ S_{xx}} + \frac{1}{n} \end{pmatrix} - t_{0}^{2} SE(Y_{0})^{2} \end{pmatrix} = 0.$$
(35d)

Divide through by  $\hat{b}_1^2$  ,

$$\begin{pmatrix} 1 - \frac{t_e^2 s_e^2}{\hat{b}_1^2 S_{xx}} \end{pmatrix} \{ X_L, X_U \}^2 - 2 \left( \hat{X}_0 - \frac{t_e^2 s_e^2}{\hat{b}_1^2 S_{xx}} \overline{x} \right) \{ X_L, X_U \} + \left( \hat{X}_0^2 - \frac{t_e^2 s_e^2}{\hat{b}_1^2 S_{xx}} \left( \overline{x}^2 + \frac{S_{xx}}{n} \right) - \frac{t_0^2 SE(Y_0)^2}{\hat{b}_1^2} \right) = 0.$$
(35e)

Then substitute  $g = \frac{t_e^2 s_e^2}{\hat{b}_1^2 S_{xx}} = \left(\frac{t_e SE(\hat{b}_1)}{\hat{b}_1}\right)^2$  into (35e),

$$(1-g) \{ X_L, X_U \}^2 + 2 (g\overline{x} - \hat{X}_0) \{ X_L, X_U \} + \left( \hat{X}_0^2 - g \left( \overline{x}^2 + \frac{S_{xx}}{n} \right) - \frac{t_0^2 SE(Y_0)^2}{\hat{b}_1^2} \right) = 0.$$
(35f)

Equation (34f) now has the form of a quadratic equation,

$$P\{X_L, X_U\}^2 + Q\{X_L, X_U\} + R = 0.$$
(35g)

The quadratic formula is used to solve for  $\{X_L, X_U\}$ ,

$$\{X_L, X_U\} = \frac{-Q \pm \sqrt{Q^2 - 4PR}}{2P}.$$
(35h)

The variables for the quadratic formula are,

$$P = 1 - g,$$

$$Q = 2(g\overline{x} - \hat{X}_0), \text{ and}$$

$$R = \hat{X}_0^2 - g\left(\overline{x}^2 + \frac{S_{xx}}{n}\right) - \frac{t_0^2 SE(Y_0)^2}{\hat{b}_1^2}.$$
(35i)

Substitution into (35h) gives the solution,

$$\{X_{L}, X_{U}\} = \frac{\left(\hat{X}_{0} - g\overline{x}\right) \pm \sqrt{\left(g\overline{x} - \hat{X}_{0}\right)^{2} - (1 - g)\left(\hat{X}_{0}^{2} - g\left(\overline{x}^{2} + \frac{S_{xx}}{n}\right) - \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}}\right)}{1 - g}.$$
 (35j)

Equation (35j) can be further simplified, as shown in (35k) and (35l),

$$\left\{X_{L}, X_{U}\right\} = \hat{X}_{0} + \frac{\left(\hat{X}_{0} - g\overline{x}\right) - (1 - g)\hat{X}_{0} \pm \sqrt{\left(g\overline{x} - \hat{X}_{0}\right)^{2} - (1 - g)\left(\hat{X}_{0}^{2} - g\left(\overline{x}^{2} + \frac{S_{xx}}{n}\right) - \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}}\right)}{1 - g}, \quad (35k)$$

and

$$\left\{ \Delta X_{L}, \Delta X_{U} \right\} = \frac{\left( \hat{X}_{0} - \overline{x} \right) g \pm \sqrt{\left( g \overline{x} - \hat{X}_{0} \right)^{2} - (1 - g) \left( \hat{X}_{0}^{2} - g \left( \overline{x}^{2} + \frac{S_{xx}}{n} \right) - \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}} \right)}{1 - g}$$
(351)

The equation beneath the radical can be simplified to

$$\begin{aligned} \left(g\overline{x} - \hat{X}_{0}\right)^{2} - (1 - g) \left(\hat{X}_{0}^{2} - g\left(\overline{x}^{2} + \frac{S_{xx}}{n}\right) - \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}}\right) \\ &= g^{2}\overline{x}^{2} - 2g\overline{x}\hat{X}_{0} + \hat{X}_{0}^{2} \\ -\hat{X}_{0}^{2} + g\overline{x}^{2} + g\frac{S_{xx}}{n} + \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}} \\ &+ g\hat{X}_{0}^{2} - g^{2}\overline{x}^{2} - g^{2}\frac{S_{xx}}{n} - g\frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}} \\ &= g\left(\overline{x}^{2} - 2\overline{x}\hat{X}_{0} + \hat{X}_{0}^{2}\right) + (1 - g)g\frac{S_{xx}}{n} + (1 - g)\frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}} \\ &= g\left(\hat{X}_{0} - \overline{x}\right)^{2} + (1 - g)\left(\frac{gS_{xx}}{n} + \frac{t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}}\right) \\ &= \frac{t_{e}^{2} SE(\hat{b}_{1})^{2}}{\hat{b}_{1}^{2}}\left(\hat{X}_{0} - \overline{x}\right)^{2} + (1 - g)\left(\frac{t_{e}^{2}s_{e}^{2}}{n \hat{b}_{1}^{2}} + t_{0}^{2} SE(Y_{0})^{2}}{\hat{b}_{1}^{2}}\right) \\ &= b_{1}^{-2}\left(t_{e}^{2} SE(\hat{b}_{1})^{2}\left(\hat{X}_{0} - \overline{x}\right)^{2} + (1 - g)\left(\frac{t_{e}^{2}s_{e}^{2}}{n \hat{b}_{1}^{2}} + t_{0}^{2} SE(Y_{0})^{2}\right)\right) \end{aligned}$$

Substitution of (35m) into (35l) gives the final equation for the confidence curves,

$$\left\{ \Delta X_{L}, \Delta X_{U} \right\} = \frac{g(\hat{X}_{0} - \overline{x}) \pm \hat{b}_{1}^{-1} \sqrt{t_{e}^{2} SE(\hat{b}_{1})^{2} (\hat{X}_{0} - \overline{x})^{2} + (1 - g) \left(\frac{t_{e}^{2} s_{e}^{2}}{n} + t_{0}^{2} SE(Y_{0})^{2}\right)}{1 - g}.$$
(35n)

The confidence interval provided by (35n) is called the prediction error for  $\hat{X}_0$ , in that it includes the error associated with estimating the calibration equation and the error associated with the measurement  $Y_0$ . There may be instances where one wants the confidence interval for the calibration error alone, which can be determined by setting  $SE(Y_0)^2$  to zero in (35n).

## SI-5. Determining the standard deviation for a group of replicate measurements

Section 3.2 of the paper uses replicated measurements of standards to characterize the variability of  $\Delta_{47}$  measurements. Consider the general case where there are i = 1 to n samples and j = 1 to  $m_i$  replicate measurements for the *i*th sample. The estimate for standard deviation  $SD(\Delta_{47})$  is given by

$$SD^{2}(\Delta_{47}) = \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \left[ \left( \Delta_{47,ij} - \sum_{j=1}^{m_{i}} \Delta_{47,ij} / m_{i} \right)^{2} / (n(m_{i}-1)) \right].$$
(36)

This equation reflects the difference of replicate analyses per sample from the mean value of that sample. Note that one degree of freedom is lost for each calculated mean. For example, in this study, our synthetic carbonates contain 7 samples and 21 replicate measurements (3 replicates for each sample, resulting in14 degrees of freedom); we subtract the mean of each sample from the individual replicate measurements and calculate the variance of these differences.

## **REFERENCES**

Brand, W. A., Assonov, S. S., & Coplen, T. B. (2010). Correction for the <sup>17</sup>O interference in  $\delta^{13}$ C measurements when analyzing CO<sub>2</sub> with stable isotope mass spectrometry (IUPAC Technical Report). Pure and Applied Chemistry, 82(8), 1719–1733. doi:10.1351/PAC-REP-09-01-05

Berkson, J. (1950). Are there two regressions? Journal of the American Statistical Association, 164–180.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). Measurement error in nonlinear models. CRC Press, 438 p.

Buonaccorsi, J. P. (2009). Measurement error. Chapman & Hall/CRC, 428 p.

Dennis, K.J., Affek, H.P., Passey, B.H., Schrag, D.P., Eiler, J.M., (2011). Defining an absolute reference frame for 'clumped' isotope studies of CO<sub>2</sub>. Geochimica et Cosmochimica Acta 75, 7117-7131.

Draper, N. R. and H. Smith (1998). Applied Regression Analysis, Third Edition. Series in Probability and Statistics. Wiley and Son, Inc., 706 p,

Eiler, J.M., (2013). The isotopic anatomies of molecules and minerals. Annual Review of Earth and Planetary Sciences. 41, 411-441

Fuller, W., (1987). Measurement error models: John Wiley and Sons.

Gonfiantini, R., Stichler, W., & Rozanski, K. (1995). Standards and intercomparison materials distributed by the International Atomic Energy Agency for stable isotope measurements (pp. 1–17). International Atomic Energy Agency.

Ghosh, P., Adkins, J., Affek, H., Balta, B., Guo, W.F., Schauble, E.A., Schrag, D., Eller, J.M., (2006a). <sup>13</sup>C<sup>18</sup>O bonds in carbonate minerals: A new kind of paleothermometer. Geochimica et Cosmochimica Acta 70, 1439-1456.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. Communications in Statistics-Theory and Methods, 6(9), 813–827.

Huntington, K.W., Eiler, J.M., Affek, H.P., Guo, W., Bonifacie, M., Yeung, L.Y., Thiagarajan, N., Passey, B., Tripati, A., Daeron, M., Came, R., (2009). Methods and limitations of 'clumped'  $CO_2$  isotope ( $\Delta_{47}$ ) analysis by gas-source isotope ratio mass spectrometry. Journal of Mass Spectrometry 44, 1318-1329.

Kalina, J., (2004). Durbin-Watson test for least weighted squares, in Antoch, J. (ed.) COMPSTAT 2004, Proceedings in Computational Statistics, p. 1287-1294.

Lavagnini, I., and Magno, F., (2006). A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique: Mass Spectrometry Reviews, v. 26, no. 1, p. 1–18, doi: 10.1002/mas.20100.

Merritt, D.A., Hayes, J.M., (1994). Factors controlling precision and accuracy in isotope-ratiomonitoring mass-spectrometry. Analytical Chemistry 66, 2336-2347. Passey, B.H., Levin, N.E., Cerling, T.E., Brown, F.H., Eiler, J.M., (2010). High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. Proceedings of the National Academy of Sciences of the United States of America 107.

Schauble, E.A., Ghosh, P., Eiler, J.M., (2006). Preferential formation of <sup>13</sup>C<sup>18</sup>O bonds in carbonate minerals, estimated using first-principles lattice dynamics. Geochimica et Cosmochimica Acta 70, 2510-2529.

Taylor, J. R. (1997). An introduction to error analysis. Univ Science Books.

Wang, Z.G., Schauble, E.A., Eiler, J.M., (2004). Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. Geochimica et Cosmochimica Acta 68, 4779-4797.

Willett, J.B., and Singer, J.D., (1988). Another cautionary note about R<sup>2</sup>: Its use in weighted least-squares regression analysis: American Statistician, p. 236–238.

York, D., Evensen, N., Martinez, M., & Delgado, J. (2004). Unified equations for the slope, intercept, and standard errors of the best straight line. American Journal of Physics, 72(3), 367–375. doi:10.1119/1.1632486