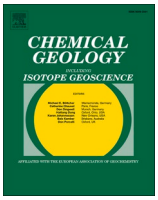
Chemical Geology 647 (2024) 121881

Contents lists available at ScienceDirect

Chemical Geology

journal homepage: www.elsevier.com/locate/chemgeo



Omnivariant Generalized Least Squares regression: Theory, geochronological applications, and making the case for reconciled Δ_{47} calibrations

M. Daëron [a,*], P. Vermeesch [b]

[a] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France
[b] Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Editor: Christian France-Lanord

Keywords:
Statistics
Regression
Least squares
Geochronology
Clumped isotopes

ABSTRACT

Least-squares regression methods are mathematically powerful, conceptually and computationally simple, and widely used in many fields. However, none of the commonly-used flavors of least-squares regression, such as York regression or Generalized Least Squares (GLS), take into account the full set of covariances between all observed (x, y) values. Here we describe the Omnivariant Generalized Least Squares (OGLS) method to fit a model of the form $y = f(x)$, accounting for the full error correlation structure of the (x, y) data, based on a first-order linear propagation of the uncertainties in all variables into errors in y residuals, followed by minimizing the vector of y residuals with respect to the Mahalanobis norm defined by its covariance matrix. This approach may be described as a generalization of both York regression and GLS. It is mathematically exact for straight-line fits, and is also suitable for many non-linear models. Here we describe the principles of OGLS regression and discuss its properties, caveats, and practical use, and provide two consistent open-source implementations in Python and R. To illustrate how various fields of geochronology and stable-isotope geochemistry may benefit from this new method, we discuss how OGLS may specifically apply to $^{40}Ar/^{39}Ar$ dating and how it provides robust mathematical evidence that Δ_{47} carbonate calibrations in the recently defined I-CDES metrological scale are statistically indistinguishable, effectively solving long-standing methodological discrepancies.

1. Introduction

When investigating the physical laws of the natural world, we often find ourselves attempting to establish a mathematical relationship linking two or more numerical variables based on a set of observations. This is usually because we wish to estimate some parameter of this relationship (e.g., the slope of a regression line) or because we hope to use this relationship as a general model linking these variables in other data sets (e.g., when calibrating a thermometer). The use of such subjective terms as *wish* and *hope* calls attention to the fact that regression analysis derives from an underlying purpose which strongly influences the choices we make when designing regression models. In a general sense, determining optimal ("best-fitting") model parameters is conceptually simple, however computationally difficult it may be, but this optimization step first requires us to decide what kind of mathematical relationship is well-suited both to our objectives and to the data available, and to define explicitly what "best-fitting" means, which usually requires assigning statistical weights to the observations. After this parameter optimization step, it is often useful to assess the quality of the resulting model, usually based on some goodness-of-fit statistic; to check, ideally using objective mathematical methods, whether some observations deviate anomalously from the model; to estimate the predictive power of the model, e.g., expressed as prediction confidence intervals; to estimate confidence intervals for the best-fitting parameters of interest.

All of the above steps depend, in one way or another, on the statistical weight assigned a priori to each datum, or, equivalently, on characterizing the uncertainties associated with a set of observations. Geochemistry has a long tradition of paying proper attention to measurement uncertainties and their propagation into quantitative interpretations. In many cases — such as virtually all of geochronology — these uncertainties are one of the primary factors limiting interpretations, leading to the creation of specialized mathematical frameworks aiming to model and propagate various sources of uncer-

* Corresponding author.
E-mail addresses: daeron@lsce.ipsl.fr (M. Daëron), p.vermeesch@ucl.ac.uk (P. Vermeesch).

https://doi.org/10.1016/j.chemgeo.2023.121881
Received 17 September 2023; Received in revised form 6 December 2023; Accepted 7 December 2023
Available online 18 December 2023

tainty as accurately as possible [e.g., Ludwig, 1998; McLean et al., 2011; Vermeesch, 2018; Daëron, 2021]. One notable feature of geochemical observations is that measurement errors are often correlated in one way of the other. In Pb-Pb dating, for instance, when measuring the isotopic ratios $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$, large uncertainties in the determination of ^{204}Pb contribute simultaneously to both ratios, and this shared source of uncertainty causes statistical errors in the two ratios to be positively correlated [Connelly et al., 2021]. Another kind of error covariance occurs when a group of analyses are corrected/standardized based on a common set of reference material measurements. If the analytical uncertainty derived from this correction step is large enough, errors in the corrected measurements may be correlated in complex but non-negligible ways [e.g., Daëron, 2021]. In the most general case, estimates of the analytical uncertainties associated with a set of N scalar measurements are described by a probability density function over the N-dimensional measurement space, but this can often be simplified to a multivariate Gaussian probability distribution which is fully characterized by a $N \times N$ "variance-covariance" matrix specifying the uncertainty (or variance, or standard error) of each measured quantity along with the the correlations (or covariances) between these measurement errors.

There is a large body of work on how to propagate these non-independent measurement uncertainties when fitting regression models — or, more generally, when solving any kind of inverse problem — using different but related mathematical formalisms such as maximum-likelihood estimation, Bayesian inference, or least-squares estimation. Among these approaches, least-squares methods are most widely used (and taught) because they are conceptually and computationally simple. Although it is well known that least squares are not robust against statistical outliers, geochemical measurements usually undergo quality-control procedures which are expected, at least in theory, to largely mitigate this issue.

Different "flavors" of least-squares regression with somewhat ambiguous denominations are widely used, each corresponding to a different assigment of statistical weights (or, equivalently, to different structures of observation uncertainties). For instance, "Ordinary Least Squares" only considers equally-weighted regression residuals on a "response" variable (the y scalar, in the classical case of fitting y as a function of x), whereas "Orthogonal Distance Least Squares" assigns the same weights to residuals on both predictor and response variables (x, y), and "Generalized Least Squares" only considers errors in the response variable but takes error correlations into account. Section 2.3 provides a brief overview of these different methods.

The least squares formulation with the most general specification of observation uncertainties, i.e. the only one allowing for arbitrarily complex correlation between all errors, is "Weighted Total Least Squares" [Kukush and Van Huffel, 2004; Markovsky and Van Huffel, 2007]. However, to the best of our knowledge, this approach is virtually never used by geochemists, most likely because it is usually described in the context — and using the jargon — of linear algebra. What's more, although Total Least Squares as traditionally defined [Golub and Van Loan, 1980; Van Huffel and Vandewalle, 1991] is, strictly speaking, a linear regression problem, it should be possible in many cases to solve non-linear regression problems using a modified weighted total least squares approach.

Here we propose a new formulation for non-linear least-squares regression of observational data with Gaussian error distributions and arbitrarily complex error correlations between all variables, based on a first-order linear propagation of the uncertainties in all variables into errors in response residuals, followed by minimizing the vector of response residuals with respect to the Mahalanobis norm defined by its covariance matrix. Because such verbiage is likely to scare off most geochemists, we will devote a large part of this article to reformulating the above description in layman's terms. For now, this "Omnivariant Generalized Least Squares" (OGLS) approach can be considered an even more general version of Generalized Least Squares accounting for arbitrary covariances between all predictor and response variables. In the first two sections we start by summarizing the broader context of least-squares methods, before describing the theoretical basis of OGLS and addressing various practical issues including goodness-of-fit tests. Based on this detailed description and the accompanying open-source implementation, it should be straightforward to apply OGLS to a wide array of scientific fields.

Section 4 describes the application of OGLS to linear regression in the geochronological context of two-dimensional isochron fitting. Although geochronologists have long recognized the importance of error correlations between the dependent and independent variables in isochron regression, they have only recently started to appreciate the existence of significant error correlations between different aliquots of the same sample. By accounting for these inter-sample error correlations, OGLS improves both the accuracy and precision of the resulting age estimates. Section 4.3 also shows how excess dispersion of the isotopic data around the best fit line can be quantified using the method of maximum likelihood, which builds on the OGLS algorithm.

Section 5 applies OGLS to stable isotope geochemistry, revisiting the thorny issue of carbonate clumped-isotope calibrations, which are characterized by sizable and blockwise-correlated uncertainties in both the predictor (temperature) and the response (Δ_{47}) variables. We conclude, based on OGLS-adapted goodness-of-fit estimates, that Δ_{47} calibration data sets (re)processed in the new I-CDES metrological scale are in statistical agreement with each other, and that a single quadratic calibration function provides a good fit to the whole data set ($N = 104$). This finding, if it holds true, marks another milestone in the progress of clumped-isotope thermometry, which has long been plagued by discrepancies in calibration data reported by different groups.

As noted above, regression analysis is rooted in purpose. We, the authors, very recently found out that we had both been working for years on a more general least-squares approach and that, despite being motivated by entirely distinct applications, our efforts had independently converged to essentially the same mathematical solution. We decided that the best course of action was to combine our findings into a single manuscript explicitly addressing both of the separate problems that initially prompted our work. This article represents an 80/20 split of contributions. MD wrote the entire paper except for Section 3 and the toy examples of Fig. 2, which were contributed by PV. MD wrote the implementation of OGLS in Python, whereas PV wrote the implementation in `R`.

2. Background

2.1. Notation

Scalars are printed in italics (e.g., "x"). Vectors are typeset in lowercase boldface and are considered to be column vectors by default, with vector elements indexed from top to bottom:

$$\mathbf{x} = [x_i] = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{x}^\top = [x_i]^\top = (x_1 \ x_2 \ \dots \ x_n)$$

Matrices are typeset in uppercase boldface, with elements indexed from top to bottom then from left to right:

$$\mathbf{X} = [x_{i,j}] = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{pmatrix}$$

A matrix can also be represented as a combination of smaller matrices. In that case, enclosing parentheses are replaced by square brackets:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix}$$

$$\begin{bmatrix} \mathbf{A} & \boldsymbol{B} \\ \mathbf{B}^{\top} & \mathbf{A} \end{bmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & b_{1,1} & b_{1,2} \\ a_{2,1} & a_{2,2} & b_{2,1} & b_{2,2} \\ b_{1,1} & b_{2,1} & a_{1,1} & a_{1,2} \\ b_{1,2} & b_{2,2} & a_{2,1} & a_{2,2} \end{pmatrix}$$

2.2. Characterizing correlations between measurement uncertainties

Here we consider random, zero-centered, Gaussian measurement errors affecting a series of measurements or analyses, hereafter called *observations*. We will mostly limit ourselves to two-dimensional observations, where each datum is described by two scalars (x,y), but OGLS applies equally well to multi-dimensional predictor and response variables.

As mentioned above, analytical errors in geochemistry are often correlated with each other. Let us consider a set of N observations noted (x_i,y_i), noting $\mathbf{x} = [x_i]$ and $\mathbf{y} = [y_i]$. In the most general case, the observational uncertainties associated with these x_i and y_i values can be described by a $2N \times 2N$ covariance matrix:

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}^{\top}} & \mathbf{V_x} \end{bmatrix} \tag{1}$$

where $\mathbf{V_y}$ is the variance-covariance matrix for $[y_i]$ values, $\mathbf{V_x}$ that for $[x_i]$ values, and $\mathbf{V_{yx}}$ is the (non-symmetric) matrix of all covariances between y_i and x_j, noted $\omega_{y_i x_j}$:

$$V_y = \begin{pmatrix} \sigma^2_{y_1} & \omega_{y_1 y_2} & \dots & \omega_{y_1 y_N} \\ \omega_{y_1 y_2} & \sigma^2_{y_2} & \dots & \omega_{y_2 y_N} \\ \vdots & \vdots & & \vdots \\ \omega_{y_1 y_N} & \omega_{y_2 y_N} & \dots & \sigma^2_{y_N} \end{pmatrix}$$

$$V_x = \begin{pmatrix} \sigma^2_{x_1} & \omega_{x_1 x_2} & \dots & \omega_{x_1 x_N} \\ \omega_{x_1 x_2} & \sigma^2_{x_2} & \dots & \omega_{x_2 x_N} \\ \vdots & \vdots & & \vdots \\ \omega_{x_1 x_N} & \omega_{x_2 x_N} & \dots & \sigma^2_{x_N} \end{pmatrix} \tag{2}$$

$$V_{yx} = \begin{pmatrix} \omega_{y_1 x_1} & \omega_{y_1 x_2} & \dots & \omega_{y_1 x_N} \\ \omega_{y_2 x_1} & \omega_{y_2 x_2} & \dots & \omega_{y_2 x_N} \\ \vdots & \vdots & & \vdots \\ \omega_{y_N x_1} & \omega_{y_N x_2} & \dots & \omega_{y_N x_N} \end{pmatrix} \tag{3}$$

The diagonal elements of $\mathbf{V_y}$ and $\mathbf{V_x}$ correspond to the uncertainties (statistical variance) associated with each y_i and x_i value, respectively. The diagonal elements of $\mathbf{V_{yx}}$ describe the covariance between x and y within each datum (x_i,y_i). When plotting observations in (x,y) space, these covariance terms are usually displayed using error ellipses instead of error bars (Fig. 1), making it obvious whether errors in x and y are independent, positively correlated, or anti-correlated.

We lack a generally accepted graphical representation for the non-diagonal terms $\omega_{y_i y_j}$ in $\mathbf{V_y}$, which characterize the covariances between these y_i values. When the measurements of y_i and y_j are statistically independent, these non-diagonal terms are equal to zero, but this is not necessarily the case, because some source(s) of uncertainty may be shared by several measurements (e.g., when taking into account uncertainties associated with correction/standardization procedures). The same applies to non-diagonal elements of $\mathbf{V_x}$.

Non-diagonal elements of $\mathbf{V_{yx}}$ ($\omega_{y_i x_j}$ with $i \neq j$) are zero in most cases because the uncertainty of the predictor variable x_j for one observation is usually not correlated with that of the response variable y_i for a different observation. Nevertheless, for the sake of generality, we also consider here non-zero $\omega_{y_i x_j}$ values.

2.3. Least-squares regression

2.3.1. Ordinary, weighted, and generalized least-squares regression

Least-squares methods were developped independently by Legendre [1805] and Gauss [1809], both of whom used them to predict the trajectories of heavenly bodies. Least-squares regression postulates that the optimal model to describe an over-determined set of (x_i,y_i) observations with $1 \leqslant i \leqslant N$, among a family of models of the form $y = f\left(x, p_1, p_2 \dots p_{N_p}\right)$, where $\left(p_1, p_2 \dots p_{N_p}\right)$ are the model parameters, is that which minimizes the sum of square residuals along the y dimension, e.g.:

$$\chi^2 = \sum_{i=1}^{N} \left(y_i - f(x_i, p_1, p_2 \dots)\right)^2 \tag{4}$$

This quantity is a χ^2 statistic summarizing the differences between observed and predicted y values. The above formula corresponds to *Ordinary Least Squares* (OLS), the simplest version of least-squares regression, in which the residuals are not scaled in any way, so that χ^2 has the same dimension as y^2.

A slighty different method named *Weighted Least Squares* (WLS) scales each residual term by a weight usually equal to the inverse squared uncertainty (or squared standard error, or variance) of each y_i observation:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i, p_1, p_2 \dots)}{\sigma_{y_i}}\right)^2 \tag{5}$$

In that case χ^2 is dimensionless and, if the weighted residuals are statistically independent and normally distributed with a variance of 1, the minimized value of χ^2 should follow a chi-squared distribution with $N_f = (N - N_p)$ degrees of freedom $\left(\chi^2_{N_f}\right)$. The mean of this distribution is equal to N_f, which leads naturally to considering the "reduced" χ^2 value — defined as χ^2/N_f —, also called mean squared weighted deviation

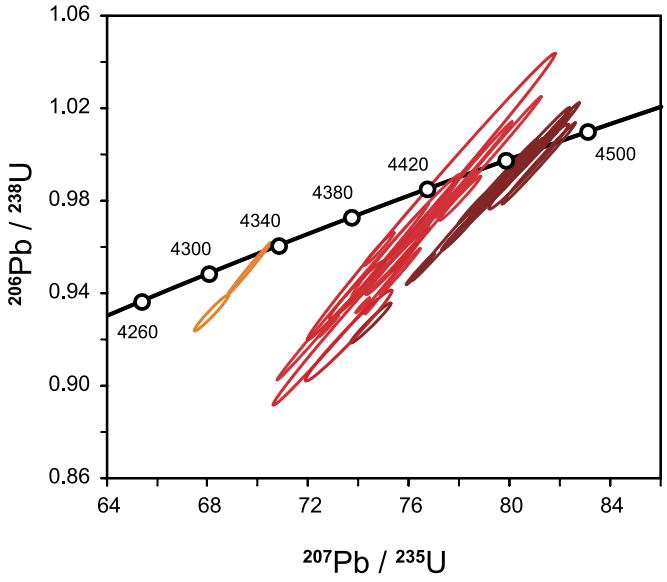


Fig. 1. Example of covariance between x and y within each observation. Projecting these ($^{207}Pb/^{235}U$, $^{206}Pb/^{238}U$) observations on the concordia line must take into account the strong covariance of the two measurements, strongly influencing the final ages. Plot modified from Fig. 1A of Costa et al. [2020].

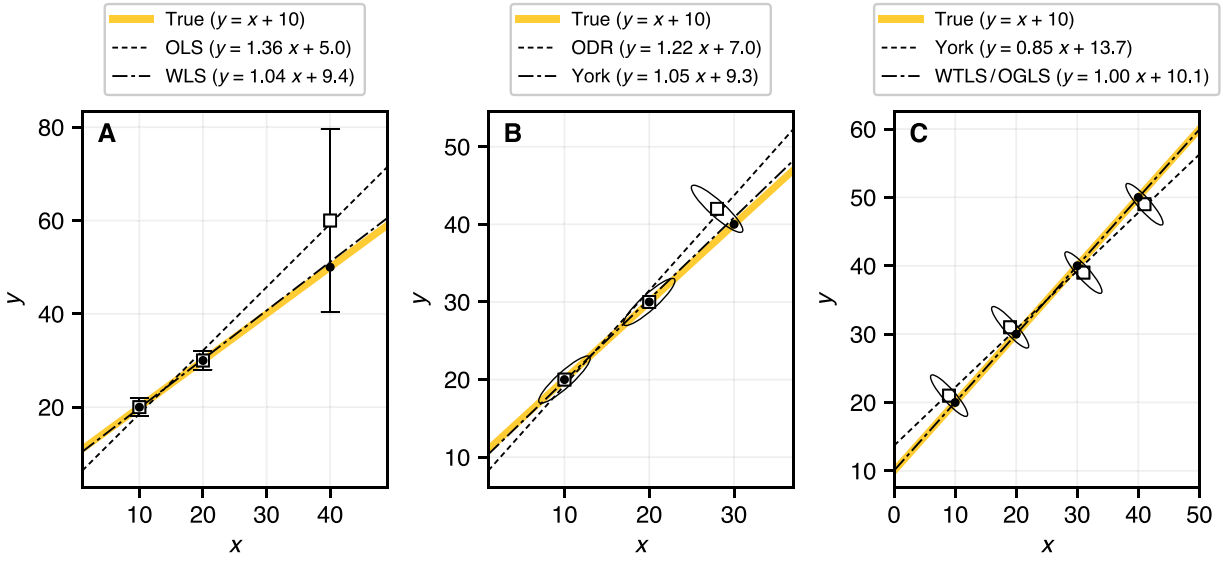


Fig. 2. Three toy examples illustrating of the benefits of error-weighted linear regression. In all threes cases the true x and y values (black circles) belong to a straight line (in yellow) with a slope of 1 and an intercept of 10. White squares represent random realizations of these samples, given Gaussian uncertainties shown as 95% confidence bars or ellipses. Panel A: given three samples with uncertainties only in the y variable, OLS regression ignores these uncertainties, resulting in a poor fit compared to WLS. Panel B: given three samples, each having correlated (x, y) uncertainties (listed in Table 2), ODR regression ignores these correlations, resulting in a poor fit compared to York regression. Panel C: given four samples whose uncertainties are not only correlated within each sample, but also *between* samples (see Table 3), York regression ignores these inter-sample correlations, resulting in a poor fit compared to the (equivalent) OGLS and WTLS regression. In each of these three cases, the most accurate results are obtained by properly taking into account all relevant sources of error correlation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Notations.

N	number of (x, y) observations
N_p	number of model parameters
N_f	number of model degrees of freedom, usually equal to $(N - N_p)$
f	model function linking x and y, with $y = f(x, \mathbf{p})$
$\mathbf{x}$	N-vector of x observations
$\mathbf{y}$	N-vector of y observations
$\boldsymbol{\xi}$	N-vector of error-free ("true") values of x
$\boldsymbol{\zeta}$	N-vector of Cholesky residuals
$\mathbf{p}$	N_p-vector of model parameters
$\mathbf{p}^*$	N_p-vector of the best-fit model parameters
$\mathbf{r_y}$	vector of model residuals in the y dimension
$\mathbf{r_x}$	vector of model residuals in the x dimension
$\mathbf{r_y^*}$	vector of the best-fit model residuals in the y dimension
σ_x^2	variance of x
ρ_{xy}	correlation of (x, y)
ω_{xy}	covariance of (x, y), equal to $\rho_{xy}\sigma_x\sigma_y$
$\mathbf{V_x}$	$N \times N$ covariance matrix of $[x_i]$
$\mathbf{V_y}$	$N \times N$ covariance matrix of $[y_i]$
$\mathbf{V_{yx}}$	$N \times N$ covariance matrix between $[y_i]$ and $[x_i]$
$\mathbf{V}$	$2N \times 2N$ covariance matrix of $(y_1, y_2, \ldots, x_1, x_2 \ldots)$
$\mathbf{V_{r_y}}$	$N \times N$ covariance matrix of $\mathbf{r_y}$
$\mathbf{V_{p^*}}$	$N_p \times N_p$ covariance matrix of $\mathbf{p}^*$
$\mathbf{J_{r_y}}$	$N \times 2N$ Jacobian matrix of $\mathbf{r_y}$ with respect to $(y_1, y_2, \ldots, x_1, x_2 \ldots)$
$\mathbf{J_{f\|p}}$	Jacobian matrix of f with respect to $\mathbf{p}$

Table 2
Synthetic data for toy example of Fig. 2B.

	True		Observed				
i	x_i	y_i	x_i	y_i	σ_{x_i}	σ_{y_i}	$\sigma^2_{x_i y_i}$
1	10	20	10	20	1	1	+ 0.9
2	20	30	20	30	1	1	+ 0.9
3	30	40	28	42	1	1	− 0.9

(MSWD) in geochronology. Because the expected mean value of χ^2/N_f is equal to one and its variance inversely proportional to N_f, it can be used as a goodness-of-fit criterion when N_f is large enough [Wendt and Carl, 1991]. Fig. 2A compares OLS and WLS regression using a simple example of three colinear points, where the y value of the third point is ten times less precise than the other y values. In this situation, WLS yields results far more accurate than OLS.

If the residuals terms in Eq. (5) are not statistically independent from each other, we may instead use *Generalized Least Squares* (GLS), where χ^2 is reformulated to account for the full covariance structure of the $[y_i]$ vector:

$$\chi^2 = \mathbf{r_y}\ \mathbf{V_y}^{-1}\ \mathbf{r_y} \quad \text{with } \mathbf{V_y} \text{ defined in (2) and } \quad \mathbf{r_y} = \begin{pmatrix} y_1 - f(x_1, p_1, p_2 \ldots) \\ y_2 - f(x_2, p_1, p_2 \ldots) \\ \vdots \\ y_N - f(x_N, p_1, p_2 \ldots) \end{pmatrix} \tag{6}$$

From this brief overview it should be clear that Eq. (4) can be viewed as a special case of (5), which is itself a special case of (6), and that Generalized Least Squares, despite its name, ignores all observational uncertainties on $[x_i]$ values.

2.3.2. Least-squares regression with multivariate uncertainties

In many cases, only considering uncertainties in the y dimension is not a valid option. Going back to Adcock [1878] and Kummell [1879], various "errors-in-variables" methods have been formulated to perform least-squares regression on data with comparably large errors in both x and y. A well-known example is *Orthogonal Distance Regression* (ODR), which minimizes the sum of squared orthogonal distances between (x, y) observations and a straight regression line. The more general straight line regression method of York et al. [2004] takes into account arbitrary uncertainties in all x_i and y_i values (the diagonal elements of $\mathbf{V_y}$ and $\mathbf{V_x}$), as well as the covariance between x_i and y_i within each observation (the diagonal elements of $\mathbf{V_{yx}}$). Fig. 2B illustrates the benefits of York

Table 3
Synthetic data for toy example of Fig. 2C. The true x and y are $(10, 20, 30, 40)$ and $(20, 30, 40, 50)$, respectively, consistent with an intercept of 10 and a slope of 1.

$$y = \begin{pmatrix} 21 \\ 31 \\ 39 \\ 49 \end{pmatrix} \qquad x = \begin{pmatrix} 9 \\ 19 \\ 31 \\ 41 \end{pmatrix} \qquad cov\left(\begin{bmatrix} y \\ x \end{bmatrix}\right) = \begin{pmatrix} 1 & 0.99 & 0 & 0 & -0.99 & -0.99 & 0 & 0 \\ 0.99 & 1 & 0 & 0 & -0.99 & -0.99 & 0 & 0 \\ 0 & 0 & 1 & 0.99 & 0 & 0 & -0.99 & -0.99 \\ 0 & 0 & 0.99 & 1 & 0 & 0 & -0.99 & -0.99 \\ -0.99 & -0.99 & 0 & 0 & 1 & 0.99 & 0 & 0 \\ -0.99 & -0.99 & 0 & 0 & 0.99 & 1 & 0 & 0 \\ 0 & 0 & -0.99 & -0.99 & 0 & 0 & 1 & 0.99 \\ 0 & 0 & -0.99 & -0.99 & 0 & 0 & 0.99 & 1 \end{pmatrix}$$

regression over ODR regression using another three-point toy example.

Beyond straight line regressions, accounting for bivariate uncertainties in least-square models can be achieved by using a modified version of WLS based on replacing each $\sigma^2_{y_i}$ variance in (5) with the corresponding *effective variance* (EV), noted σ^2_e [Tellinghuisen, 2020]:

$$\sigma^2_e(x_i, y_i) = \sigma^2_{y_i} + \left(\frac{\partial f}{\partial x}(x_i)\right)^2 \cdot \sigma^2_{x_i} \tag{7}$$

This formula assumes that x_i and y_i are not correlated, but otherwise the effective variance may be reformulated as:

$$\sigma^2_e(x_i, y_i) = \sigma^2_{y_i} + \left(\frac{\partial f}{\partial x}(x_i)\right)^2 \cdot \sigma^2_{x_i} - 2 \cdot \frac{\partial f}{\partial x}(x_i) \cdot \omega_{x_i y_i} \tag{8}$$

which corresponds exactly to York's regression in the case of a straight-line fit. Once again, ODR can be viewed as a special case of York regression, itself a special case of the EV method, all of which take only partially into account the full covariance structure of observations.

2.3.3. Total Variance methods

All of the previous methods seek to minimize, with different weighting choices, the squared sum of residuals in the response variable (y). A different approach, usually attributed to Deming [1943] and more recently labeled *Weighted Total Least Squares* (WTLS), almost entirely forgoes the distinction between predictor and response variables by aiming to minimize the *total variance* (TV), defined as:

$$\mathrm{TV} = \mathbf{r_y}^\top \cdot \mathbf{V_y}^{-1} \cdot \mathbf{r_y} + \mathbf{r_x}^\top \cdot \mathbf{V_x}^{-1} \cdot \mathbf{r_x} \tag{9}$$

where $\mathbf{r_x}$ and $\mathbf{r_y}$ are the residual vectors or x and y observations relative to their error-free ("true") values:

$$\begin{aligned} \mathbf{r_x} &= [x_i - \xi_i] \qquad \xi_i \text{ being the error} - \text{free value of } x_i \\ \mathbf{r_y} &= [y_i - f(\xi_i, p_1, p_2 \ldots)] \end{aligned} \tag{10}$$

WTLS can be slightly reformulated to also account for $\mathbf{V_{yx}}$ terms:

$$\mathrm{TV} = \begin{bmatrix} \mathbf{r_y} \\ \mathbf{r_x} \end{bmatrix}^\top \cdot \begin{bmatrix} \mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}}^\top & \mathbf{V_x} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{r_y} \\ \mathbf{r_x} \end{bmatrix} \tag{11}$$

Total least-squares methods were reviewed by Markovsky and Van Huffel [2007]. Fig. 2C illustrates their benefits over York regression using the four-sample toy dataset shown in Table 3. To say the least, they have never been very popular among geochemists, among other factors because they are defined and discussed using the specialized idiom of linear algebra. What's more, whereas solving a traditional least-squares problem requires optimizing only the model parameters $[p_i]$, WTLS needs to find an optimal combination of $[p_i]$ *and* $[\xi_i]$, considerably increasing the dimensionality of the parameter space to explore. As a result, many "simple" regressions with a small number of model parameters and a few tens of observations quickly become challenging from a computational point of view. A notable exception is the case of fitting a straight line, which may be solved analytically using the maximum likelihood algorithm described below (Section 4.3).

3. Omnivariant Generalized Least-Square regression

3.1. General formulation

Going back to the GLS definition of χ^2 (Eq. (6) above), this formulation corresponds to the squared *Mahalanobis distance* between $\mathbf{r}$ and a zero-centered statistical distribution of $\mathbf{r}$ based on $\mathbf{V_y}$. The Mahalanobis distance is a multivariate generalization of a Z-score, or "number of standard deviations" [cf De Maesschalck et al., 2000]. In three-dimensional Euclidean space, the isosurface corresponding to a Euclidean distance of one from the null vector is a sphere. By contrast, given any covariance matrix $\mathbf{V}$, we can define the corresponding Mahalanobis distance $M_\mathbf{V}$, and the isosurface defined by $M_\mathbf{V} = 1$ is an ellipsoid whose half principal dimensions are the square-root eigenvalues of $\mathbf{V}$, and whose principal axes are oriented according to the correlation structure of $\mathbf{V}$. By design, minimizing the Mahalanobis length of $\mathbf{r}$ thus accounts for scaling the residuals according to their observational standard errors (as in simple WLS) but also for the statistical correlations between them.

Here, we propose to combine the Mahalanobis distance criterion of GLS with a generalized version of effective variance to take into account the entire covariance structure of (x, y). To clearly distinguish this approach from existing methods, we propose, somewhat presumptuously, the denomination "Omnivariant Generalized Least Squares" (OGLS).

In essence, the key feature of OGLS is to express $\mathbf{V_{r_y}}$, the covariance matrix of $\mathbf{r_y}$, using the classical first-order expansion for error propagation:

$$\mathbf{V_{r_y}} = \mathbf{J_{r_y}}^\top \cdot \begin{bmatrix} \mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}}^\top & \mathbf{V_x} \end{bmatrix} \cdot \mathbf{J_{r_y}} \tag{12}$$

with $\mathbf{J_{r_y}}$ being the Jacobian of $\mathbf{r_y}$ relative to $[y_1, y_2, \ldots, x_1, x_2, \ldots]$. For example, in the simple case of three observations (x_1, y_1), (x_2, y_2), and (x_3, y_3):

$$\mathbf{J_{r_y}} = \begin{pmatrix} 1 & 0 & 0 & -\partial f/\partial x_1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -\partial f/\partial x_2 & 0 \\ 0 & 0 & 1 & 0 & 0 & -\partial f/\partial x_3 \end{pmatrix}^\mathrm{T} \tag{13}$$

In practice, $\mathbf{r_y}$ and $\mathbf{J_{r_y}}$ both depend on the regression parameters. Minimizing the squared Mahalanobis distance $\mathbf{r_y}^\top \mathbf{V_{r_y}}^{-1} \mathbf{r_y}$ thus needs to be done numerically, using one of many available optimization methods. Leaving the implementation details aside, it bears pointing out that each of the least-squares methods considered in Sections 2.3.1 and 2.3.2 (OLS, WLS, GLS, ODR, York/EV regression) can be expressed as restricted cases of OGLS regression (Fig. 3).

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(a) Ordinary Least Squares

$$\begin{pmatrix} \sigma^2_{y_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{y_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{y_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(b) Weighted Least Squares

$$\begin{pmatrix} \sigma^2_{y_1} & \omega_{y_1y_2} & \omega_{y_1y_3} & 0 & 0 & 0 \\ \omega_{y_1y_2} & \sigma^2_{y_2} & \omega_{y_2y_3} & 0 & 0 & 0 \\ \omega_{y_1y_3} & \omega_{y_2y_3} & \sigma^2_{y_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(c) Generalized Least Squares

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(d) Orthogonal Distance Regression

$$\begin{pmatrix} \sigma^2_{y_1} & 0 & 0 & \omega_{y_1x_1} & 0 & 0 \\ 0 & \sigma^2_{y_2} & 0 & 0 & \omega_{y_2x_2} & 0 \\ 0 & 0 & \sigma^2_{y_3} & 0 & 0 & \omega_{y_3x_3} \\ \omega_{y_1x_1} & 0 & 0 & \sigma^2_{x_1} & 0 & 0 \\ 0 & \omega_{y_2x_2} & 0 & 0 & \sigma^2_{x_2} & 0 \\ 0 & 0 & \omega_{y_3x_3} & 0 & 0 & \sigma^2_{x_3} \end{pmatrix}$$

(e) York regression with correlated (x_i, y_i) pairs

Fig. 3. Various classical least-squares regression methods, which differ primarily in how they account for observation uncertainties, may be described as special cases of a general formulation taking into account the full covariance matrix $\mathbf{V}$ of all x and y values (Eq. (1)), as illustrated here in the simple case of three observations (x_1, y_1), (x_2, y_2), and (x_3, y_3).

3.2. Properties

In the following section we consider a few simple examples of polynomial regression and briefly discuss some general properties of OGLS models.

3.2.1. Equivalence with York regression

York et al. [2004] only account for the diagonal elements of $\mathbf{V_x}$, $\mathbf{V_y}$ and $\mathbf{V_{yx}}$ (Fig. 3e). In this approach, the weight W_i assigned to each (x_i, y_i) observation is the inverse effective variance, depending on $\sigma^2_{x_i}$, $\sigma^2_{y_i}$, $\omega_{x_iy_i}$, and b, the slope of the regression line:

$$W_i = \frac{1}{\sigma^2_e(x_i, y_i)} = \frac{1}{\sigma^2_{y_i} + b\sigma^2_{x_i} - 2b\omega_{x_iy_i}} \quad \text{from Table I of } \textit{York et al.} \left[2004\right] \tag{14}$$

Going back to the OGLS formulation, the Jacobian for the residuals of a straight-line model of slope b is (e.g., for $N = 3$ observations):

$$\mathbf{J_{r_y}} = \begin{pmatrix} 1 & 0 & 0 & -b & 0 & 0 \\ 0 & 1 & 0 & 0 & -b & 0 \\ 0 & 0 & 1 & 0 & 0 & -b \end{pmatrix}^{\mathrm{T}} \tag{15}$$

As a result, the covariance matrix of the residuals is:

$$\mathbf{V_{r_y}} = \begin{pmatrix} \sigma^2_e(x_1,\ y_1) & 0 & 0 \\ 0 & \sigma^2_e(x_2,\ y_2) & 0 \\ 0 & 0 & \sigma^2_e(x_3,\ y_3) \end{pmatrix} \tag{16}$$

York regression is thus mathematically equivalent to a straight-line OGLS regression with diagonal $\mathbf{V_y}$, $\mathbf{V_x}$ and $\mathbf{V_{yx}}$ matrices, and, as illustrated in Fig. 4, both methods yield identical results within numerical precision.

3.2.2. Effects of different types of covariance

Two different kinds of observation covariance are frequently encountered in geochemical data. In the first case, the (x_i, y_i) pairs of observations are statistically independent from each other, but errors in

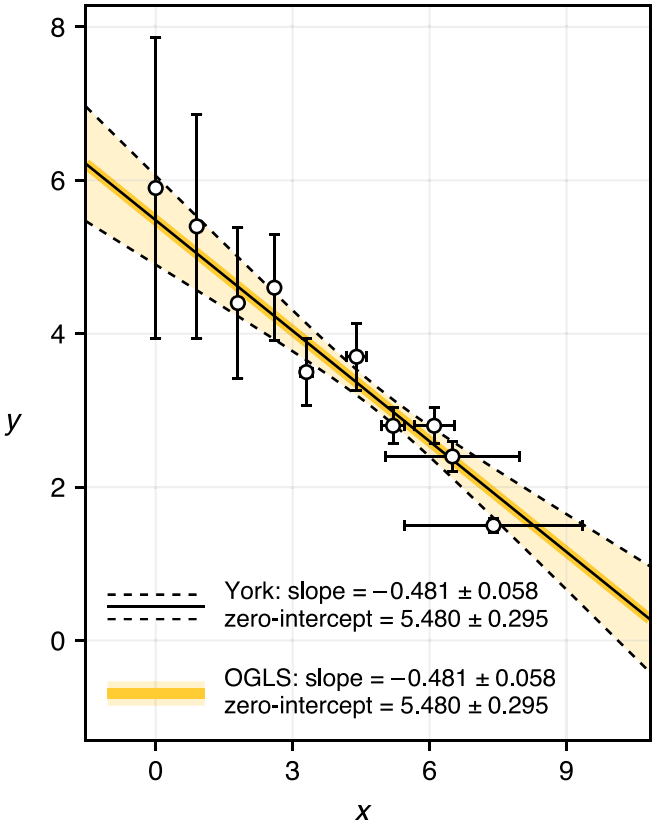


Fig. 4. OGLS and York regression yield identical results. The (x, y) data shown here correspond to the (in)famous Pearson/York data set [York, 1966]. Processing these data using OGLS regression or the straight-line fitting method of York et al. [2004] yields identical results. Error bars, dashed lines and the yellow area indicate 95% confidence regions of data points, York best-fit line, and OGLS line, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

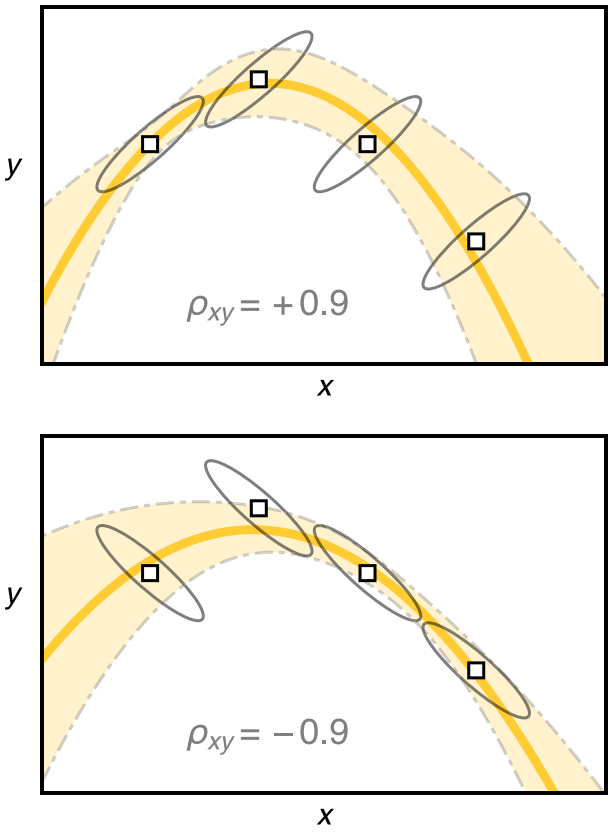


Fig. 5. Influence of covariance between x and y within each observation. Square markers correspond to (x_i, y_i) observations, with slanted 95% confidence ellipses indicating positive or negative correlation ρ_{xy} between x_i and y_i. All other covariance terms $\omega_{x_i y_j}$ $(i \neq j)$ are zero. Changing the sign of ρ_{xy} displaces the best-fit OGLS regression curve (yellow line) and has a strong influence on its 95% confidence band. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

x_i and y_i may be correlated (e.g., Fig. 1). These correlations not only affect the best-fit values of regression parameters, but may also strongly influence the final model uncertainties, as illustrated in Fig. 5. In geochronology, properly accounting for this type of covariance is a crucial requirement for many dating methods [e.g., Vermeesch, 2018].

Another situation is that when x or y values belonging to different observations (e.g., x_i and y_j, with $i \neq j$) are not statistically independent. This situation arises naturally when the uncertainty introduced by our correction/standardization procedures are large enough compared to other sources of analytical errors [e.g. Daëron, 2021]. Once again, these correlations may strongly affect regression results and model uncertainties. Fig. 6 illustrates, for example, how the standard error for the slope of a straight-line regression constrained by three data points depends quite strongly on the sign and magnitude of the correlation between the y values.

OGLS is able to properly propagate these different kinds of covariance, with a small number of caveats to consider. In the following section, we briefly discuss these limitations and how to mitigate them.

3.2.3. Caveats and best practices

3.2.3.1. Limitations of a first-order linear error propagation. The contributions of x uncertainties to the residuals covariance $\mathbf{V}_{r_y}$ depend critically on the *local* values of $\partial f/\partial x$. Fig. 7 illustrates how this may be problematic in some cases: in that example, one of the observations (x_3, y_3) has a large x uncertainty. A truly optimal result would account for the fact that the 95% confidence region for (x_3, y_3) is entirely

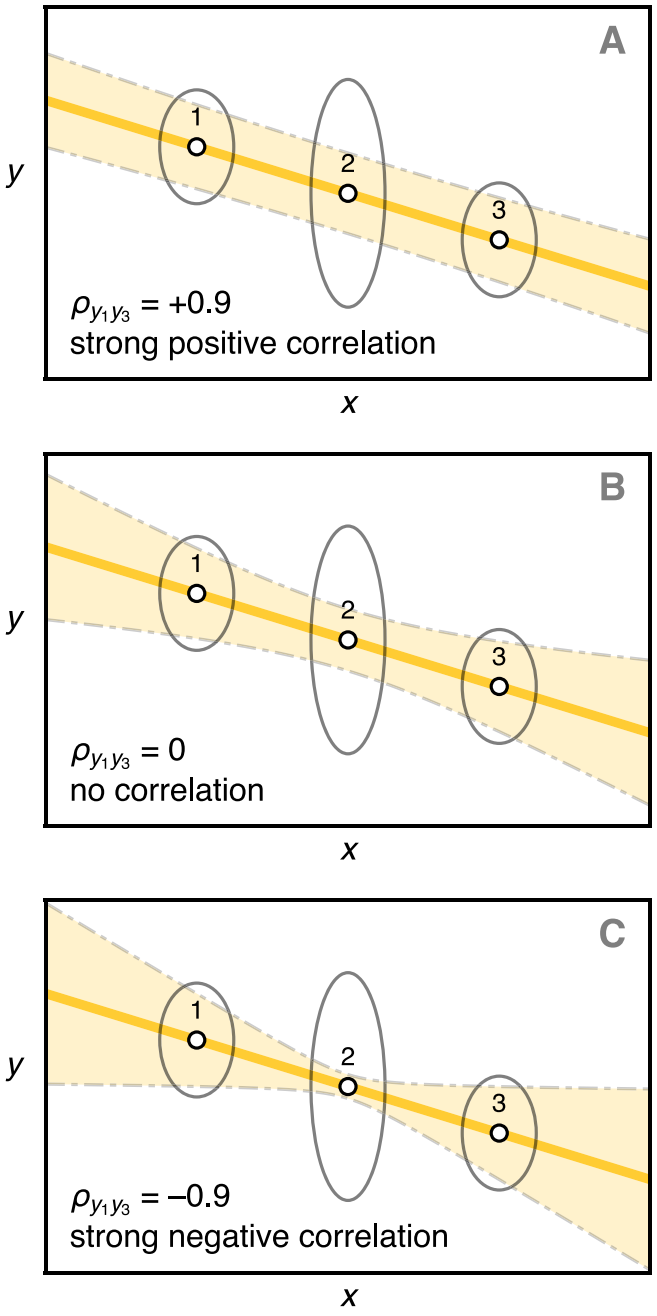


Fig. 6. Influence of y covariance across observations. These three (x, y) observations have the same covariance structure in the three panels, except for $\omega_{y_1 y_3}$, with $\rho_{y_1 y_3}$ varying from -0.9 to $+0.9$, exerting a strong influence on the standard error of the regression slope.

consistent with the true function f_{true} (dashed line in Fig. 7), so that the influence of that observation on the best-fit function should ideally be negligible. Yet, in this case, the best-fit OGLS model (yellow line in Fig. 7) deviates strongly from the true function. This is because, over the range of likely x_3 values, the slope of f_{true} varies substantially and $\partial f_{\text{true}}/\partial x(x_3)$ happens to be zero. As a result, as the regression approaches an optimal version of f, the contribution of the uncertainty in x_3 tends to zero, so that the fitting procedure only takes the y uncertainty into account for that observation.

In practice, this issue only matters in the case of large relative variations in $\partial f/\partial x$ at a scale of $\pm\sigma_x$. When in doubt, one may check whether the OGLS approximation still yields valid results by using Monte Carlo methods to randomly resample the observations and investigate how the

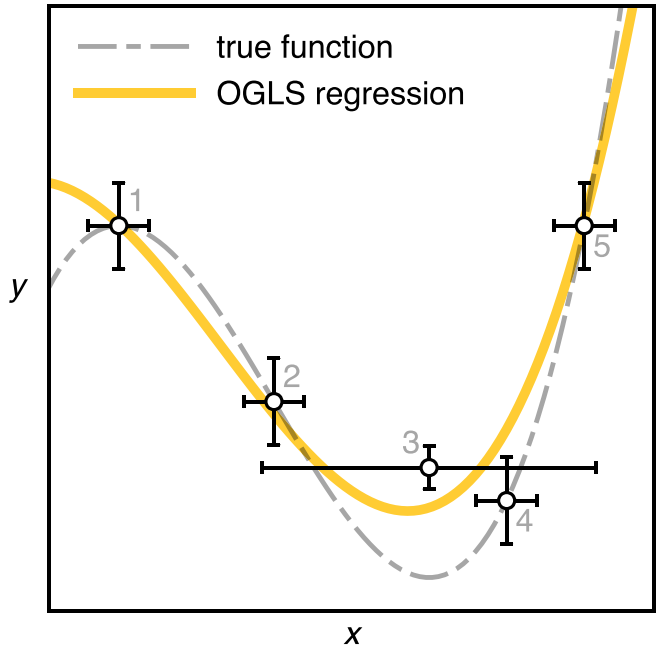


Fig. 7. Limitations of a first-order linear error propagation. In this synthetic example, all of the observations (white circles with 95% error bars) lie on the true curve (dashed gray line), except for observation number 3, whose very large x uncertainty imply that it does not diverge significantly from the true curve. One would thus expect that the best least-squares polynomial fit for these data would be very close to the true curve. This is not the case here, because OGLS scales the influence of x errors by the *local* value of $\partial f/\partial x$ (cf Section 2.2.3), so that as the optimization procedure approaches the true solution, the influence of these x errors for observation number 3 tends to zero, shifing the OGLS best-fit curve (yellow line) away from the truly optimal solution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

best-fit parameters respond to these perturbations.

3.2.3.2. Assignment of response and predictor variables. In some cases, such as fitting of a straight line, there is no natural choice for which quantity should be treated as a response (y) or a predictor (x) variable. Because OGLS accounts for x uncertainties using a first-order linear approximation, it is generally safer to assign the status of response variable to the quantity with larger observation uncertainties. When x and y do not have the same dimension (e.g, distance vs time), comparing their respective uncertainties is still possible by scaling x errors according to the expected magnitude of $\partial f/\partial x$ [Tellinghuisen, 2020].

It also bears pointing out that, although OGLS regression yields equivalent results whether x or y observations are treated as the predictor/response variable *when fitting a straight line*, this is not generally the case for models where $\partial f/\partial x$ is not constant. This is a direct consequence of estimating $\partial f/\partial x$ — and thus weighing the relative contributions of $\mathbf{V_x}$, $\mathbf{V_y}$ and $\mathbf{V_{yx}}$ — at fixed values of the predictor variable.

3.2.3.3. Convergence issues. OGLS regression, like most numerical optimization problems, requires some degree of care to ensure convergence, i.e. to find the vector of model parameters $\left(p_1, p_2 \ldots p_{N_p}\right)$ minimizing $\chi^2 = \mathbf{r_y}^\top \cdot \mathbf{V_{r_y}^{-1}} \cdot \mathbf{r_y}$. Convergence is facilitated when the initial choice of parameters is reasonably close to the optimal vector. A generally good choice of initial parameters can be obtained from one of the approaches described in Section 2.3. Although it has been pointed out that the iterative algorithms used by some effective variance methods may, in some cases, fail to reach an optimal solution when the initial model parameters are those computed from OLS [Chandler, 1972; Lybanon, 1984], this issue does not affect gradient descent algorithms such as the classical Levenberg-Marquardt method [Levenberg, 1944; Marquardt, 1963].

3.2.4. Standard errors of the best-fit regression parameters

Most software implementations of nonlinear optimization also provide estimates of the covariance matrix and/or joint confidence regions of the best-fit model parameters, usually by computing the inverse Hessian matrix of χ^2, or by explicitly mapping χ^2 over the model parameter space. For strongly nonlinear problems, a proper characterization of the joint probability distribution of the best-fit parameters may require some version of a Monte Carlo simulation. In many cases, however, the regression problem is linear enough that the vector of best-fit parameters $\mathbf{p}^* = [p_i^*]$ may be treated as having a multivariate Gaussian distribution caracterized by its covariance matrix $\mathbf{V_{p^*}}$. In that case, the usual propagation rules may be used to estimate, for an arbitrary value of x, the standard error of the model prediction:

$$\sigma^2_{\text{model}}\left(f\left(x, p_1^*, p_2^* \ldots\right)\right) = \mathbf{J}_{f|\mathbf{p}}^\top \cdot \mathbf{V_{p^*}} \cdot \mathbf{J}_{f|\mathbf{p}}$$
$$\mathbf{J}_{f|\mathbf{p}} = \left[\frac{\partial f}{\partial p_i}\left(x, p_1^*, p_2^* \ldots\right)\right] \tag{17}$$

For example, for a best-fit model of the form $f(x) = ax + b$:

$$\sigma^2_{\text{model}}(ax+b) = \begin{pmatrix} x & 1 \end{pmatrix} \cdot \begin{pmatrix} \sigma_a^2 & \omega_{ab} \\ \omega_{ab} & \sigma_b^2 \end{pmatrix} \cdot \begin{pmatrix} x \\ 1 \end{pmatrix} = x^2\sigma_a^2 + 2x\omega_{ab} + \sigma_b^2 \tag{18}$$

3.2.5. Goodness-of-fit and Cholesky residuals

Many tests for goodness-of-fit provide estimates of the likelihood ("p-values") that an observed sample (in the statistical sense, i.e. a set of values) was drawn from a single assumed underlying probability distribution, such as a Gaussian distribution whose mean and variance were postulated a priori. In the context of least-squares regression, a common objective is to establishes whether the distribution of best-fit residuals $\mathbf{r_y^*}$ differs from an expected N-dimensional Gaussian distribution. Even in the relatively simple case of WLS (i.e. only considering uncorrelated uncertainties on y_i values), this requires transforming $\mathbf{r_y^*}$ so that the expected variance in each dimension is one. In the context of WLS, this is done by computing the *weighted residuals* (or "weighted deviates", or "Z-scores") defined as:

$$z_i = \frac{y_i - f\left(x_i, p_1^*, p_2^*, \ldots\right)}{\sigma_{y_i}} \tag{19}$$

If the observation uncertainties and the model are both accurate, the z_i values are expected to be normally distributed around zero with a variance of one, and we can test this prediction in various ways.

One common approach is to apply Pearson's chi-squared test, by computing again the χ^2 statistic, here equal to $\sum z_i^2$. Under the assumption of z_i normality, $\sum z_i^2$ follows a χ^2 distribution with $N_\text{f} = \left(N - N_p\right)$ degrees of freedom, with an expected value equal to N_f. If $\sum z_i^2$ is much greater that N_f, with a corresponding p-value below an arbitrary threshold, the observed residuals are said to be over-dispersed relative to the assumed uncertainties. Conversely, under-dispersed residuals correspond to the case when $\sum z_i^2$ is much less than N_f. This test is often formulated in terms of the reduced χ^2 or MSWD, defined as χ^2/N_f, whose expected value is one. But however useful the reduced χ^2 statistic may be to characterize under/over-dispersion, what should be done based on this evidence depends entirely on the scientific context (cf Vermeesch, 2018 and Section 4.3 below).

Pearson's chi-squared test applies equally well to OGLS, which simply redefines the covariance matrix used to compute χ^2 from $\mathbf{r_y}$, as to other flavors of least-squares regression. Things are not as simple when it comes to other tests of normality, including the popular Kolmogorov-

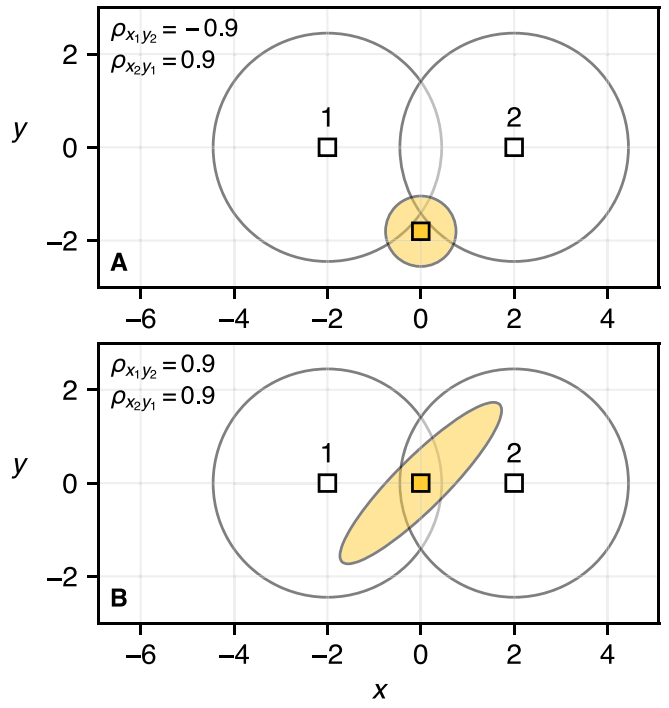


Fig. 8. Two counter-intuitive examples of weighted averages using OGLS. White squares represent the two observations, encircled by their respective 95% confidence ellipse. Yellow square is the corresponding OGLS-weighted average, with associated 95% confidence ellipse shaded in yellow. Inter-sample correlations between x and y errors are listed in upper left corners but not shown graphically. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Smirnoff or Shapiro–Wilk tests, which consider the whole N-dimensional vector $\mathbf{r}_\mathbf{y}^*$ but require that its elements are statistically independent from each other. In such cases, we can go back to the definition of OGLS in terms of a Mahalanobis distance: the best-fiting model being that which minimizes $\mathbf{r}_\mathbf{y}^\top \cdot \mathbf{V}_{\mathbf{r}_\mathbf{y}}^{-1} \cdot \mathbf{r}_\mathbf{y}$, we can rewrite $\mathbf{V}_{\mathbf{r}_\mathbf{y}}^{-1}$, using Cholesky decomposition, as the product of an upper triangular matrix $\mathbf{U}$ and its transpose:

$$\mathbf{V}_{\mathbf{r}_\mathbf{y}}^{-1} = \mathbf{U}^\top \cdot \mathbf{U} \quad \Rightarrow \quad \chi^2 = \mathbf{r}_\mathbf{y}^\top \cdot \mathbf{V}_{\mathbf{r}_\mathbf{y}}^{-1} \cdot \mathbf{r}_\mathbf{y} = \left(\mathbf{U} \cdot \mathbf{r}_\mathbf{y}\right)^\top \cdot \left(\mathbf{U} \cdot \mathbf{r}_\mathbf{y}\right) \tag{20}$$

This leads naturally to considering the *Cholesky residuals* $\mathbf{U} \cdot \mathbf{r}_\mathbf{y}$, noted $[\zeta_i]$ and analogous to the z_i values defined above. Conversion between $[r_i]$ and $[\zeta_i]$ is a bijective (i.e. reversible) linear transformation, and the linear transformation of a multivariate normal variable is also a multivariate normal variable. Thus $[r_i]$ is normally distributed if and only if $[\zeta_i]$ is as well. More specifically, the transformation defined by $\mathbf{U}$ corresponds to a N-dimensional rotation and rescaling which, applied to $\mathbf{r}_\mathbf{y}$ (or to any other vector whose statistical covariance matrix is $\mathbf{V}_{\mathbf{r}_\mathbf{y}}$), yields transformed residuals with *independent* distributions of variance equal to one [Houseman et al., 2004]. Applying classical tests of normality to $[\zeta_i]$ thus provides us with a simple way to test the normality of $[r_i]$, to assess the optimal model's goodness-of-fit, and to quantify the under/over-dispersion of our observations.

3.3. Extension of OGLS to weighted averages

The ideas described above can be naturally extended to computing weighted averages of (x, y) observations, taking into account their full covariance structure. In practice, given observations $(\mathbf{x}, \mathbf{y})$, one should minimize the Mahalanobis length of the corresponding residuals, with $(\bar{x}, \bar{y})$, defined as the weighted average point, being the parameter vector to optimize. Note that the covariance matrix used to define the Mahalanobis length to minimize is the covariance of the residuals, which is identical to the covariance the observations themselves:

$$\chi^2 = \begin{bmatrix} \mathbf{y} - \bar{y} \\ \mathbf{x} - \bar{x} \end{bmatrix}^\top \cdot \begin{bmatrix} \mathbf{V}_\mathbf{y} & \mathbf{V}_\mathbf{yx} \\ \mathbf{V}_\mathbf{yx}^\top & \mathbf{V}_\mathbf{x} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{y} - \bar{y} \\ \mathbf{x} - \bar{x} \end{bmatrix} \tag{21}$$

Fig. 8 illustrates two non-intuitive situations where taking into account inter-sample correlations between x and y uncertainties (off-diagonal terms in $\mathbf{V}_\mathbf{yx}$), with notable effects on the best-fit location $(\bar{x}, \bar{y})$ of the weighted average (Fig. 8A) and/or on the size and orientation of the corresponding confidence ellipse (Fig. 8B).

A slightly different situation, already described by Vermeesch [2015] in the case of a single sample, occurs when considering measurements of the same quantity, noted x, in different samples noted $a, b, c, \ldots$, with each sample analyzed an arbitrary number of times. In that case, the vector of observations is defined as:

$$\mathbf{x} = \begin{bmatrix} x_{a_1} & x_{a_2} & \ldots & x_{a_{N_a}} & x_{b_1} & x_{b_2} & \ldots & x_{b_{N_b}} & \ldots \end{bmatrix}^\top \tag{22}$$

The OGLS-weighted sample averages, noted $\bar{\mathbf{x}} = (\bar{x}_a, \bar{x}_b, \ldots)$, is the vector which, again, minimizes the χ^2 of the residuals for each sample:

$$\chi^2 = \begin{bmatrix} x_{a_1} - \bar{x}_a \\ x_{a_2} - \bar{x}_a \\ \vdots \\ x_{b_1} - \bar{x}_b \\ x_{b_2} - \bar{x}_b \\ \vdots \end{bmatrix}^\top \cdot V_x^{-1} \cdot \begin{bmatrix} x_{a_1} - \bar{x}_a \\ x_{a_2} - \bar{x}_a \\ \vdots \\ x_{b_1} - \bar{x}_b \\ x_{b_2} - \bar{x}_b \\ \vdots \end{bmatrix} \quad \text{with } V_x \text{ being the covariance matrix of } x \tag{23}$$

As before, the optimal vector $\bar{\mathbf{x}}$ may be found using numerical methods, usually also yielding an estimate of the covariance matrix of the best-fit $\bar{\mathbf{x}}$, thus fully characterizing inter-sample error correlations. This approach is useful when computing sample mean values for measurements known to have correlated errors (e.g., as is the case for Δ_{47} measurements, see Daëron, 2021 and Section 5 below).

3.4. Open-source Python implementation (ogls)

The `ogls` Python module provides an efficient implementation of OGLS regression based on the methods discussed above and released under a MIT license. Regression models may be specified in a very general way, by providing functional definitions for f and its partial derivatives relative to x and $[p_i]$. The module offers a simpler API to work with polynomial functions of x with arbitrarily configurable degrees (e.g., $f(x) = ax^5 + bx^2 + c$), or with polynomial functions of inverse absolute temperature (e.g., $f(T) = a/T^2 + b$). It also implements the two weighted average methods described in Section 2.3. Under the hood, χ^2 minimization uses a Trust Region Reflective algorithm as implemented by `scipy.optimize.least_squares`, but alternative methods may be used instead. The module also provides built-in methods to access various best-fit statistics such as the covariance matrix of fit parameters, the OGLS χ^2 value, or the best-fit Cholesky residuals, as well as common plotting functions for regression data, error bars, confidence regions, and best-fit function. Documentation and development branches of the ogls module are currently hosted at https://github.com/mdaeron/ogls. The source code for each release is archived on Zenodo (https://doi.org/10.5281/zenodo.8357230).

4. Application to geochronology

In the following section, the straight line case of OGLS regression — which is mathematically equivalent to WTLS regression — is applied to the problem of isochron fitting. Isochrons are a cornerstone of high precision geochronology. They are used to determine the radiogenic isotope ratio in the presence of an inherited (nonradiogenic) component, by analyzing multiple aliquots from the same sample. Isochrons come in

Table 4
Examples of two-dimensional isochron fits, using the generic formula $\mathbf{y} = a + b\mathbf{x}$. J is a calibration constant, λ_* is the decay constant of the parent nuclide, i.e., ^{40}K, ^{235}U or ^{238}U. $(Z)_i$ is the inherited component of Z, and t is geologic time.

isochron type	**x**	**y**	a	b
conventional ^{40}Ar/^{39}Ar	$\frac{^{39}\mathrm{Ar}}{^{36}\mathrm{Ar}}$	$\frac{^{40}\mathrm{Ar}}{^{36}\mathrm{Ar}}$	$\left(\frac{^{40}\mathrm{Ar}}{^{36}\mathrm{Ar}}\right)_i$	$J(exp[\lambda_{40}t] - 1)$
inverse ^{40}Ar/^{39}Ar	$\frac{^{39}\mathrm{Ar}}{^{40}\mathrm{Ar}}$	$\frac{^{36}\mathrm{Ar}}{^{40}\mathrm{Ar}}$	$\left(\frac{^{36}\mathrm{Ar}}{^{40}\mathrm{Ar}}\right)_i$	$J(1 - exp[\lambda_{40}t])\left(\frac{^{36}\mathrm{Ar}}{^{40}\mathrm{Ar}}\right)_i$
conventional ^{207}Pb/^{206}Pb	$\frac{^{206}\mathrm{Pb}}{^{204}\mathrm{Pb}}$	$\frac{^{207}\mathrm{Pb}}{^{204}\mathrm{Pb}}$	$\left(\frac{^{207}\mathrm{Pb}}{^{204}\mathrm{Pb}}\right)_i$	$\left(\frac{^{235}\mathrm{U}}{^{238}\mathrm{U}}\right)\frac{exp[\lambda_{235}t] - 1}{exp[\lambda_{238}t] - 1}$
inverse ^{207}Pb/^{206}Pb	$\frac{^{204}\mathrm{Pb}}{^{206}\mathrm{Pb}}$	$\frac{^{207}\mathrm{Pb}}{^{206}\mathrm{Pb}}$	$\left(\frac{^{235}\mathrm{U}}{^{238}\mathrm{U}}\right)\frac{exp[\lambda_{235}t] - 1}{exp[\lambda_{238}t] - 1}$	$\left(\frac{^{207}\mathrm{Pb}}{^{206}\mathrm{Pb}}\right)_i$

several forms. See Table 4 for some examples.

Conventional isochron regression is done using the York et al. [2004] algorithm (see Sections 1.3.2 and 2.2.1 above). In fact, York regression was developed for this very purpose [York, 1969]. However, the following sections will show that York regression is unable to handle the complex covariance structures that underly many geochronological data sets.

4.1. Random and systematic uncertainty in geochronology

Geochronological uncertainties can be divided into two main categories. Random errors, also known as internal errors, arise from factors such as electronic noise in ion detectors and counting statistics. These uncertainties can be quantified by investigating the statistical properties of replicate measurements in time-resolved mass spectra. Analytical precision is assessed using the standard error of these measurements. Increasing the number of measurements leads to a decrease in the standard error, enhancing precision.

In contrast, systematic errors, which are also known as external errors, stem from uncertainties linked to assumptions made during the calculation of isotopic ratios from the raw mass spectrometer data. These encompass effects like the uncertainty of decay constants (λ in Table 4), the isotopic composition of reference materials, laboratory blanks, calibration factors (J in Table 4) and so forth. Unlike random uncertainties, systematic uncertainties cannot be characterized by repeated measurements and are not reducible by increasing measurement frequency.

In many cases, the distinction between random and systematic uncertainties is clear, and they can be treated separately using a so-called "hierarchical error propagation" approach [Renne et al., 1998]. Consider the example of ^{40}Ar/^{39}Ar isochron regression, in which all the aliquots in the isochron are associated with the same calibration constant J. Then the uncertainty of the isochron age t is obtained in two steps. First, a straight line regression is made between the **x** and **y** data, using the random uncertainties only, and ignoring the systematic uncertainty associated with J. Second, the uncertainty of the J-factor is added to the estimated uncertainty of the slope b in quadrature. This then allows the uncertainty of the isochoron age t to be estimated by standard error propagation.

Unfortunately, in some cases the distinction between random and systematic uncertainties is not clear, and the hierarchical approach does not work. Revisiting the ^{40}Ar/^{39}Ar example, consider a hypothetical dataset combining aliquots from two neutron irradiations, corresponding to two J-factors. The hierarchical error propagation approach is unable to deal with this situation, which requires the full covariance matrix of uncertainties to be taken into account. Such "hybrid uncertainties" are pervasive throughout geochronology. Examples include:

1. Blank correction of ^{40}Ar/^{39}Ar or ^{207}Pb/^{206}Pb data, in which multiple aliquots share the same blank measurements [Vermeesch, 2015; Connelly et al., 2021].
2. Sample-standard bracketing in U-Pb geochronology by laser ablation inductively coupled plasma mass spectrometry, in which spline interpolation of laboratory drift creates complex inter-sample error correlations [McLean et al., 2011].
3. Construction of a Pb/U vs UO/U calibration curve in U-Pb geochronology by secondary ion mass spectrometry, whereby the Pb/U-ratio uncertainties of a sample's aliquots vary depending on their UO/U-ratios relative to those of the calibration data [Vermeesch, 2022].

Geochronologists have only recently started capturing these inter-sample error correlations. However, the statistical tools to analyse them have not yet been developed. The application of OGLS to isochron regression is an important first step in this direction.

4.2. An ^{40}Ar/^{39}Ar example

Let us consider the inverse isochron ratios of the ^{40}Ar/^{39}Ar system [Vermeesch, 2015]:

$$\left[\frac{36}{40}\right] = \frac{A - B - C + D - E}{1 - F + G} \text{ and } \left[\frac{39}{40}\right] = \frac{H - I}{1 - F + G} \tag{24}$$

where '36', '39' and '40' stand for ^{36}Ar, ^{39}Ar and ^{40}Ar, respectively, whilst

$$A = \left[\frac{36}{40}\right]_m, B = \left[\frac{37}{40}\right]_m \left[\frac{36}{37}\right]_{ca}, C = \left[\frac{38}{40}\right]_m \left[\frac{36}{38}\right]_{cl}, D = \left[\frac{39}{40}\right]_m \left[\frac{38}{39}\right]_k \left[\frac{36}{38}\right]_{cl},$$
$$E = \left[\frac{37}{40}\right]_m \left[\frac{39}{37}\right]_{ca} \left[\frac{38}{39}\right]_k \left[\frac{36}{38}\right]_{cl}, F = \left[\frac{39}{40}\right]_m \left[\frac{40}{39}\right]_k,$$
$$G = \left[\frac{37}{40}\right]_m \left[\frac{39}{37}\right]_{ca} \left[\frac{40}{39}\right]_k, H = \left[\frac{39}{40}\right]_m \text{ and } I = \left[\frac{37}{40}\right]_m \left[\frac{39}{37}\right]_{ca};$$

in which m stands for the 'measured ratio' of the sample, ca and k for the isotopic composition of co-irradiated Ca- and K-salts (to quantify the interferences on ^{36}Ar, ^{39}Ar and 40Ar), and cl stands for the reactor-specific interferences from Cl (on ^{36}Ar and ^{38}Ar). The complexity of Eq. (24) introduces hybrid uncertainties that cannot be separated into random and sytematic components.

Fig. 9 shows an example of such a complex ^{40}Ar/^{39}Ar-dataset, using 12 sets of argon measurements on Quaternary sanidine, acquired by a Nu Instruments Noblesse mass spectrometer at the University of Wisconsin (courtesy of Dr. Allen Schaen). Using OGLS instead of York regression changes the slope and intercept by 1% and 0.7%, respectively, whilst improving their precision by a third. The accuracy and precision of the corresponding isochron age changes accordingly, from 1.030 ± 0.14 to 1.033 ± 0.10 Ma.

4.3. Dealing with overdispersion

The degree to which the analytical uncertainties account for the observed dispersion of the isotopic ratio measurements around the best

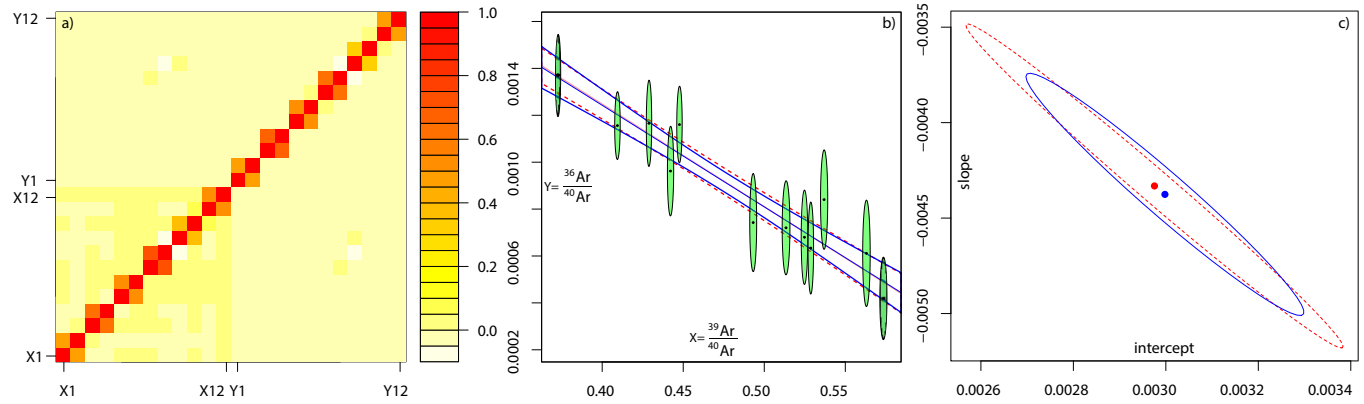


Fig. 9. a) Correlation matrix of 12 pairs of inverse $^{40}Ar/^{39}Ar$ isochron ratios; b) inverse isochron regression using the York algorithm (dashed red) and OGLS algorithms (solid blue); c) 95% confidence ellipses of the best-fit parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fit line can be quantified with the χ^2 statistic, and formally assessed with a chi-square test. If the p-value of this test is less than some predefined significance cutoff (typically $\alpha = 0.05$) or, equivalently, if the MSWD is greater than $1 + 2\sqrt{2/N_f}$, then the data are said to be 'overdispersed' with respect to the analytical uncertainties, meaning that the goodness-of-fit is poor and that the parameteric assumptions of the OGLS are no longer valid.

One way to deal with overdispersed datasets is to inflate the analytical uncertainties by a *factor* of $\sqrt{\text{MSWD}}$. This removes the statistical problem but does not address the root cause of the overdispersion. A scientifically more useful solution is to attribute the excess dispersion to one of the fit parameters. Thus, in the case of isochron regression, the overdispersion can be attributed to excess scatter of the intercept (a) or slope (b). This excess scatter has scientific meaning. Overdispersion of the intercept can be attributed to heterogeneity of the inherited component. Overdispersion of the slope, on the other hand, can be attributed to diachronous isotopic closure [Rioux et al., 2012].

Overdispersion can be formally quantified using the the method of maximum likelihood instead of the (equivalent) least squares approach. To estimate the overdispersion of the intercept (as a standard deviation σ_a), maximize the goodness-of-fit parameter L_a, where:

$$-2{\cdot}L_a = ln\left|\begin{matrix}\mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}^\top} & \mathbf{V_x} + \mathbf{I}_N\sigma_a^2\end{matrix}\right| + \begin{bmatrix}\mathbf{r_x} \\ \mathbf{r_y}\end{bmatrix}^T \begin{bmatrix}\mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}^\top} & \mathbf{V_x} + \mathbf{I}_N\sigma_a^2\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{r_x} \\ \mathbf{r_y}\end{bmatrix} \quad (25)$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Eq. (25) can be efficiently maximized using an iterative two step process:

1. For any given value of σ_a, find the optimal values of a and b by applying ordinary OGLS regression to the second term of (25), ignoring the first term. Then plug these optimal values of a and b into (25) to calculate the corresponding log-likelihood.
2. Explore the parameter space of σ_a until the log-likelihood is maximized.

The overdispersion of the slope (σ_b) can be estimated by maximizing an alternative log-likelihood function, L_b, where:

$$-2{\cdot}L_b = ln\left|\begin{matrix}\mathbf{V_y} & \mathbf{V_{yx}} \\ \mathbf{V_{yx}^\top} & \mathbf{V_x} + [\mathbf{diag}(\boldsymbol{\xi})\boldsymbol{\sigma}_\mathbf{b}]^2\end{matrix}\right| + \begin{bmatrix}\mathbf{r_x} \\ \mathbf{r_y}\end{bmatrix}^T \begin{bmatrix}\mathbf{V_y} & \\ \mathbf{V_{yx}^T} & V_\mathbf{x} + [\mathbf{diag}(\boldsymbol{\xi})\boldsymbol{\sigma}_\mathbf{b}]^2\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{r_x} \\ \mathbf{r_y}\end{bmatrix} \quad (26)$$

in which $diag(\boldsymbol{\xi})$ is a diagonal matrix with the true (but unknown) values of the independent variable (where $\boldsymbol{\xi} = [\xi_i]$, see Section 2.3.3). Eq. (26) has $N + 3$ unknowns (a, b, σ_b and the elements of $\boldsymbol{\xi}$). Unfortunately, unlike the σ_a estimation, it is not possible to divide the parameter estimation into two clean steps, so computationally more intensive numerical methods are required.

4.4. Implementation in IsoplotR

`IsoplotR` is an `R` package for radiometric geochronology [Vermeesch, 2018]. It includes functions for $^{40}Ar/^{39}Ar$, U-Pb and many other chronometers, as well as generic datasets. WTLS/OGLS was added as a bivariate linear regression option in version 5.4 of the package, which is available online from the Comprehensive R Archive Network (CRAN, R Core Team, 2013) at https://cran.r-project.org/package=IsoplotR. The easiest way to get started with IsoplotR is through the graphical user interface (GUI), which is available online at several mirror sites (e.g., https://isoplotr.es.ucl.ac.uk).

The online GUI consists of various pull-down menus. The OGLS functionality can be accessed by selecting `other` and `regression` from the first two of these menus, and selecting the third input format in the Options menu. This format consists of a matrix with $2N$ rows and $2N + 1$ columns, in which the first column contains the X-values on top of the Y-values, and the subsequent $2N$ columns contain the full covariance matrix of those X and Y-values. Clicking the PLOT button at the bottom of the GUI triggers the OGLS regression. In the presence of excess dispersion, the overdispersion parameter σ_a can be estimated by selecting 'Model-3' regression in the Options menu. Ordinary OGLS regression corresponds to 'Model-1', whereas 'Model-2' regression corresponds to orthogonal least squares regression, which ignores the analytical uncertainties.

The following code snippet provides a self-contained example of OGLS regression from the command-line, using the $^{40}Ar/^{39}Ar$ example of Section 4.2. The file path to this dataset, which is included with `IsoplotR`, can be retrieved by `R`'s `system.file()` function:

```
library(IsoplotR)
fn <- system.file("UW137.csv",package="IsoplotR")
dat <- read.data(fn,method="other",format=6)
fit <- ogls(dat) # numerical calculations
isochron(dat) # graphical results
```

To estimate the overdispersion parameter σ_a:

```
fit <- ogls(dat,random.effects=TRUE)
isochron(dat,model=3)
```

Generating large covariance matrices such as the one stored in `UW137.csv` requires a new generation of upstream data processing

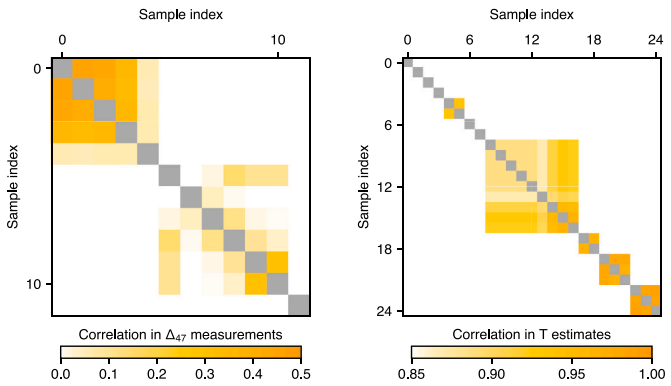


Fig. 10. Example of error correlations in Δ_{47} calibration data. Left panel: Example of error correlations between Δ_{47} measurements. Gray squares correspond to $\rho = 1$. Raw data from Jautzy et al. [2020], reprocessing details in Appendix A. Right panel: Example of error correlations between temperature estimates. Correlation in the uncertainties for calcification temperatures of foraminifera from various core-tops, derived from $\delta^{18}O$ thermometry. Samples from the same core top likely experienced similar seawater compositions, so that the error contribution from seawater $\delta^{18}O$ is common to all samples from that core top. Raw data from Peral et al. [2018], reprocessing details in Appendix A. Gray squares correspond to $\rho = 1$.

software such as `Ar-Ar_Redux` [Vermeesch, 2015] and `simplex` [Vermeesch, 2022].

5. Application to Δ_{47} calibrations

In this section, OGLS regression is used to investigate whether different calibrations of carbonate "clumped isotopes" (Δ_{47}) paleothermometry agree with each other within analytical uncertainties. For context, Δ_{47} is a quantity used by stable isotope geochemists to characterize the statistical excess of $^{16}O^{13}C^{18}O$ isotopologues in CO_2 [Eiler and Schauble, 2004; Eiler, 2007]. Because $^{13}C - ^{18}O$ bonds in carbonate minerals are thermodynamically favored at low crystallization temperatures, precise measurements of Δ_{47} may be used to constrain the formation conditions of well-preserved carbonates up to hundreds of million years old [Schauble et al., 2006; Ghosh et al., 2006; Eiler, 2011].

Although certain types of carbonates, such as stalagmites or corals, are known to present large departures from equilibrium Δ_{47} values [e.g., Daëron et al., 2011; Saenger et al., 2012], several research groups over the years have independently observed that many other carbonates, both biogenic and abiotic, appear to follow the same relationship between Δ_{47} and formation temperature despite being formed under very diverse chemical conditions. Early calibration studies aiming to constrain this equilibrium relationship, however, yielded large discrepancies between research groups, with differences up to ±10 °C at ambient temperatures [cf Fig. 4 of Spencer and Kim, 2015]. Since then, improved data processing methods have greatly reduced these inter-laboratory discrepancies but not entirely eliminated them [Petersen et al., 2019]. More recently, the clumped-isotope community collectively defined a new metrological scale anchored to carbonate reference materials, the *InterCarb Carbon Dioxide Equilibrium Scale* (I-CDES), which solves long-standing inter-laboratory standardization issues [Bernasconi et al., 2021]. Published shortly afterwards, two independent calibration studies based on re-analyzed and newly-analyzed carbonates of various types appear to yield consistent I-CDES Δ_{47} values as a function of temperature [Anderson et al., 2021; Fiebig et al., 2021].

The issue we aim to address here is whether reprocessing previously published Δ_{47} calibration studies in the I-CDES reference scale yields statistically consistent calibration relationships, or, in other words, whether the degree of scatter observed in a large Δ_{47} data set is commensurate (neither smaller nor greater) than predicted based on the analytical and environmental uncertainties. This problem plays to the strengths of OGLS regression, because calibration studies must take into account analytical uncertainties in Δ_{47} measurements as well as uncertainties on independently constrained crystallization temperatures, both of which may be rather large and potentially correlated across observations.

5.1. Covariance in Δ_{47} measurements and temperature estimates

Compared to many other stable isotope tracers, carbonate clumped isotopes stand out in that analytical uncertainties remain large relative to the range of Δ_{47} values typical of natural samples. What's more, a sizable component of measurement error arises from standardization corrections designed to compensate for various sources of analytical nonlinearity, and the resulting errors are shared by samples analyzed together. As a result, Δ_{47} measurements from a single study are not statistically independent but correlated (most often positively, cf. Fig. 10), to an extent that depends on the temporal distribution of analyses, on the ratio of standards to unknown samples and on their relative compositions [Daëron, 2021]. Computing the full covariance matrix for a set of Δ_{47} measurements requires access to the "raw" (uncorrected) analytical data, using for instance the `D47crunch` library for Python.

Regarding temperature estimates, two different kinds of covariance are to be considered. For one thing, several calibration samples considered below (natural inorganic calcites from Devils Hole and Laghetto Basso, and eight samples purposefully re-equilibrated at high temperatures) were analyzed independently by different groups [Jautzy et al., 2020; Anderson et al., 2021; Fiebig et al., 2021]. Although their Δ_{47} values measured in different studies are statisticaly independent, neither their true formation/equilibration temperatures nor their assigned temperature estimates vary between studies. Because of this, the errors in the assigned formation temperature for a given sample analyzed by different groups are all equal, with correlation coefficients equal to one.

Another case to consider is that of planktic foraminifera recovered from sedimentary core tops [e.g., Peral et al., 2018; Meinicke et al., 2020], whose calcification temperatures were estimated from oxygen-18 thermometry using seawater $\delta^{18}O$ values from a global gridded data set. Because samples collected from the same core top presumably experienced similar water compositions, the error contribution from seawater $\delta^{18}O$ is shared by all samples from that core top, resulting in strongly positive correlation coefficients (Fig. 10).

5.2. Compilation and reprocessing of published calibration data

In the first calibration study processed in the I-CDES framework, Anderson et al. [2021] compared their own results with those of previous studies, independently reprocessed to the I-CDES scale, and concluded that the corresponding regression lines were in close agreement. Shortly thereafter, Fiebig et al. [2021] reported new measurements of inorganic calcites presumed to have achieved clumped-isotope equilibrium, with state-of-the-art analytical precision and covering a

Table 5
Δ_{47} calibration data sets reprocessed and combined in this study.

Label	Original publication	Laboratory	Sample Type	N	T (°C)
[B18]	Breitenbach et al. [2018]	Univ. Cambridge	Cave pearls	6	3–47
[P18]	Peral et al. [2018]	LSCE	Planktic foraminifera	25	3–23
[J20]	Jautzy et al. [2020]	Geol. Surv. Canada	Synthetic Heated	11 1	5–250 727
[A21·MIT]	Anderson et al. [2021]	MIT	Synthetic Tufa, Travertines Lacustrine (Proto-)dolomite Heated	17 12 6 4 2	6–80 5–95 0–4 80–350 1100
[A21·LSCE]	Anderson et al. [2021]	LSCE	Slow-growing calcite	2	8–34
[F21]	Fiebig et al. [2021]	Goethe-Univ.	Synthetic Slow-growing calcite Heated	5 3 3	120–250 8–34 727–1100
[H22]	Huyghe et al. [2022]	LSCE	Marine bivalves	7	−2–27
[OGLS23]	this study	all of the above	all of the above	104	−2–1100

Table 6
Δ_{47} model selection: The performances of various polynomial model functions of inverse absolute temperature were compared by applying OGLS regression to the combined data set `[OGLS23]` before computing the root mean squared weighted deviation (RMSWD), the Bayes information criterion (BIC) and the Akaike information criterion (AIC) for each model. Both the BIC and the AIC can be used as a benchmark of model performance by considering the trade-off between goodness-of-fit and model simplicity, with lower BIC or AIC values implying better models [Stoica and Selen, 2004]. The results confirm that a model with degrees $(0, 2)$ performs much better than with degrees $(0, 1)$, which is hardly unexpected. Adding a first-degree (a_1) or a third-degree (a_3) term further decreases BIC and AIC and slightly decreases RMSWD. Based on both information criteria, the best model is that with degrees $(0, 1, 2)$.

Degrees	Model	RMSWD	BIC	AIC
$(0, 1)$	$\Delta_{47} = a_0 + a_1/T$	3.41	262.7	257.4
$(0, 2)$	$\Delta_{47} = a_0 + a_2/T^2$	0.96	− 0.5	− 5.8
$(0, 1, 2)$	$\Delta_{47} = a_0 + a_1/T + a_2/T^2$	0.93	− 3.7	− 11.6
$(0, 2, 3)$	$\Delta_{47} = a_0 + a_2/T^2 + a_3/T^3$	0.94	− 1.6	− 9.5
$(0, 1, 2, 3)$	$\Delta_{47} = a_0 + a_1/T + a_2/T^2 + a_3/T^3$	0.93	− 0.5	− 11.1

wide range of temperatures (8–1100 °C). Among other findings, they also concluded that their equilibrium Δ_{47} calibration was in good agreement with Anderson et al. [2021] and some of the older reprocessed studies. These results bode well for the hypothesis that the I-CDES scale may fully reconcile the calibration data from different groups, and OGLS offers an opportunity to test this hypothesis quantitatively, based on a large reprocessed data set and accounting for error covariance in Δ_{47} and in temperature.

Calibration data to include here were selected based on the following criteria. First and foremost, proper reprocessing requires access to the original analytical data, and the I-CDES conversion also requires that these studies include measurements of several carbonate standards whose I-CDES values are well known, as opposed to relying on equilibrated CO_2 standards. Furthermore, estimating Δ_{47} covariance is done here using a "pooled regression" approach as implemented by the `D47crunch` library [Daëron, 2021], using the I-CDES nominal values assigned to the ETH-1/2/3/4 standards, from Figs. 2 and 4 of Bernasconi et al. [2021] for ETH-1/2/3 and ETH-4, respectively. In a few studies initially considered, Δ_{47} measurements were standardized based on reference materials analyzed in a moving time window rather than grouping analyses in discrete analytical sessions [e.g., Meinicke et al., 2020]. Although this approach is entirely valid in itself, the statistical treatment implemented in `D47crunch` does not properly apply in the case of a moving window, and to the best of our knowledge there is no published method to reliably propagate full standardization uncertainties for that approach. Moving-window studies were thus excluded, except for that of Jautzy et al. [2020], which provides crucial high-temperature constraints and whose results are not particularly sensitive to the use of discrete versus moving-window sessions.

In the end, the reprocessed data sets meeting all of the above criteria are those of Breitenbach et al. [2018], Peral et al. [2018], Jautzy et al. [2020], Anderson et al. [2021], Fiebig et al. [2021], and Huyghe et al. [2022]. The combined data set is summarized in Table 5 and comprises 104 sample measurements based on ∼ 2000 unknown and ∼ 3000 standard analyses performed by five different research groups, with formation temperatures ranging from −2 °C to 1100 °C, including synthetic carbonates, experimentally heated samples, very slow-growing natural calcites, travertines, tufa, cave pearls, foraminifera, marine bivalves, glacial lake carbonates, and (proto-)dolomite precipitates. Reprocessing details for each data set are provided in appendix, and the complete reprocessing code and data are included in the source code of the `D47calib` library (Section 5.4).

5.3. Reconciled calibrations of carbonate Δ_{47}

5.3.1. Choice of model function and best-fit model for the combined data set

The combined data set ([OGLS23] in Table 5) was subjected to OGLS regression, modeling Δ_{47} as second-degree polynomial of inverse absolute temperature:

$$\Delta_{47} = a_0 + a_1/T + a_2/T^2 \quad (27)$$

Alternative polynomial functions have been used, historically, with the most frequent choice being an equation of the form $a_0 + a_2/T^2$. Another popular option is to use a polynomial of degree greater than two, introducing additional terms such as a_3/T^3 and a_4/T^4. After comparing the outcomes of different models, we conclude that Eq. (27) qualifies as the best polynomial model function based on the Bayes and Akaike information criteria, two different benchmarks often used for

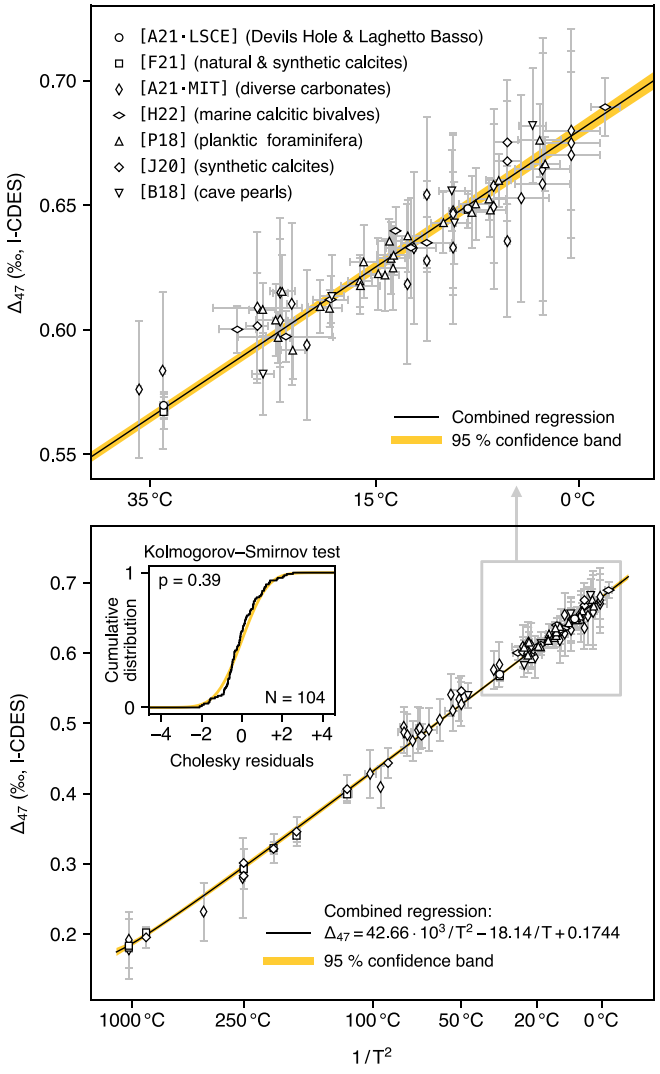


Fig. 11. Combined Δ_{47} calibration plot. A second-degree polynomial regression yields a good fit to the [OGLS23] data set combining all data from the prior studies listed in Table 5, with Cholesky residuals statistically indistinguishable from the expected Gaussian distribution (Section 5.3.2).

model selection (Table 6).

The second-degree model function yielding the optimal fit for the combined data set is:

$$\Delta_{47} = 0.1744 - \frac{18.14}{T} + \frac{42.66 \times 10^3}{T^2} \quad \text{(OGLS23 calibration)} \quad (28)$$

In practice, to compute T from Δ_{47}, one may invert (28) algebraically (appendix B), but it is also possible to do so numerically, for example using the D47calib library Section 5.4 which also propagates calibration and/or measurement uncertainties into the final temperature estimate. The combined calibration is shown in Fig. 11 along with all measurements constraining it, and its standard model errors and confidence limits are reported in appendix B.

5.3.2. *Overall goodness of fit and statistical consistency between individual data sets*

As a first test, we may check how well the combined regression function agrees with the confidence regions for each of the independent data sets. Eq. (28) remains well within the intersection of the 95% confidence limits for all individual studies, from 0 °C to 1200 °C (cf fig-S1.pdf in code repository). We may also assess how well the individual regressions agree at low temperature, by plotting the corresponding

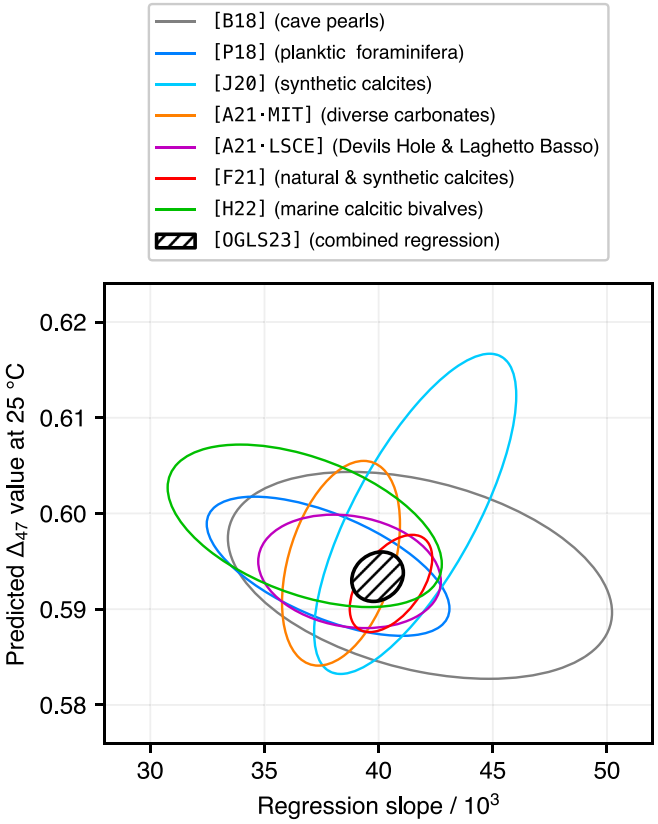


Fig. 12. Comparison of the Δ_{47} value and sensitivity to T at 25 °C for the different calibrations. For historical reasons, the regression slope shown here is defined as the local derivative of Δ_{47} with respect to $1/T^2$. All ellipses represent 95% confidence regions. See Appendix A for reprocessing details.

joint confidence ellipses for the Δ_{47} value predicted at 25 °C and the local calibration slope. As shown in Fig. 12, all of these 95% confidence ellipses overlap.

Considering now the regression residuals, the combined calibration has a RMSWD of 0.93. This value lies within the 95% confidence interval (0.86–1.14) for the RMSWD of a χ^2 distribution with 101 degrees of freedom, suggesting that the uncertainties assigned to our observations are not grossly inaccurate. As described in Section 3.2.5, we may also call upon a Kolmogorov-Smirnoff test to determine whether the Cholesky residuals of our best-fit model are unlikely to be drawn independently from a normal distribution with a variance of one. The resulting p-value of 0.39 (Fig. 11) implies that there is no compelling evidence that the distribution of our fit residuals differs significantly from that expected from the covariance of our observations, further strengthening the case that Eq. (28) provides a good fit to the combined data set.

Leaving aside this overall goodness of fit, however, we may still ask whether each of the individual data sets is in good statistical agreement with the combined calibration. To do so, we start by computing, for each of the partial data sets listed in Table 5, the Δ_{47} residuals relative to Eq. (28). We then convert these Δ_{47} residuals to Cholesky residuals based on the covariance matrix of the observations for that data set, and check, using a Kolmogorov-Smirnoff test of normality, whether their observed distributions are likely to deviate from the expected (normal) distribution. As shown in Fig. 13, none of the data sets considered here have Cholesky residuals deviating significantly from a zero-centered normal distribution with unit variance, implying that each study's results are neither biased nor under/over-scattered compared to the dispersion predicted from their respective environmental and analytical uncertainties.

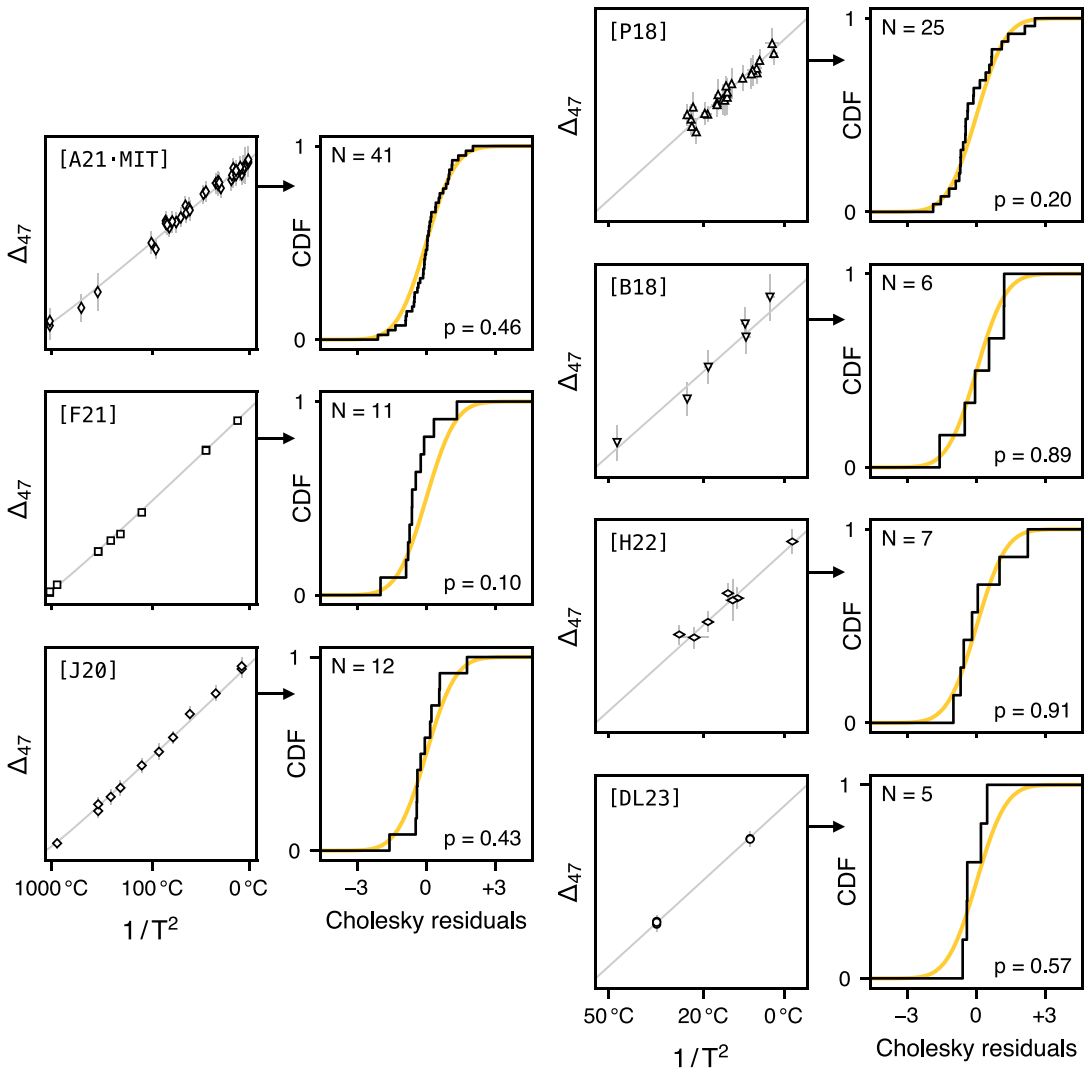


Fig. 13. Kolmogorov-Smirnoff tests of normality for the Cholesky residuals of each individual calibration relative to the *combined* regression `[OGLS23]`. Left-hand plots show the data points, with 95% error bars, for each data set, along with the combined regression line. Right-hand plots show the cumulative distribution of Cholesky residuals, with the number of observations and *p*-value for a Kolmogorov-Smirnoff test of normality with mean of zero and variance of one. All data sets have a distribution of Cholesky residuals which is statistically indistinguishable from the expected Gaussian distribution (Section 5.3.2). See Appendix A for reprocessing details.

5.3.3. *Comparison with theoretical models*

The combined Δ_{47} calibration may also be compared with theoretical predictions of the Δ_{63} clumped-isotope anomaly in calcite [Hill et al., 2014] and of isotopic fractionations associated with the conversion of $CaCO_3$ to CO_2 by phosphoric acid reaction [Guo et al., 2009]. As shown in Fig. 14, the observed Δ_{47} values increase linearly with Δ_{63} predictions based on the known formation/equilibration temperatures, with a slope slightly greater than one (1.043 ± 0.011, 95% confidence). This slope is consistent with the original prediction of Guo et al. [2009] that the net effect (Δ^*_{47}) of the acid reaction should increase as a function of (equilibrium) Δ_{63} with a sensitivity of 0.035. Based on the high-temperature calcites in the combined data set, Δ^*_{47} for an acid reaction at 90 °C (since the I-CDES scale is by convention defined as Δ_{47} values corresponding to a 90 °C reaction) is on the order of 0.184 ± 0.004 ‰ (95% confidence). Because the high-temperature Δ_{47} constraints in the combined data set are primarily from measurements performed by Fiebig et al. [2021], this value is virtually identical to theirs.

5.3.4. *Implications*

The statistical consistency among I-CDES calibration studies and their agreement with theoretical predictions support a conceptually simple worldview where:

(a) the calibration materials compiled here share a common relationship linking their Δ_{47} value to their crystallization or equilibration temperature, most likely reflecting thermodynamic equilibrium between carbonate isotopologues;
(b) this equilibrium relationship is closely approximated by Eq. (28), with a precision better than ±1 °C (95% CL) for temperatures below 50 °C (Fig. 15);
(c) Δ_{47} measurements performed in different laboratories, when (re) processed in the I-CDES metrological scale, are directly comparable with each other, without introducing additional uncertainties beyond those explicitly accounted for using current analytical error models. This point was previously established by Bernasconi et al. [2021] based on the InterCarb results. The findings reported here further strengthen their conclusions, based on an equally large ($N \approx 5000$) yet completely independent data set.
(d) I-CDES calibration observations are quantitatively consistent with the combined theoretical models of Hill et al. [2014], who

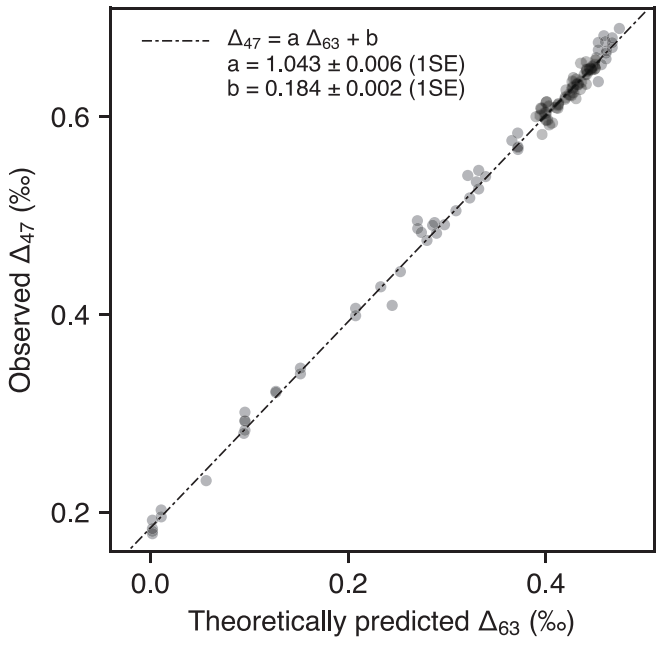


Fig. 14. Comparison between Δ_{47} measurements and theoretical calcite Δ_{63} values for the whole [OGLS23] data set. Theoretical Δ_{63} is from Hill et al. [2014]. The observed slope is statistically consistent with that predicted based on the acid fractionation model of Guo et al. [2009] (cf Section 5.3.3).

predicted how Δ_{63} should vary with temperature in carbonate minerals, and of Guo et al. [2009], who predicted how Δ^*_{47} should increase with (equilibrium) Δ_{63}.

These claims are not new, having been scrutinized for well over a decade. Yet it is worth pointing out that each of the statements above is now unambiguously backed up by strong evidence, thanks to the persistent, concerted efforts of the clumped-isotope community. This is a milestone for Δ_{47} thermometry, not because the framework summarized above is perfect and universally applicable, but because we are now much better-equipped to investigate where and how the statements above cease to apply or require qualification.

For one thing, we are in fact well aware that not all carbonates follow an equilibrium Δ_{47} calibration. Being able to precisely compare $\Delta_{47\ (\text{I-CDES})}$ measurements from any laboratory to the equilibrium calibration reported here — or to future, better versions of it — improves our ability to quantify Δ_{47} disequilibrium in natural and synthetic carbonates. This, along with the progress of Δ_{48} metrology, will in turn greatly improve our ability to determine which carbonates have equilibrium clumped-isotope signatures and which do not, but also to study quantitatively how Δ_{47} disequilibria vary with other parameters beyond temperature, and how they co-vary with other isotopic and elemental tracers, with obvious applications to the study of inorganic and biologic processes [Guo, 2020; Uchikawa et al., 2021; Watkins and Devriendt, 2022]. This prediction should be tempered by a reminder that independent $\Delta_{47(\text{I-CDES})}$ measurements are expected to be consistent *within analytical uncertainties*. Meaningful inter-laboratory comparisons still depend critically on the accuracy of our analytical error estimates, which should, as pointed out in previous studies, account for the (correlated) errors introduced by correction/standardization procedures [Kocken et al., 2019; Daëron, 2021; Bernasconi et al., 2021].

Furthermore, the combined data set considered here is almost exclusively based on calcite samples. According to the theoretical models of Hill et al. [2014], different carbonate mineralogies should have slightly different equilibrium Δ_{47} values at a given temperature. These differences remain small, on the order of 10 ppm, at ambient temperatures, but should be detectable based on current instrumental limits and the newly improved constraints on equilibrium values of calcite $\Delta_{47\ (\text{I-CDES})}$. An additional concern is that phosphoric acid reactions at different temperatures may fractionate different mineralogies in different ways, yielding different results $\Delta_{47\ (\text{I-CDES})}$ when aragonite or dolomite samples are standardized using the calcite reference materials underpinning the I-CDES scale. It is likely that this will ultimately require defining new reference materials with mineralogies other than calcite [Anderson et al., 2023].

Finally, we now have very tight contraints, consistent across reseach groups, on $\Delta_{47(\text{I-CDES})}$ values for slow-growing inorganic calcites from Devils Hole (∼ 34 °C) and Laghetto Basso (∼ 8 °C) on one hand [Daëron et al., 2019; Anderson et al., 2021; Fiebig et al., 2021], and for biogenic or abiotic calcites charactarized by strikingly different mineralization

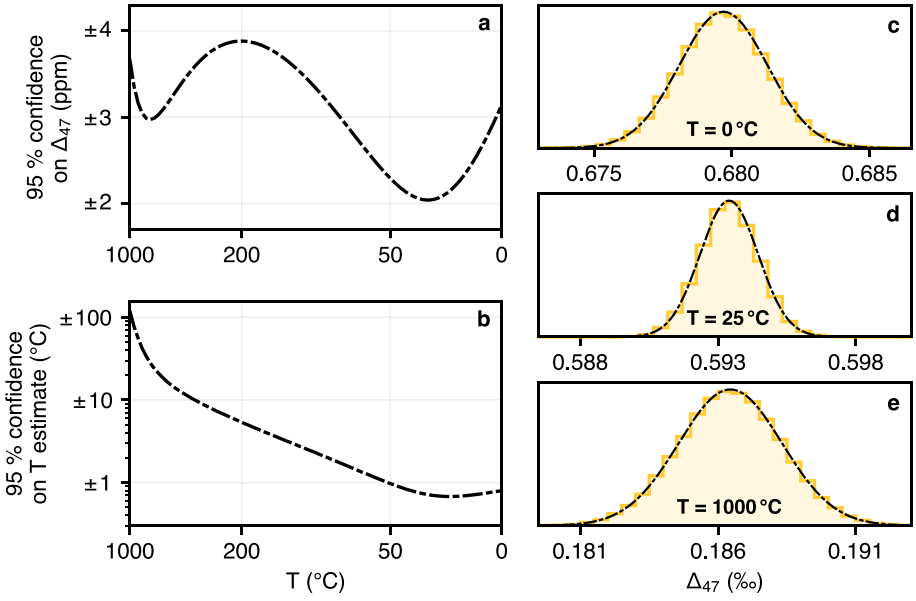


Fig. 15. Regression confidence limits on the [OGLS23] calibration. Left panels: confidence limits on Δ_{47} and T estimates as a function of T (x axis is scaled as $1/T^2$). Right panels: Quasi-Monte Carlo simulations of Δ_{47} uncertainties at three different values of T, confirming the applicability of our first-order propagation methods. See Appendix B.2 for computation details.

conditions [e.g., Daëron and Gray, 2023, and references therein for planktic foraminifera]. The fact that these different types of calcite have virtually identical Δ_{47} signatures, despite clear differences in their apparent fractionation oxygen-18 factors relative to their parent waters, and despite the fact that their mineralization conditions differ in many important ways (pH, Mg/Ca, mineralogies, overall cristallization rates, chemical flux rates, activities of organic and inorganic catalysts…), is an observation which should tell us something fundamental about how exactly a carbonate mineral inherits its clumped-isotope composition from those of the DIC species it precipitated from. It may be useful here to consider two end-member "toy models". According to the first one, the net effect of all fractionations, due to chemistry and/or transport, between the mineral phase and the DIC species preserves clumping to the first order, either because of quantitative precipitation of a DIC pool [e.g., Tripati et al., 2015], or because fractionation factors are inherently "stochastic-like" [cf Watkins and Hunt, 2015]. In this case, the fact that planktic foraminiferal tests, for instance, have equilibrium Δ_{47} values whereas coral carbonate does not may be used to constrain DIC compositions in the calcification micro-environment. According to another end-member model, heterogenous oxygen-18 equilibrium (between water and DIC as well as between DIC and the precipitating mineral) may be mechanistically decorrelated to some extent from single-phase clumped-isotope equilibrium, for example through rapid breaking and reforming of C-O bonds at the crystal growth surface, and/or during the transition from amorphous to crystalline phases. This might explain why some rapidly forming biocarbonates, whose DIC-water equilibration reactions should be slow due to high presumed calcification pH, may still display equilibrium Δ_{47} signatures. Both of these models are naive and obviously lack a proper mechanistic grounding, but are meant to encourage future process/model studies to take advantage of our newfound ability to pinpoint equilibrium $\Delta_{47(\text{I-CDES})}$ values with unprecendented precision.

5.4. Open-source Python implementation (D47calib)

The `D47calib` Python module provides a general framework for processing Δ_{47} calibration data sets using OGLS regression, e.g., defining new calibrations from (T, Δ_{47}) data; combining two or more data sets with or without samples in common; simple plotting of observations and best-fit regression line along with the corresponding confidence limits as a function of (inverse) temperature. `D47calib` also offers a simple function `T47()` to compute T values from Δ_{47} measurements and vice versa, along with fully propagated standard errors (optionally including calibration uncertainties, measurement uncertainties, or both), also potentially computing the the full covariance matrix of T estimates based on the covariance matrix of a Δ_{47} vector, with or without accounting for calibration uncertainties. All of the calibration data sets of Table 5 are already defined in the module, so that applying Eqs. (28) and (38) only requires a call to `D47calib.OGLS23.T47()`. Documentation and development branches of the D47calib module are currently hosted at https://github.com/mdaeron/D47calib, and the source code for each release is and will be archived on Zenodo (https://doi.org/10.5281/zenodo.8357232).

6. Conclusions

Based on the overview of existing least-squares methods, each with their own strengths and simplifications, it should be evident that Omnivariant Generalized Least Squares can be viewed as a missing link between Weighted Total Least Squares, whose high dimensionality of model parameter space makes for challenging computational problems when dealing with hundreds of observations or more, and the complementary approaches of Generalized Least Squares, which accounts for general covariance in **y** and ignores errors in **x**, and effective variance methods including York's straight line regression, which treat (x, y) observations as statistically independent from each other. Our aim here has been to describe the background, core principle, and the practical uses of OGLS in a widely accessible manner, so that researchers in diverse fields may readily apply and/or extend this approach.

As stated in introduction, regression analysis stems from purpose. For the purpose of geochronology, new software is needed to take full advantage of the improved precision and accuracy offered by the OGLS method. Currently, most low-level software that is used to process geochronological data ignores inter-sample error correlations. Only a handful of new and little-used packages capture these rich covariance structures [e.g., Vermeesch, 2015; McLean et al., 2016; Vermeesch, 2022]. Those tools need to become more widespread and linked together with higher order software such as `IsoplotR` [Vermeesch, 2018] in order to create an internally consistent ecosystem of geochronological software with OGLS regression at its heart.

In the context of clumped isotope geochemistry, it is now clear that the definition of the I-CDES metrological scale, having virtually eliminated inter-laboratory discrepancies [Bernasconi et al., 2021], also solves our long-standing (calcite) Δ_{47} calibration problem. Future studies will doubtlessly refine, qualify, extend, and perhaps challenge this finding. However, we now have constraints on equilibrium Δ_{47} values which are accurate and precise enough for the vast majority of known applications, including virtually all paleo-environmental reconstructions.

CRediT authorship contribution statement

M. Daëron: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **P. Vermeesch:** Conceptualization, Methodology, Software, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data and code used in this study are available at https://github.com/mdaeron/ogls-paper-code and archived at https://doi.org/10.5281/zenodo.10418915.

Acknowledgements

We are grateful to Ilja Kocken and a second, anonymous reviewer for their thoughtful comments, and to C. France-Lanord's editorial handling of this article. MD would also like to thank J. Jautzy for checking the reprocessed results of the [J20] data set. PV's research was partly funded by NERC standard grant #NE/T001518/1.

Appendix A. Reprocessing Δ_{47} calibrations in the I-CDES reference frame

Note: The complete reprocessing code and data described below are included in the source code of the `D47calib` library, and the corresponding calibration plots are shown in Fig. 16. Data sets including samples equilibrated at $T > 700$ °C were fit as quadratic functions of the form $a/T^2 + b/T + c$, while those whose formation temperatures did not exceed 50 °C were fit as $a/T^2 + c$.

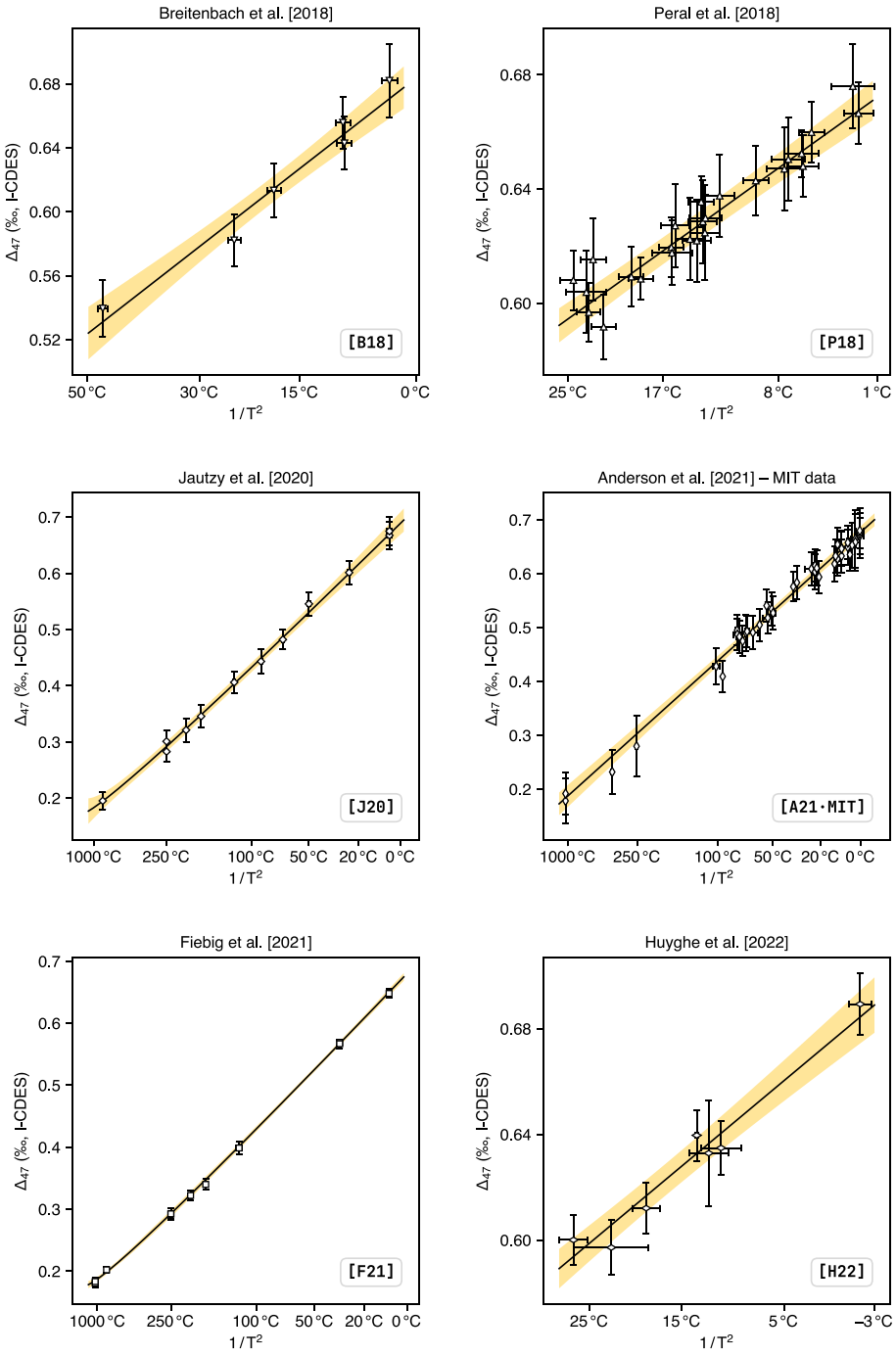


Fig. 16. Regression plots for the individual calibration data sets. Error bars are 95% confidence limits. Yellow shaded regions are 95% confidence bands of the best-fit functions. See Appendix A for reprocessing details. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[B18] Cave pearls [Breitenbach et al., 2018]

Raw data were obtained from the original study's supplementary information. The original publication processed data according to two sessions, each 4–5 months long, separated by 2 months. After reprocessing the original raw data using `D47crunch`, visual inspection of the standardization residuals defined revealed the presence of substantial drifts in both sessions. We thus assigned modified session boundaries defining four continuous measurement periods separated by 21 to 52 days, with new session lengths ranging from 24 to 80 days. The original data was not modified in any other way. Formation temperatures are from Table 1 of the original study. We assigned arbitrary 95% uncertainties of ±1°C, which seem reasonable for cave environments. Combining these temperature constraints with the `D47crunch`-reprocessed Δ_{47} values yields the following OGLS best-fit regression:

$$\Delta_{47} = 41.81 \times 10^3 / T^2 + 0.123 \qquad (\texttt{D47calib.breitenbach\_2018}) \tag{29}$$

[P18] Planktic foraminifera [Peral et al., 2018]

Peral et al. [2018] reported Δ_{47} values of foraminifera from core-tops, both planktic and benthic, whose calcification temperature estimates were recently reassessed by Daëron and Gray [2023]. Here we only consider *Peral* et al.'s planktic data, excluding two benthic samples (cf *Daëron & Gray* for reasons why we only consider planktic samples for now). In our reprocessing, as in the original study, "samples" were defined by default as a unique combination of core site, species, and size fraction. Δ_{47} values were then standardized in the usual way, before using `D47crunch`'s built-in `combine_samples()` method to combine all size fractions with the same core and species, except for *G. inflata* samples (cf *Daëron & Gray* and accompanying GitHub repository). By properly accounting for analytical error covariance between the Δ_{47} values to combine, this two-step approach avoids underestimating the final standardization errors. This yields the following OGLS best-fit regression:

$$\Delta_{47} = 37.79 \times 10^3 / T^2 + 0.169 \qquad (\texttt{D47calib.peral\_2018}) \tag{30}$$

[J20] Synthetic and heated calcites [Jautzy et al., 2020]

Jautzy et al. [2020] reported data from a continuous period spanning 10 months, and used a moving-window approach to standardize their measurements. We assigned sessions defined, whenever possible, as periods of one or more complete weeks enclosing one of more unknown sample analyses. The resulting Δ_{47} residuals, on the order of 40 ppm (1SD), do not display evidence of instrumental drift. Formation temperatures are from table S2 of the original study. We assigned arbitrary 95% uncertainties of ± 1°C, which seem reasonable for laboratory experiments. Combining these temperature constraints with the `D47crunch`-reprocessed Δ_{47} values yields the following OGLS best-fit regression:

$$\Delta_{47} = 45.44 \times 10^3 / T^2 - 25.78 / T + 0.175 \qquad (\texttt{D47calib.jautzy\_2020}) \tag{31}$$

[A21·MIT] Natural and synthetic carbonates analyzed at MIT [Anderson et al., 2021]

Raw IRMS data and temperature constraints were obtained from the original study's supplementary information (tables S01 and S02). When reprocessing the IRMS data we made no changes to the session defintions, but we excluded sessions 5 and 25 because they did not include any unknown sample analyses. This yields the following OGLS best-fit regression:

$$\Delta_{47} = 38.35 \times 10^3 / T^2 + 0.163 \qquad (\texttt{D47calib.anderson\_2021\_mit}) \tag{32}$$

[A21·LSCE] Slow-growing calcites analyzed at LSCE [Anderson et al., 2021]

Raw IRMS data were obtained from the original study's supplementary information (SI-S02). Temperature contraints are from Table 1 in Daëron et al. [2019]. This yields the following OGLS best-fit regression:

$$\Delta_{47} = 38.72 \times 10^3 / T^2 + 0.158 \qquad (\texttt{D47calib.anderson\_2021\_lsce}) \tag{33}$$

[F21] Synthetic, heated and slow-growing calcites [Fiebig et al., 2021]

Temperature contraints are duplicated from the earlier publications where the corresponding samples were first described [Daëron et al., 2019; Jautzy et al., 2020; Anderson et al., 2021]. Raw IRMS data and were obtained from the original study's supplementary information, and processed as described by Fiebig et al. [2021], jointly using (a) heated and 25 °C -equilibrated CO_2 to constrain the scrambling effect and compositional nonlinearity associated with each session, and (b) ETH-1 and ETH-2 reference materials to anchor unknown samples to the I-CDES scale. This data processing strategy was designed to overcome compositional drift in the batch of ETH-3 used in some of *Fiebig* et al.'s analytical sessions, and is validated by the fact that the resulting "pseudo-I-CDES" Δ_{47} values for slow-growing calcite from Devils Hole (DVH-2; DHC2–8) and Laghetto Basso (LGB-2) agree almost perfectly with the I-CDES values obtained by Anderson et al. [2021] in [A21·LSCE]. This yields the following OGLS best-fit regression:

$$\Delta_{47} = 44.43 \times 10^3 / T^2 - 26.08 / T + 0.180 \qquad (\texttt{D47calib.fiebig\_2021}) \tag{34}$$

[H22] Marine bivalves [Huyghe et al., 2022]

Huyghe et al. [2022] reported Δ_{47} values of modern calcitic bivalves collected from localities with good environmental constraints. As was done in the original publication, different bivalve individuals were initially treated as distinct analytical samples. In some sites with strong seasonality, individuals were sub-sampled into winter-calcified a summer-calcified fractions. Δ_{47} values were then standardized in the usual way, before using `D47crunch`'s built-in `combine_samples()` method to combine all samples from the same locality. Calcification temperature estimates are from the original study. This yields the following OGLS best-fit regression:

$$\Delta_{47} = 36.76 \times 10^3 / T^2 + 0.185 \qquad (\texttt{D47calib.huyghe\_2022}) \tag{35}$$

[DL23] Devils Laghetto

Finally, one may wish to combine the measurements of Devils Hole and Laghetto Basso calcite reported by Anderson et al. [2021] and Fiebig et al. [2021], for a total of 76 replicates with an external Δ_{47} repeatability of 0.009 ‰. These independent measurements yield statistically indistinguishable values (RMSE = 2.6 ppm at the sample level), yielding the following OGLS best-fit regression (see Fig. 17):

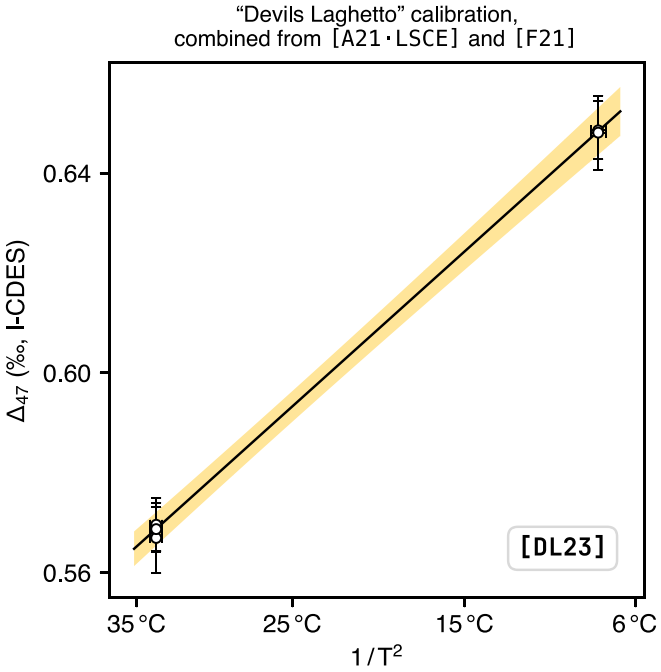


Fig. 17. Regression plot for the "Devils Laghetto" calibration data, based on slow-growing calcites from Devils Hole and Laghetto Basso. Error bars are 95% confidence limits. Yellow shaded regions are 95% confidence bands of the best-fit functions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\Delta_{47} = 39.04 \times 10^3 / T^2 + 0.154 \qquad (\texttt{D47calib.devils\_laghetto\_2023}) \tag{36}$$

Note that is equation differs slightly from that reported by Daëron and Gray [2023]. This is because they used York regression, ignoring correlations in Δ_{47} measurement uncertainties.

Appendix B. Algebraic inversion and confidence limits of the combined Δ_{47} calibration

B.1. Inversion

Inverting $\Delta 47 = a_0 + a_1/T + a_2/T^2$ yields:

$$\frac{1}{T} = \frac{\sqrt{a_1^2 - 4a_2(a_0 - \Delta_{47})} - a_1}{2a_2} \quad \Leftrightarrow \quad \frac{1}{T} = \sqrt{\frac{\Delta_{47}}{a_2} + \left(\frac{a_1}{2a_2}\right)^2 - \frac{a_0}{a_2}} - \frac{a_1}{2a_2} \tag{37}$$

In the case of the combined regression of Eq. (28), this results in:

$$\mathrm{T} = \frac{10^3}{0.21265 + \sqrt{23.4427 \cdot \Delta_{47} - 4.0427}} \tag{38}$$

B.2. Standard model errors and confidence limits of the combined regression

As seen in Section 2.2.4, the model standard error for Eq. (28) may be computed from the covariance matrix of the best-fit model parameters and the Jacobian of the model function:

$$\sigma^2_{\text{model}}\left(a_0+\frac{a_1}{T}+\frac{a_2}{T^2}\right) = \mathbf{J}_{f|\mathbf{p}}^{\top}\cdot\mathbf{C}_{\mathbf{p}}\cdot\mathbf{J}_{f|\mathbf{p}}$$

$$\text{with}: \quad \mathbf{J}_{f|\mathbf{p}} = \begin{pmatrix} 1 & -1/T^2 & -2/T^3 \end{pmatrix}^{\mathrm{T}}$$

$$\text{and}: \quad \mathbf{C}_{\mathbf{p}} = \begin{pmatrix} \sigma^2_{a_0} & \omega_{a_0a_1} & \omega_{a_0a_2} \\ \omega_{a_0a_1} & \sigma^2_{a_1} & \omega_{a_1a_2} \\ \omega_{a_0a_2} & \omega_{a_1a_2} & \sigma^2_{a_2} \end{pmatrix} \tag{39}$$

The resulting confidence limits on Δ_{47} values and the corresponding temperature estimates are shown in Fig. 15. Although it is assumed that the probability distribution of best-fit regression parameters is a multivariate normal distribution, as discussed in Section 3.2.3 this is only a first-order approximation. We may test the validity of that assumption by using a Quasi-Monte Carlo approach [Roy et al., 2023] to generate 2^{16} slightly different versions of the 104 (T, Δ_{47}) pairs in the combined calibration data set, distributed according to their full covariance matrix. Each of these 2^{16} versions of the data set is subjected to an OGLS regression yielding 2^{16} versions of the best-fit parameters (a_0, a_1, a_2). As shown in Fig. 15, the distribution of Δ_{47} values computed from these Quasi-Monte Carlo outputs is visually indistinguishable, for any given temperature, from the normal distribution of Δ_{47} derived from our first-order linear approximation.

B.3. Confidence limits and covariance of Δ_{47}-derived temperatures estimates

When analyzing a group of N samples, with $\mathbf{y} = [y_i]$ being the vector of measured Δ_{47} values and the corresponding temperature estimates being noted $\mathbf{T} = [T_i]$:

$$\begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_N \end{pmatrix} = \begin{pmatrix} g(y_1) \\ g(y_2) \\ \vdots \\ g(y_N) \end{pmatrix} \quad \text{with } g \text{ being for example } \textit{defined} \text{ as in (38)} \tag{40}$$

The covariance matrix of $\mathbf{T}$, noted $\mathbf{C}_{\mathbf{T}}$, may then be derived from $\mathbf{J}_{\mathbf{T}}$, the Jacobian matrix of $\mathbf{T}$ with respect to $\mathbf{y}$ and $\mathbf{C}_{\mathbf{y}}$, the covariance matrix of $\mathbf{y}$, itself estimated using for instance the `D47crunch` library [Daëron, 2021]:

$$\mathbf{C}_{\mathbf{T}} = \mathbf{J}_{\mathbf{T}|\mathbf{y}}\cdot\mathbf{C}_{\mathbf{y}}\cdot\mathbf{J}_{\mathbf{T}|\mathbf{y}}{}^{\top} \quad \text{with}: \quad \boldsymbol{J}_{\mathbf{T}|\mathbf{y}} = \begin{pmatrix} \frac{\partial g}{\partial y}(y_1) & 0 & \cdots & 0 \\ 0 & \frac{\partial g}{\partial y}(y_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\partial g}{\partial y}(y_N) \end{pmatrix} \tag{41}$$

Note that the formulation above neglects the model uncertainties computed in (18). To include them, we may use the complete covariance and Jacobian matrices:

$$\mathbf{C}_{\mathbf{T}} = [\,\mathbf{J}_{\mathbf{T}|\mathbf{p}} \quad \mathbf{J}_{\mathbf{T}|\mathbf{y}}\,]\cdot\begin{bmatrix} \mathbf{C}_{\mathbf{p}} & 0 \\ 0 & \mathbf{C}_{\mathbf{y}} \end{bmatrix}\cdot[\,\mathbf{J}_{\mathbf{T}|\mathbf{p}} \quad \mathbf{J}_{\mathbf{T}|\mathbf{y}}\,]^{\top}$$

$$\text{with}: \quad \boldsymbol{J}_{\mathbf{T}|\mathbf{p}} = \begin{pmatrix} 1 & -1/T_1^2 & -2/T_1^3 \\ 1 & -1/T_2^2 & -2/T_2^3 \\ \vdots & \vdots & \vdots \\ 1 & -1/T_N^2 & -2/T_N^3 \end{pmatrix} \tag{42}$$

The standard errors for T_i estimates are computed as usual from the diagonal elements of $\mathbf{C}_{\mathbf{T}}$. Note that when comparing different T_i values, correlated errors due to calibration uncertainties ($\mathbf{C}_{\mathbf{p}}$) and/or to positive covariance in the Δ_{47} measurements ($\mathbf{C}_{\mathbf{y}}$) will tend to cancel out if the full $\mathbf{C}_{\mathbf{T}}$ errors are properly propagated.

References

Adcock, R.J., 1878. A problem in least squares. Analyst 5 (2), 53. https://doi.org/10.2307/2635758.

Anderson, N.T., Kelson, J.R., Kele, S., Daëron, M., Bonifacie, M., Horita, J., Mackey, T.J., John, C.M., Kluge, T., Petschnig, P., Jost, A.B., Huntington, K.W., Bernasconi, S.M., Bergmann, K.D., 2021. A unified clumped isotope thermometer calibration (0.5–1,100 °C) using carbonate-based standardization. Geophys. Res. Lett. 48 (7) https://doi.org/10.1029/2020gl092069.

Anderson, N.T., Bonifacie, M., Jost, A.B., Siebert, J., Bontognali, T., Horita, J., Müller, I. A., Bernasconi, S.M., Bergmann, K.D., 2023. Re-assessing the need for apatite- and dolomite-specific calibrations of the carbonate clumped isotope thermometer. Geochem. Geophys. Geosyst. https://doi.org/10.1029/2023GC011049 in press.

Bernasconi, S.M., Daëron, M., Bergmann, K.D., Bonifacie, M., Meckler, A.N., Affek, H.P., Anderson, N., Bajnai, D., Barkan, E., Beverly, E., Blamart, D., Burgener, L., Calmels, D., Chaduteau, C., Clog, M., Davidheiser-Kroll, B., Davies, A., Dux, F., Eiler, J.M., Elliot, B., Fetrow, A.C., Fiebig, J., Goldberg, S., Hermoso, M., Huntington, K.W., Hyland, E., Ingalls, M., Jaggi, M., John, C.M., Jost, A.B., Katz, S., Kelson, J., Kluge, T., Kocken, I.J., Laskar, A., Leutert, T.J., Liang, D., Lucarelli, J., Mackey, T.J., Mangenot, X., Meinicke, N., Modestou, S.E., Müller, I.A., Murray, S.,

Neary, A., Packard, N., Passey, B.H., Pelletier, E., Petersen, S., Piasecki, A., Schauer, A., Snell, K.E., Swart, P.K., Tripati, A., Upadhyay, D., Vennemann, T., Winkelstern, I., Yarian, D., Yoshida, N., Zhang, N., Ziegler, M., 2021. InterCarb: a community effort to improve inter-laboratory standardization of the carbonate clumped isotope thermometer using carbonate standards. Geochem. Geophys. Geosyst. 22 (5) https://doi.org/10.1029/2020GC009588.

Breitenbach, S.F.M., Mleneck-Vautravers, M.J., Grauel, A.-L., Lo, L., Bernasconi, S.M., Müller, Inigo A., Rolfe, James, Gázquez, Fernando, Greaves, Mervyn, Hodell, D.A., 2018. Coupled Mg/Ca and clumped isotope analyses of foraminifera provide consistent water temperatures. Geochim. Cosmochim. Acta. https://doi.org/10.1016/j.gca.2018.03.010.

Chandler, J.P., 1972. On an iterative procedure for estimating functions when both variables are subject to error. Technometrics 14 (1), 71–76. https://doi.org/10.1080/00401706.1972.10488884.

Connelly, J.N., Bollard, J., Costa, M.M., Vermeesch, P., Bizzarro, M., 2021. Improved methods for high-precision Pb–Pb dating of extra-terrestrial materials. J. Anal. At. Spectrom. 36 (12), 2579–2587.

Costa, M.M., Jensen, N.K., Bouvier, L.C., Connelly, J.N., Mikouchi, T., Horstwood, M.S. A., Suuronen, J.-P., Moynier, F., Deng, Z., Agranier, A., Martin, L.A.J., Johnson, T.E., Nemchin, A.A., Bizzarro, M., 2020. The internal structure and geodynamics of Mars inferred from a 4.2-Gyr zircon record. Proc. Natl. Acad. Sci. 117 (49), 30973–30979. https://doi.org/10.1073/pnas.2016326117.

Daëron, M., 2021. Full propagation of analytical uncertainties in Δ_{47} measurements. Geochem. Geophys. Geosyst. 22 (5) https://doi.org/10.1029/2020gc009592.

Daëron, M., Gray, W.R., 2023. Revisiting oxygen-18 and clumped isotopes in planktic and benthic foraminifera. Paleoceanography and Paleoclimatology 38 (10). https://doi.org/10.1029/2023PA004660.

Daëron, M., Guo, W., Eiler, J., Genty, D., Blamart, D., Boch, R., Drysdale, R., Maire, R., Wainer, K., Zanchetta, G., 2011. $^{13}C^{18}O$ clumping in speleothems: Observations from natural caves and precipitation experiments. Geochim. Cosmochim. Acta 75 (12), 3303–3317. https://doi.org/10.1016/j.gca.2010.10.032.

Daëron, M., Drysdale, R.N., Peral, M., Huyghe, D., Blamart, D., Coplen, T.B., Lartaud, F., Zanchetta, G., 2019. Most Earth-surface calcites precipitate out of isotopic equilibrium. Nat. Commun. 10 (1) https://doi.org/10.1038/s41467-019-08336-5.

De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L., 2000. The Mahalanobis distance. Chemom. Intell. Lab. Syst. 50 (1), 1–18. https://doi.org/10.1016/s0169-7439(99)00047-7.

Deming, W.E., 1943. Statistical Adjustment of Data. Wiley, New York.

Eiler, J.M., 2007. "Clumped-isotope" geochemistry—the study of naturally-occurring, multiply-substituted isotopologues. Earth Planet. Sci. Lett. 262, 309–327. https://doi.org/10.1016/j.epsl.2007.08.020.

Eiler, J.M., 2011. Paleoclimate reconstruction using carbonate clumped isotope thermometry. Quat. Sci. Rev. 30, 3575–3588. https://doi.org/10.1016/j.quascirev.2011.09.001.

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

Fiebig, J., Daëron, M., Bernecker, M., Guo, W., Schneider, G., Boch, R., Bernasconi, S.M., Jautzy, J., Dietzel, M., 2021. Calibration of the dual clumped isotope thermometer for carbonates. Geochim. Cosmochim. Acta 312, 235–256. https://doi.org/10.1016/j.gca.2021.07.012.

Gauss, C.F., 1809. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Perthes, F. & Besser, J. H., Hambourg.

Ghosh, P., Adkins, J., Affek, H., Balta, B., Guo, W., Schauble, E.A., Schrag, D., Eiler, J.M., 2006. ^{13}C–^{18}O bonds in carbonate minerals: a new kind of paleothermometer. Geochim. Cosmochim. Acta 70, 1439–1456. https://doi.org/10.1016/j.gca.2005.11.014.

Golub, G.H., Van Loan, C.F., 1980. An analysis of the total least squares problem. SIAM J. Numer. Anal. 17 (6), 883–893.

Guo, W., 2020. Kinetic clumped isotope fractionation in the DIC-H2O-CO2 system: patterns, controls, and implications. Geochim. Cosmochim. Acta 268, 230–257. https://doi.org/10.1016/j.gca.2019.07.055.

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

Hill, P.S., Tripati, A.K., Schauble, E.A., 2014. Theoretical constraints on the effects of pH, salinity, and temperature on clumped isotope signatures of dissolved inorganic carbon species and precipitating carbonate minerals. Geochim. Cosmochim. Acta 125, 610–652. https://doi.org/10.1016/j.gca.2013.06.018.

Houseman, E., Andrés, Ryan, Louise, M., Coull, Brent A., 2004. Cholesky residuals for assessing normal errors in a linear model with correlated outcomes. J. Am. Stat. Assoc. 99 (466), 383–394. https://doi.org/10.1198/016214504000000403.

Huyghe, D., Daëron, M., de Rafelis, M., Blamart, D., Sébilo, M., Paulet, Y.-M., Lartaud, F., 2022. Clumped isotopes in modern marine bivalves. Geochim. Cosmochim. Acta 316, 41–58. https://doi.org/10.1016/j.gca.2021.09.019.

Jautzy, J.J., Savard, M.M., Dhillon, R.S., Bernasconi, S.M., Smirnoff, A., 2020. Clumped isotope temperature calibration for calcite: Bridging theory and experimentation. Geochem. Perspecti. Lett. 36–41. https://doi.org/10.7185/geochemlet.2021.

Kocken, Ilja J., Müller, Inigo A., Ziegler, Martin, 2019. Optimizing the use of carbonate standards to minimize uncertainties in clumped isotope data. Geochem. Geophys. Geosyst. 20 (11), 5565–5577. https://doi.org/10.1029/2019gc008545.

Kukush, A., Van Huffel, S., 2004. Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model AX=B. Metrika 59 (1), 75–97. https://doi.org/10.1007/s001840300272.

Kummell, C.H., 1879. Reduction of observation equations which contain more than one observed quantity. Analyst 6 (4), 97. https://doi.org/10.2307/2635646.

Legendre, A.-M., 1805. Nouvelles méthodes pour la détermination des orbites des comètes. Didot, F., Paris.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2 (2), 164–168. https://doi.org/10.1090/qam/10666.

Ludwig, K.R., 1998. On the treatment of concordant uranium-lead ages. Geochim. Cosmochim. Acta 62 (4), 665–676. https://doi.org/10.1016/s0016-7037(98)00059-3.

Lybanon, M., 1984. Comment on "Least squares when both variables have uncertainties". Am. J. Phys. 52 (3), 276–278. https://doi.org/10.1119/1.13713.

Markovsky, I., Van Huffel, S., 2007. Overview of total least-squares methods. Signal Process. 87 (10), 2283–2302. https://doi.org/10.1016/j.sigpro.2007.04.004.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11 (2), 431–441. http://www.jstor.org/stable/2098941.

McLean, N.M., Bowring, J.F., Bowring, S.A., 2011. An algorithm for U-Pb isotope dilution data reduction and uncertainty propagation. Geochem. Geophys. Geosyst. 12 (6) https://doi.org/10.1029/2010gc003478.

McLean, N.M., Bowring, J.F., Gehrels, G., 2016. Algorithms and software for U-Pb geochronology by LA-ICPMS. Geochem. Geophys. Geosyst. 17 (7), 2480–2496. https://doi.org/10.1002/2015gc006097.

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

Peral, Marion, Daëron, Mathieu, Blamart, Dominique, Bassinot, Franck, Dewilde, Fabien, Smialkowski, Nicolas, Isguder, Gulay, Bonnin, Jérôme, Jorissen, Frans, Kissel, Catherine, Michel, Elisabeth, Riveiros, Vázquez, Natalia, Waelbroeck, Claire, 2018. Updated calibration of the clumped isotope thermometer in planktonic and benthic foraminifera. Geochim. Cosmochim. Acta 239, 1–16. https://doi.org/10.1016/j.gca.2018.07.016.

Petersen, S.V., Defliese, W.F., Saenger, C., Daëron, M., John, C.M., Huntington, K.W., Kelson, J.R., Bernasconi, S.M., Colman, A.S., Kluge, T., Olack, G.A., Schauer, A.J., Bajnai, D., Bonifacie, M., Breitenbach, S.F.M., Fiebig, J., Fernandez, A.B., Henkes, G. A., Hodell, D., Katz, A., Kele, S., Lohmann, K.C., Passey, B.H., Peral, M., Petrizzo, D. A., Rosenheim, B.E., Tripati, A., Venturelli, R., Young, E.D., Wacker, U., Winkelstern, I.Z., 2019. Effects of improved ^{17}O correction on interlaboratory agreement in clumped isotope calibrations, estimates of mineral-specific offsets, and temperature dependence of acid digestion fractionation. Geochem. Geophys. Geosyst. https://doi.org/10.1029/2018gc008127.

R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Renne, Paul R., Swisher, Carl C., Deino, Alan L., Karner, Daniel B., Owens, Thomas L., DePaolo, Donald J., 1998. Intercalibration of standards, absolute ages and uncertainties in $^{40}Ar/^{39}Ar$ dating. Chem. Geol. 145 (1), 117–152.

Rioux, Matthew, Lissenberg, C. Johan, McLean, Noah M., Bowring, Samuel A., MacLeod, Christopher J., Hellebrand, Eric, Shimizu, Nobumichi, 2012. Protracted timescales of lower crustal growth at the fast-spreading East Pacific rise. Nat. Geosci. 5 (4), 275–278.

Roy, P.T., Owen, A.B., Balandat, M., Haberland, M., 2023. Quasi-Monte Carlo methods in Python. J. Open Source Softw. 8 (84), 5309. https://doi.org/10.21105/joss.05309.

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. https://doi.org/10.1016/j.gca.2012.09.035.

Schauble, E.A., Ghosh, P., Eiler, J.M., 2006. Preferential formation of ^{13}C–^{18}O bonds in carbonate minerals, estimated using first-principles lattice dynamics. Geochim. Cosmochim. Acta 70, 2510–2529. https://doi.org/10.1016/j.gca.2006.02.011.

Spencer, C., Kim, S.-T., 2015. Carbonate clumped isotope paleothermometry: a review of recent advances in CO_2 gas evolution, purification, measurement and standardization techniques. Geosci. J. 19 (2), 357–374. https://doi.org/10.1007/s12303-015-0018-1.

Stoica, P., Selen, Y., 2004. Model-order selection. IEEE Signal Process. Mag. 21 (4), 36–47. https://doi.org/10.1109/msp.2004.1311138.

Tellinghuisen, J., 2020. Least squares methods for treating problems with uncertainty in x and y. Anal. Chem. 92 (16), 10863–10871. https://doi.org/10.1021/acs.analchem.0c02178.

Tripati, A.K., Hill, P.S., Eagle, R.A., Mosenfelder, J.L., Tang, J., Schauble, E.A., Eiler, J. M., Zeebe, R.E., Uchikawa, J., Coplen, T.B., Ries, J.B., Henry, D., 2015. Beyond temperature: clumped isotope signatures in dissolved inorganic carbon species and the influence of solution chemistry on carbonate mineral composition. Geochim. Cosmochim. Acta 166, 344–371. https://doi.org/10.1016/j.gca.2015.06.021.

Uchikawa, J., Chen, S., Eiler, J.M., Adkins, J.F., Zeebe, R.E., 2021. Trajectory and timescale of oxygen and clumped isotope equilibration in the dissolved carbonate

system under normal and enzymatically-catalyzed conditions at 25 °C. Geochim. Cosmochim. Acta 314, 313–333. https://doi.org/10.1016/j.gca.2021.08.014.
Van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problem. Society for Industrial & Applied Mathematics. https://doi.org/10.1137/1.9781611971002.
Vermeesch, P., 2015. Revised error propagation of $^{40}Ar/^{39}Ar$ data, including covariances. Geochim. Cosmochim. Acta 171, 325–337.
Vermeesch, P., 2018. IsoplotR: a free and open toolbox for geochronology. Geosci. Front. 9 (5), 1479–1493. https://doi.org/10.1016/j.gsf.2018.04.001.
Vermeesch, P., 2022. An algorithm for U–Pb geochronology by secondary ion mass spectrometry. Geochronology 4 (2), 561–576.
Watkins, J.M., Devriendt, L.S., 2022. A combined model for kinetic clumped isotope effects in the $CaCO_3$–DIC–H_2O system. Geochem. Geophys. Geosyst. 23 (8) https://doi.org/10.1029/2021gc010200.
Watkins, J.M., Hunt, J.D., 2015. A process-based model for non-equilibrium clumped isotope effects in carbonates. Earth Planet. Sci. Lett. 432, 152–165. https://doi.org/10.1016/j.epsl.2015.09.042.
Wendt, I., Carl, C., 1991. The statistical distribution of the mean squared weighted deviation. Chem. Geol. 86 (4), 275–285. https://doi.org/10.1016/0168-9622(91)90010-T.
York, D., 1966. Least-squares fitting of a straight line. Can. J. Phys. 44 (5), 1079–1086. https://doi.org/10.1139/p66-090.
York, D., 1969. Least squares fitting of a straight line with correlated errors. Earth Planet. Sci. Lett. 5, 320–324.
York, D., Evensen, N.M., López Martínez, M., De Basabe Delgado, J., 2004. Unified equations for the slope, intercept, and standard errors of the best straight line. Am. J. Phys. 72 (3), 367–375. https://doi.org/10.1119/1.1632486.