



# **The quantification of downhole fractionation for laser ablation mass spectrometry**

Jarred C. Lloyd<sup>1,2</sup>, Carl Spandler<sup>1</sup>, Sarah E. Gilbert<sup>3</sup>, Derrick Hasterok<sup>1</sup>

<sup>1</sup>Department of Earth Science, The University of Adelaide, Australia

5 2Geological Survey of South Australia, Department for Energy and Mining, Government of South Australia, Australia 3Adelaide Microscopy, The University of Adelaide, Australia

*Correspondence to*: Jarred C. Lloyd (jarred.lloyd@adelaide.edu.au)

**Abstract.** Downhole fractionation (DHF), a known phenomenon in static spot laser ablation, remains one of the most significant sources of uncertainty for laser-based geochronology. A given DHF pattern is unique to a set of conditions, 10 including material, inter-element analyte pair, laser conditions, and spot volume/diameter. Current modelling methods (simple or complex linear models, spline-based modelling) for DHF do not readily lend themselves to uncertainty propagation, nor do they allow for quantitative inter-session comparison, let alone inter-laboratory or inter-material comparison.

In this study, we investigate the application of orthogonal polynomial decomposition for quantitative modelling of LA-ICP-MS DHF patterns with application to an exemplar U–Pb dataset across a range of materials and analytical sessions. We outline

15 the algorithm used to compute the models and provide a brief interpretation of the resulting data. We demonstrate that it is possible to quantitatively compare the DHF patterns of multiple materials across multiple sessions accurately, and use uniform manifold approximation and projection (UMAP) to help visualise this large dataset.

We demonstrate that the algorithm presented advances our capability to accurately model LA-ICP-MS DHF and may enable reliable decoupling of the DHF correction for non-matrix matched materials, improved uncertainty propagation, and inter-

20 laboratory comparison. The generalised nature of the algorithm means it is applicable not only to geochronology but also more broadly within the geosciences where predictable linear relationships exist.

# **1 Introduction**

LA-ICP-MS of geological materials has significantly advanced since its adoption at the end of the 20<sup>th</sup> century and is today the technique of choice for most applications for mineral geochronology. Initial geochronological studies on zircon were only

- 25 able to produce tens of individual data per session due to technical, time, and computing limitations (e.g. Hirata and Nesbitt, 1995). Nowadays high-precision individual analyses on multiple minerals using a range of geochronologic systems (e.g. U– Pb, Rb–Sr, Lu–Hf, Re–Os) can be rapidly and accurately acquired (Chew et al., 2019; Gehrels et al., 2008; Glorie et al., 2023; Hogmalm et al., 2017; Kendall-Langley et al., 2020; Larson et al., 2024; McFarlane, 2016; Mohammadi et al., 2024; Roberts et al., 2020; Simpson et al., 2021; Subarkah et al., 2021; Tamblyn et al., 2024; Zack et al., 2011). However, to achieve accurate,
- 30 high-quality data, correction procedures need to be implemented, including calibration to reference material(s) (RM), corrections for matrix offsets, and inter-element downhole fractionation (DHF) (Agatemor and Beauchemin, 2011; Allen and Campbell, 2012; Gilbert et al., 2017; Günther et al., 2001; McLean et al., 2016; Ver Hoeve et al., 2018). As DHF is a volume-dependant spot-ablation phenomena unique to each material and (inter-element) analyte ratio pair (e.g. NIST glasses, zircon, 206Pb/238U, 207Pb/235U, 87Rb/87Sr) it cannot be corrected via ICP-MS tuning (Mank and Mason, 1999).
- 35 However, DHF can be optimised to balance the overall count rate against the magnitude of DHF by adjusting the spot size and fluence parameters for analysis of a given material (Guillong and Günther, 2002; Košler et al., 2005). Researchers have proposed several methods for minimising DHF, typically by surface rastering or limiting acquisition times to short intervals; however, these compromise either the spatial or temporal resolution of analysis (Horstwood et al., 2003; Paton et al., 2010). Alternatively, regression modelling can be used to correct DHF during the data reduction process either using a predetermined





- 40 empirical model or fitting a model to the observed data (e.g. Horn et al., 2000; Paton et al., 2010). In the current state of LA-ICP-MS methodology the latter method of fitting a regression model to the observed DHF of an RM is used by commercial data reduction software (e.g. LADR: Norris and Danyushevsky, 2018; Iolite: Paton et al., 2011). In this process a matrixmatched RM is used (or a glass RM if no suitable matrix-matched material is available) and requires the assumption that the unknowns (i.e. target samples) have the same DHF as the RM. If the modelled RM poorly matches the DHF of the unknown
- 45 it will introduce some artefact (user induced error) that may either increase the uncertainty of the observed ratio of the unknowns if using a time matched signal, or result in over- or underestimation of the observed ratio if using a subset of the signal [Fig. 1]. Additionally, these modelling methods (i) do not readily lend themselves to arithmetic propagation of their uncertainties (Paton et al., 2010), (ii) require an arbitrary user choice for the fitting method (simple or complex linear regression, polynomial order, or non-linear regression), and (iii) cannot be quantitatively compared between laboratories or 50 even intra-laboratory analytical sessions. Thus, inter-element DHF remains one of the largest sources of uncertainty for LA-
- ICP-MS analyses and is difficult to either qualitatively or quantitatively compare (Horstwood et al., 2016). not corrected for DHF DHF corrected using NIST DHF corrected using GJ1



**Figure 1 – Demonstration of the impact of using an inappropriate material to correct downhole fractionation (DHF). The two rows are 206Pb-238U ratios for NIST610 glass (a–c) and GJ1 zircon (d–f), respectively, collected in a single session. The columns from left to right are, not corrected for DHF (grey points, raw ratios), DHF corrected using NIST610 as the model, and DHF corrected using GJ1 as the model. Blue points indicate an appropriate DHF correction (e.g. GJ1 by GJ1) while maroon points indicate an inappropriate DHF correction (e.g. GJ1 by NIST610). Black horizontal spans indicate the geometric mean and its asymmetric lower and upper 2 standard error for each data set. The percent error shown is 2-standard error level of the larger geometric standard error (usually upper uncertainty). Note the significantly larger error for the inappropriate corrections (maroon points) compared to the appropriate (blue) corrections. These uncertainties would then propagate into further processing steps to obtain accurate LA-ICP-MS ratios (e.g. calibration to a known value).** 

Here we present a method to numerically quantify DHF using the coefficients of orthogonal polynomials fit to the signals. We develop upon the algorithms for modelling rare earth element patterns by O'Neill (2016) and Anenburg & Williams (2022). 55 Our modified implementation of these algorithms accounts for input uncertainties and can be more generally applied to other purposes where there is a predictable x-to-y relationship, such as varying analyte signal intensity versus time, depth, or cumulative volume (i.e. during progressive pit ablation). Using a LA-ICP-MS dataset (time, analyte counts per second), an orthogonal decomposition is used to fit a polynomial to the signal data by generalised least squares [Fig. 2]. This method of using orthogonal polynomial decomposition during model fitting enables the quantitative comparison of DHF patterns via





- 60 their coefficients for data from different analyte ratio pairs, for varied materials, and for inter-session [Fig. 2] and interlaboratory datasets. Furthermore, this fitting method allows accurate computation of the uncertainty for each coefficient (and thus overall fit) and its covariance which can then be propagated into the result. In this research we highlight the utility of the algorithm for numeric comparison of the U–Pb DHF patterns from a range of natural and synthetic reference materials (RMs) used for U–Pb age date calibration. We envisage that this algorithm could be
- 65 used to self-correct individual analyses for DHF (or as spline bases for signals that display zonation, e.g. different age domains), with a fall back to a known RM in the case of particularly noisy (analytical) data. This would allow reliable decoupling of the DHF correction for non-matrix matched materials, and propagation of the uncertainty in the downhole model into the result, thereby improving the accuracy and precision of LA-ICP-MS analyses.







**Figure 2 – (a) Centred, pointwise ratios for all zircon GJ1 analyses 206Pb/238U analyses measured in multiple analytical sessions from 2020 to 2024. All analyses were collected using a 30 µm spot, 5 Hz repetition rate, and nominal fluence of 2.0 Jcm-2 . The black line is the fitted orthogonal polynomial, with grey shading indicating the 95% confidence interval of the fit. The shaded confidence interval is barely visible due to its small absolute range. (b) Individual orthogonal components of the fit in the upper panel excluding λ<sup>0</sup> (mean ratio, i.e. vertical central tendancy) as this is a function of mass-spectrometer tuning and not a result of downhole fractionation. For this aggregated GJ1 data, different mass-spectrometer tuning parameters is accounted for by centring the data resulting in λ<sup>0</sup> ≈ 0. Components λ1 and higher represent the increasing shape component (linear slope, quadratic curvature, etc) of the observed downhole fractionation pattern. The sum of all components forms the polynomial in the upper panel.**

### 70 **2 Experimental**

To improve our ability to assess, quantify, and compare downhole fractionation we implemented a method of orthogonal polynomial decomposition to perform linear regression model fitting to time-resolved analytical data from ICP-MS instruments. In keeping with the terminology of O'Neill (2016) and Anenburg & Williams (2022) we use the term lambda (*λ*) coefficients to denote the polynomial coefficients derived from the fit to the model using orthogonal decomposition.





- 75 Unlike regular linear regression fitting, the implementation of orthogonal polynomial decomposition for regression modelling imparts the property of independence on lower order polynomial coefficients from their higher order counterparts (Bevington and Robinson, 2003). That is, the first coefficient is independent of the second and higher order coefficients, while the second coefficient is independent of the third and higher order coefficients and so forth. In practice, this means that the values of the polynomial coefficients are stable (you can fit a first or second order polynomial to some data and the value of first coefficient
- 80 will not change) and they have some physical meaning in relation to the data. In the case of DHF data,  $\lambda_0$  (the first coefficient) represents the (arithmetic) mean ratio, while  $\lambda_1$  and higher are the shape parameters that represent the DHF pattern [Fig. 2]. For example,  $\lambda_1$  represents the linear slope while  $\lambda_2$  represents the quadratic curve of the DHF pattern [Fig. 2]. The independence and physical meaning of the lambda coefficients allows them to be used to quantitatively compare independent fits (e.g. single analyses, materials, analytical sessions, differing laboratories) so long as other parameters (e.g. fluence, spot
- 85 diameter/volume, laser wavelength) are considered. Further details on the fitting algorithm and mathematics behind this process are outlined below in Sect. 2.2, detailed in Appendix A, and described in Bevington & Robinson (2003), O'Neill (2016), and Anenburg & Williams (2022). The raw data ingestion and required preprocessing steps are outlined below in section Sect. 2.1.
- Code written to perform the data ingestion, preprocessing, and fitting were written in the Julia (version 1.10) programming 90 language (Bezanson et al., 2017). The code forms part of an in-development Julia package, GeochemistryTools.jl that will be formally released in the future. Should users want early access to the package or source code for the algorithms outlined in this manuscript they are available via GitHub (see code availability). Figures were generated using the Makie.jl plotting library (Danisch and Krumbiegel, 2021) using "Scientific colour maps" (Crameri et al., 2020) implemented in ColorSchemes.jl.

#### **2.1 Data ingestion and preprocessing**

95 Data were imported from the raw ICP-MS CSV files; in this case from Agilent ICP-MS instruments. The algorithm written for this task processes the individual files (which can be a single analysis, a specified sample, or entire session) and performs several operations, outlined in Fig. 3, on user-specified mass counts per

100 second (CPS) columns (i.e. channels, analyte signals). An automated procedure for determining gas blank and signal windows, as well as laser fire time, and aerosol arrival was developed to aid in rapid data processing and quality control. This algorithm is outlined in Fig. S1 with examples of the 105 automatically determined time points and signal windows shown

in Fig. S2.

Gas blank determination uses a geometric mean adapted for zeros [Eq. (1)] as per Habib (2012),

for 
$$
G \ge 0
$$
,  $G = \frac{n_+G_+ + n_0G}{N} = \frac{n_+}{N}G_+$ 

where G is the geometric mean, N is the total count of data,  $x_i$  is the *i*-th input value from  $i = 1$  to  $i = N$ ,  $G_{+}$  is the geometric mean of all  $x_i > 0$ ,  $n_+$  is the count of  $x_i > 0$ ,  $G_0$  is the geometric mean of all  $x_i = 0$  (i.e. 0), and  $n_0$  is the count of all  $x_i = 0$ .



**Figure 3 – Outline of data ingestion algorithm. CPS = counts per second of analyte (e.g. 238U).**

 $110$  (1)





- 115 As previously outlined (Sect. 2) the nature of orthogonal polynomial decomposition means that lambda coefficients (and their uncertainties) one and higher should not change with respect to a centred or non-centred dataset for single analysis. For sessionbased, sample-spot size aggregation of data there is a slight difference in the lambda coefficients, usually at the fifth decimal place, between the fits of centred and non-centred data. The slight discrepancy between the centred and non-centred for session fits suggests the central tendencies (i.e. geometric mean in our case) of the underlying analyses are slightly different to each
- 120 other. Intersession sample aggregation requires the centring of data for accurate numeric comparison, otherwise large differences in the coefficient values may occur as the central tendency (*λ*0) of data from a given sample can vary across sessions as it is a function of ICP-MS tuning. When aggregating data, it is grouped by the sample, spot size, and fluence to control for these variables.

Centring of the data aligns the central tendencies of each individual analysis to  $\sim$ 0 while retaining the scale and shape [Fig. 4]

125 of the change in raw ratios over the time-resolved signal, i.e. the DHF. Centring is performed by subtracting the central tendency (i.e. geometric mean, or a user could alternatively use the arithmetic mean, or median) from the time-resolved ratios for each individual file (i.e. analysis). This centring enables visual comparison [Fig. 4] and accurate numeric assessment of DHF behaviours across samples, sessions, spot sizes, and fluences without the need to calibrate data to a known reference material ratio for each session.







Figure  $4 - (a)$  Pointwise computed  $206Pb-238U$  ratios for each measurement of NIST610 glass in the dataset. (b) Centred pointwise Figure 4 – (a) Pointwise computed <sup>206</sup>Pb-<sup>238</sup>U ratios for each measurement of NIST610 glass in the dataset. (b) Centred pointwise<br><sup>206</sup>Pb-<sup>238</sup>U ratios for each measurement of NIST610 glass in the dataset. Centring prese **aliging their central tendancies while retaing the shape and scale of the DHF pattern thus allowing accurate comparison between measurements of differing spot size, measurements on different materials, and across ICP-MS sessions.**

130

# **2.2 Outline of the fitting algorithm**

In the following equations  $x$  denotes the independent variable (e.g. time in seconds) and the dependant variable (e.g. counts per second) is represented by y for observed values or  $\hat{y}$  for predicted values of some function of x, i.e.  $f(x)$ . For clarity, we use  $\varphi$  to differentiate the orthogonal function terms and  $\lambda$  to differentiate the coefficients of the orthogonal polynomial from 135 their regular counterparts ( $f$  and  $\beta$  respectively). Function terms and coefficients use  $j$  and  $l$  as indices with  $k$  denoting the polynomial order. The indicator *i* is used for the *i*-th value of x, y,  $\hat{y}_i$  for  $i = 1$  to  $i = N$ .

Polynomials of order *k* [Eq. (2)] can be orthogonal to each other under some inner product [Eq. (3)]. Orthogonal polynomials [Eq. (4)] have the important property that a lower order coefficient is independent of all higher order coefficients and thus each have a specific meaning with relation to the input data. The property of independence also enables efficient computation of





140 the polynomial statistics (errors, variances and covariances, residual sum of squares etc) at each order. The efficient computation is enabled because only the coefficient matrix of the highest order (five in our case) needs to be calculated, rather than the entire solution for each polynomial order. Once computed, subsequent calculations can use specific subsets of the operator and coefficient matrices to determine predictions and statistics for each polynomial order.

$$
f(x) = \beta_0 + \sum_{j=1}^{k} [\beta_j f_j(x)]
$$

 $145$  (2)

$$
\mathcal{L}^{\mathcal{L}}(\mathcal{L}
$$

$$
\varphi(x) = \lambda_0 + \sum_{j=1}^k [\lambda_j \varphi_j(x)]
$$

 $\langle f_i(x), f_i(x) \rangle = 0$  for  $j \neq l$ 

(4)

(3)

150 In our case, we apply this to the DHF observed during LA-ICP-MS using the orthogonal property in Eq. (5).

$$
\sum [\varphi_j(x_i)\varphi_l(x_i)] = 0 \text{ for } j \neq l
$$

(5)

We construct a fourth-order polynomial function [Eq. (6)] using the sum of a set of orthogonal polynomials (Anenburg and Williams, 2022; Bevington and Robinson, 2003; O'Neill, 2016).

155 
$$
\hat{y}(x_i) = \lambda_0 + \lambda_1 \varphi_1(x_i) + \lambda_2 \varphi_2(x_i) + \lambda_3 \varphi_3(x_i) + \lambda_4 \varphi_4(x_i)
$$

(6)

The orthogonal functions are solved for using the methods outlined by Anenburg & William (2022) which make use of Vietta's formula and polynomial root finding. See Anenburg & William (2022) or the source code for further detail. Once the orthogonal polynomial forms are computed, the algorithm generates the operator matrix **X** and uses generalised least-squares

160 regression [Eq. (7) to Eq. (9)] to solve the vector  $\Lambda$ , and matrix  $\Sigma$ , while accounting for analytical uncertainties. The vector  $\Lambda$ contains the polynomial coefficients  $(\lambda_0 \dots \lambda_4)$ , while the matrix  $\Sigma$  contains the variances and covariances.

$$
\hat{\mathbf{y}} = \mathbf{X}\mathbf{\Lambda} + \mathbf{\Sigma}
$$

$$
\boldsymbol{\Lambda} = (\mathbf{X}^{\mathrm{T}} \boldsymbol{\Omega} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \boldsymbol{\Omega} \mathbf{y}
$$

 $165$  (8)

(9)

(7)

A graphical representation of the individual orthogonal polynomial components and an example of the resulting fit is shown in Fig. 2.

 $\Sigma = (X^T \Omega X)^{-1}$ 

170 Our algorithm also allows the user to employ automated outlier removal with an outlier considered to have a studentised residual ≥ 3. The residuals used in outlier identification are computed from the orthogonal polynomial order with the minimum Akaike Information Criteria corrected (AICc) value [Eq. (10)] (Akaike, 1974; Burnham and Anderson, 2002). If outlier removal is enabled, it will iterate over the input data until no studentised residuals  $\geq 3$  remain in the fit or it has gone through ten iterations. The minimum AICc is used to determine the 'best fit' (i.e. minimum information loss) polynomial order for 175 each dataset.

$$
AICc = n\log\left(\frac{rss}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1}
$$
\n(10)

where  $k$  is the order plus two,  $n$  is the count of data, and rss is the residual sum of squares for the model.





Other measures of fit are computed and accessible to the user via the object created during the fitting algorithm. These measures 180 include the Bayesian Information Criteria corrected (BICc), the AICc and BICc weights, the reduced chi-squared statistic  $\chi^2_r$ , the multiple correlation coefficient ( $\rho^2$  or  $R^2$ ) value, and an unbiased Olkin-Pratt adjusted estimator ( $\rho_{OP}^2$ ) of  $\rho^2$  (Karch, 2020; Olkin and Pratt, 1958). Further detail can on these measures of fit be found in Appendix A. An additional consideration in this algorithm is that floating point computation is inherently inexact, has finite precision and can incur round-off errors (Fernández et al., 2003). While round-off errors are not an issue with most geochemical

185 computations generally, due to the often-small absolute values and numerous sequential computations being performed in this algorithm, these errors can lead to inaccuracy in the result. To overcome these potential accuracy issues we utilise the extended precision library MultiFloats.jl (Zhang, 2024). If, after the final calculation, the absolute value of a coefficient is less than the default machine rounding tolerance for the Float64 type it is rounded to zero.

# **2.3 Reference materials analysis by LA-ICP-MS**

- 190 Several reference materials for apatite (401, KO, MAD, Durango, Wilberforce), baddeleyite (BADPHE, G15874, G18650), monazite (TS1MNZ, 222, RW1, MAdel, MtGar, Ambat), rutile (R10, R19), titanite (Mt Painter, MKED), xenotime (MG1, BS1), and zircon (Mud Tank, Plešovice, GJ1, 91500, Temora, Rak17) were analysed for U–Pb isotope ratios and trace element concentrations using optimal methods for the specific mineral (Bockmann et al., 2022; Fletcher et al., 2004; Gain et al., 2019; Glorie et al., 2020; Hall et al., 2018; Horstwood et al., 2016; Liu et al., 2011; Lloyd et al., 2022; Payne et al., 2008; Sláma et
- 195 al., 2008; Spandler et al., 2016; Thompson et al., 2016; Wiedenbeck et al., 2004; Yang et al., 2024). An additional monazite (Pilbara) and several potential cassiterite (in-house) reference materials were also measured. Samples were analysed using a RESOlution-LR 193 nm ArF excimer laser ablation system coupled to an Agilent 8900 ICP-MS/MS. Both instruments are housed at the University of Adelaide within the analytical facilities at Adelaide Microscopy. Full metadata for LA-ICP-MS analysis can be found in Appendix B.





- 200 Additional zircon, rutile, and baddeleyite reference material data from prior studies were added to supplement the dataset and provide the ability to assess inter-session variability (Lloyd et al., 2020, 2022, 2023, 2024; van der Wolff, 2020; Yang et al., 2024). These supplementary data were collected using the same LA
- 205 system, coupled to either an Agilent 7900 ICP-MS (prior to November 2021) or the Agilent 8900 ICP-MS/MS in single quadrupole mode (from November 2021). Analytical conditions for these additional data can be found in the relevant references.

### **3 Results and Discussion**

- 210 In total, 5478 analyses (CSV files) were processed, and their DHF patterns were modelled with the orthogonal polynomial decomposition outlined above. This results in 5478 analysis fits, 188 session (sample per session) fits, and 58 sample fits [Fig. 5, Fig. 6] accounting for differences in spot size diameter across 29
- 215 unique materials. Linear slopes for the sample fits  $(\lambda_1)$  range from +3.17E-5 to +0.0217, quadratic curvatures  $(\lambda_2)$  range from -3.12E-5 to +3.17E-5, cubic curvatures  $(\lambda_3)$  range from -1.08E-5 to +2.15-5, and quartic curvatures  $(\lambda_4)$  range from -3.6E-6 to +3.9E-7. Given the small numbers and the large quantity of fits,
- 220 it is not feasible to display a table with all parameters nor is it intuitive for the reader. Instead, we provide the visual representation of  $\lambda_1$  plotted against  $\lambda_2$  and their best fit uncertainties (2-standard error) in Fig. 5, as well as the visual polynomial fit and its uncertainty for the sample fits in Fig. 6.



**Figure 5** - **Scatter plot of**  $\lambda_1$  (x) and  $\lambda_2$  (y) for sample (a), **session (b), and analysis (c) orthogonal polynomial fits. Uncertainty bars are 2-standard error. Fits with greater positive linear slope will plot further to the right, and fits with greater quadratic curvature will plot to higher (positive) or lower (negative) on the y-axis. The middle and lower panels are zoomed in to show greater detail in the areas where most data lie.**







**Figure 6 - Orthogonal polynomial fits of downhole fractionation (206Pb/238U) grouped by sample material. Note that the baddeleyite has a significantly larger y-axis scale due to the steeper linear fractionation component. Shaded areas show the 95% confidence interval of the individual fit. Increasing line solidity corresponds to increasing spot diameters.**





# **3.1 Data interpretation and visualisation**

Without needing reference material calibration, the derived coefficients  $(\lambda_1$  and higher) represent numerical parameters of downhole fractionation for any given analysis. The coefficients can be plotted as x-y scatter plots with uncertainty bars (e.g. Fig. 5) or used in dimensional reduction visualisation methods such as principal component analysis, multidimensional scaling, 230 and uniform manifold approximation and projection (UMAP). Specifically, we use UMAP to visualise [Fig. 7] the structure of a dataset (McInnes et al., 2020). UMAP takes the input multi-dimensional data (5478 x 4  $\lambda$  coefficients in this study) and tries to find a common embedding space (using the manifold assumption) to represent the local and global data topology in a lower dimension. UMAP has several variables called hyperparameters which control the output of the algorithm. The *n*-nearest neighbours hyperparameter is the most important for finding the balance between the global (low *n-n*) and local (high *n-n*)

235 data structure. We use a value of 10 for the *n*-nearest neighbours hyperparameter, and a value of 0.4 for the minimum distance hyperparameter to help alleviate overplotting. In simpler terms, UMAP works by constructing a graph in high-dimensional space (e.g. 4-dimensional) that is projected onto a lower-dimensional space (e.g. 2- dimensional) where points are connected based on their closeness in higher-dimensional space.

### 240 **3.1.1 Interpreting the (sample aggregated) data fits**

Oxides generally have a greater affine DHF component (*λ*1) compared to other materials analysed, cassiterite being an exception [Fig. 5 and Fig. 6]. The glass (NIST610), cassiterite, and xenotime samples show the least overall downhole fractionation. Of note with the glass (NIST610) is we can clearly see the impact of spot diameter (thus geometry) on DHF. With decreasing spot size, the overall magnitude of DHF is increased [Fig. 6], but it also exacerbates the complex shape 245 parameters (quadratic and cubic curvatures etc).

There are two distinct groupings (accounting for different spot sizes) for baddeleyite, with one group being the samples from Phalaborwa, and the other group being the Kovdor sample [Fig. 6]. The cause of this stark disparity in baddeleyite DHF is unknown, but the Ti concentration is 9–14x higher in the Phalaborwa samples than in the Kovdor samples. The obvious outliers in the apatite, cassiterite, and monazite subpanels of Fig. 6 are the Durango and Wilberforce apatite samples, CstT4370

- 250 cassiterite, and the Pilbara monazite sample. For the Pilbara monazite, some analyses show considerable variation in their Pb and U concentration (proxied by count rate; see signal plot supplementary figures available at figshare). Excluding these analyses from the sample/session base fit for the Pilbara monazite will reduce the polynomial confidence interval, and would improve the quality of the fit. However, this data is still analytically relevant and indicates that the Pilbara monazite is not suitable to be a reference material due to the variable ratio of Pb and U. For the Durango apatite, the flat DHF fit and larger
- 255 uncertainty are due to low Pb counts, and for Wilberforce, the steeper linear DHF component and larger uncertainty are due to inclusion of several points from some analyses that are highly leveraging the fit, even with automated outlier removal being applied. For cassiterite, the relatively flat fits, and outlier (CstT4370) are generally due to low Pb and/or U counts, and greater scatter in the underlying data. The discrepancies and/or larger uncertainties in the lambda coefficients for materials of the same type provide a way to numerically check for outliers in reference material data, prior to further data reduction, rather than
- 260 needing a user to review all the data graphically (although the user should still review their data via graphical means as well to check for spurious results).







**Figure 7 – (a) Uniform manifold approximation and project (UMAP) of λ coefficients 1…4 for all analysis data in the study. In a practical sense, the closer two points are to each other the more similar they are. (b) subset of (a) only including zircon analyses.**

# **3.1.2 Application of UMAP as a data visualisation aid**

Techniques like UMAP offer an efficient and robust way to deal with high-dimensionality and overplotting for data 265 visualisation purposes. UMAP is a data hungry method and performs better with higher data density (McInnes et al., 2020). For this study, a UMAP diagram [Fig. 7] helps to visualize the relative similarities between  $\lambda_{1...4}$  for the 5478 analyses. The resulting map [Fig. 7] clearly differentiates the various materials better than to the simple biplot [Fig. 5c], which suffers from overplotting, obscuring any similarities/trends.





In Fig. 7, it is obvious that baddeleyite (pale blue markers in lower left of upper panel) DHF is significantly different to all 270 other materials analysed in this study. The deviation by baddeleyite is also reflected in Fig. 5 where baddeleyite analyses are plotting to the far right of each subplot, indicating a much stronger affine fractionation component. What is also noticeable is that there is a large diversity in the DHF patterns of zircon [Fig. 7], which in part is due to analytical noise as the individual points represent a single analysis and the fits have greater uncertainty as they are more susceptible to that noise. Nevertheless, in general 91500 is behaving most differently of all the zircons at a given spot size [Fig. 8], and Plešovice is behaving the most

- 275 like NIST610 of all the zircons [Fig. 7]. We do not suggest that NIST610 glass is a suitable alternative to correct DHF for zircon (as seen in Fig. 6, and Fig. 1), rather that there is significant variation in the DHF of zircon standards, and therefore careful consideration should be given to applying appropriate zircon standards for analytical sessions depending on the unknown zircons to be analysed and the time period used for signal integration (Guillong and Günther, 2002; Hergenröder, 2006; Košler et al., 2005; Paton et al., 2010).
- 280 We can see that spot geometry has a significant impact on DHF [Fig. 6 (glass and zircon subpanels), Fig. 7, and Fig. 8], which is a known phenomenon (Horn et al., 2000; Mank and Mason, 1999; Paton et al., 2010). However, all zircons except 91500 have remarkably similar DHF patterns at 29 µm (and 30 µm) [Fig. 8]. The affine component (*λ*1) of the DHF pattern appears to be relatively constant between laser sessions for a given zircon and spot size, while the quadratic component varies (Fig. 8b) but has greater uncertainty associated with it. The exact mechanism as to why 91500 shows greater DHF than the other
- 285 analysed zircons is unknown, and not the focus of this manuscript; however, prior studies have investigated the potential causes of differing DHF patterns in zircon and glass (Košler et al., 2005), and it is possible that radiation damage plays a role (Allen and Campbell, 2012; Marillo-Sialer et al., 2014; Solari et al., 2015; Thompson et al., 2018).







**Figure 8 – (a) Orthogonal polynomial fitting of downhole fractionation for zircon data in this study. The visualised polynomials represent the best-order fit for the aggregated data of a given sample at a single spot size, e.g. GJ1 at 30 µm, GJ1 at 19 µm, …. Shaded areas show the 95% confidence interval of the individual fit. The inset graph (b) shows the λ<sup>1</sup> and λ<sup>2</sup> coefficients and their 2-standard error uncertainty for each of the session-based fits for the seven zircons analysed. Fits with a greater affine slope plot further to the right and fits with a greater negative quadratic curvature (i.e. the slope flattens at a faster rate) will plot toward the bottom.**

# 290 **3.2 Applications**

The algorithm defined in this manuscript offers a way to numerically quantify the DHF patterns of varied materials during LA-ICP-MS with uncertainty quantification. We envisage that this algorithm could be implemented in data reduction software to self-correct the DHF pattern of well-behaved materials (i.e. fitting a single geochemical zone), or as splines to geologically meaningful zonation, with a fallback to a known homogenous material where the analysis signal is complex (e.g. due to 295 complex zonation or inclusions). In the latter case, the user would need to select an appropriate material, and it would likely result in somewhat inaccurate correction in any case, thus impacting the accuracy of the final result [Fig. 1].





The quantification of the DHF pattern enables numeric assessment of fits to unknowns against fits of knowns to find the most similar material with respect to DHF and of the appropriate model order. For example, the AICc is minimised for the fourth order polynomial for NIST610 and the third order polynomial for most zircon RMs, we can see this reflected visually in Fig.

300 2, Fig. 4, Fig. 6, and Fig. 8. Using the measures of fit provides users with a way to choose an appropriate model that accurately reflects the uncertainty and scatter, while avoiding overfitting and the introduction of artefact errors. Additionally, this algorithm provides a way for laboratories to quantitatively compare the downhole fractionation behaviour of their reference materials and analytical setup against other laboratories. Furthermore, the generalised nature of this algorithm allows it to be used for orthogonal polynomial fitting, up to fourth order, of any data where is it sensible, i.e. there exists a

305 linear polynomial of order *k* (0…4) that can model the input data.

# **4 Conclusions**

Quantitative modelling of DHF patterns observed during static spot LA-ICP-MS is made possible by the algorithm we developed in the Julia programming language that uses orthogonal polynomial decomposition. We apply it to an exemplar dataset of U–Pb reference materials and guide the reader through data visualisation and interpretation of the derived

310 coefficients.

The algorithm can be used to quantitatively compare downhole fractionation for the same reference materials across laboratories and implemented into data reduction programs to numerically assess the similarities and fit qualities of DHF correction. It is probable that the algorithm could be used to self-correct a given analysis (within a single signal zone), with a fall back to a known reference material by quantitatively choosing the most similar fit in the case where the fit has high

315 uncertainty or is covering multiple signal zones. We also envisage that the algorithm could be more widely used for other polynomial modelling in the geosciences where there is a predictable x-to-y relationship in a linear space (e.g. element volatilities as the x-axis).

### **Appendix A: Algorithmic Mathematics**

- LA-ICP-MS count data are strictly positive integers and follow a (discrete) Poisson distribution at low count level (e.g. gas 320 blank), although eventually approximate a normal distribution at high count rate as implied by the central limit theorem (Bevington and Robinson, 2003). Data from a mass-spectrometer used for geochronology and/or elemental analysis are generally output in counts-per-second (CPS) not counts and violate the integer requirement of a discrete probability distribution. This combined with the less intuitive and asymmetric scale parameters of geometric means has led to the use of normal statistics (e.g. arithmetic means) for ratio computations. The use of normal statistics for gas blank measurements
- 325 generally leads to overestimation (arithmetic mean) or underestimation (median if lots of 0 counts), and to a violation of the equality  $A/B = (B/A)^{-1}$ . This latter violation is commonly seen in U–Pb geochronology, where <sup>206</sup>Pb/<sup>238</sup>U ≠ (<sup>238</sup>U/<sup>206</sup>Pb)<sup>−1</sup>. Given that elemental count data and elemental ratios are compositional data and are strictly positive real values (i.e. positively skewed) which often follow a log-normal distribution, a geometric mean is a more appropriate measure of the central tendency of the data. In our algorithm we use the geometric mean to compute the gas blank value.
- 330 While geometric means are often used to address the problems above, we use a modified geometric mean to incorporate valid zero values which a standard geometric mean cannot. We will first review the standard geometric mean before detailing the modified version.

The geometric mean is defined for a set of positive real numbers as Eq. (A1) (Habib, 2012),







 $335$  (A1)

where G is the geometric mean, N is the total count of data, and  $x_i$  is the *i*-th input value from  $i = 1$  to  $i = N$ . Alternatively, it can be calculated by the arithmetic mean of the logarithm of the values [Eq. (A2)] and then raising the result to e [Eq. (A3)] (Habib, 2012; Kirkwood, 1979).

$$
\log G = \frac{1}{N} \sum_{i=1}^{N} \log x_i
$$

 $340$  (A2)

$$
G=e^{\log(G)}
$$

(A3)

(A4)

As computing the product of an arbitrarily large series of numbers can lead to overflow errors (Fernández et al., 2003; Polhill 345 et al., 2006), and therefore inaccurate results, most geometric mean algorithms implement the second form where the arithmetic mean is calculated from the logarithm transformed data and raised to e. The logarithmic transformation requires that all  $x_i$ 0. Additionally, as this is a multiplicative mean, when any  $x_i = 0$ ,  $G = 0$ .

To overcome these limitations and obtain a more accurate estimate of the mean gas blank, we implement a geometric mean that accounts for zeros [Eq. (A4)]. This equation is effectively a weighted geometric mean of the values > 0, and the geometric 350 mean of the zeros, i.e. 0, with the weights equal to the number of values in each category (Habib, 2012),

for G 
$$
\geq
$$
 0, G =  $\frac{n_+G_+ + n_0G_0}{N} = \frac{n_+}{N}G_+$ 

where  $G_+$  is the geometric mean of all  $x_i > 0$ ,  $n_+$  is the count of  $x_i > 0$ ,  $G_0$  is the geometric mean of all  $x_i = 0$  (i.e. 0), and  $n_0$ is the count of all  $x_i = 0$ .

355 We also implement a geometric variance [Eq. (A5)], geometric standard deviation [Eq. (A6)], and standard error of the mean [Eq. (A7)] (Habib, 2012; Kirkwood, 1979).

for 
$$
G \ge 0
$$
,  $\sigma^2 = \left(\frac{n_+}{N}\right)^2 e^{\sigma_{\log x_+}^2}$   
\nfor  $G \ge 0$ ,  $\sigma = \sqrt{\left(\frac{n_+}{N}\right)^2 e^{\sigma_{\log x_+}^2}} = \sqrt{\sigma^2}$  (A5)

360 (A6)

for 
$$
G \ge 0
$$
,  $SE = \left(\frac{n_+}{N}\right)^2 e^{SE_{\log x_+}}$  (A7)

Where  $\sigma_{\log x_+}^2$  and  $SE_{\log x_+}$  are, respectively, the variance and standard error of the mean of the logarithms of  $x_i > 0$ , and again weighted by the ratio:  $n_{+}/N$ . In geometric statistics, the measures of variance, standard deviation, and standard errors are a 365 scale parameter. The corresponding uncertainty range of G for these statistics is asymmetric and denoted by

 $(G \times u; G \div u)$  where *u* is the corresponding statistic (e.g. variance,  $\sigma^2$ ).

# **Orthogonal Polynomial Decomposition**





We use orthogonal polynomial decomposition to fit a polynomial to analyte ratio data where the coefficients have physical 370 meaning in relation to that data. This decomposition enables quantification of the shape parameters of a downhole fractionation pattern. Integral to this process is the calculation of the orthogonal polynomials [Eq. (A8)] to be used to fit the final model.

$$
\hat{y}(x_i) = \lambda_0 + \lambda_1 \varphi_1(x_i) + \lambda_2 \varphi_2(x_i) + \lambda_3 \varphi_3(x_i) + \lambda_4 \varphi_4(x_i)
$$

$$
(A8)
$$

Our implementation can fit up to a fourth order polynomial and requires solving the orthogonal property of Eq. (A9) for each 375 orthogonal function, φ.

$$
\sum [\varphi_j(x_i)\varphi_k(x_i)] = 0 \text{ for } j \neq k
$$
\n(A9)

The functions  $\varphi_{1...4}$  are as follows:

$$
\varphi_1 = x_i - \beta
$$
  
\n
$$
\varphi_2 = (x_i - \gamma_1)(x_i - \gamma_2)
$$
  
\n
$$
\varphi_3 = (x_i - \delta_1)(x_i - \delta_2)(x_i - \delta_3)
$$
  
\n
$$
\varphi_4 = (x_i - \epsilon_1)(x_i - \epsilon_2)(x_i - \epsilon_3)(x_i - \epsilon_4)
$$

Where β,  $γ_n$ ,  $δ_n$ , ε<sub>n</sub> are predetermined constants calculated prior to fitting the final model. The solution to β is simply the arithmetic mean of the x values [Eq. (A10)].

385 
$$
\sum_{i=1}^{N} (x_i - \beta) = 0 \text{ leading to } \frac{1}{N} \sum_{i=1}^{N} (x_i) = \beta
$$
 (A10)

The solutions to  $\gamma_{1...2}, \delta_{1...3}, \epsilon_{1...4}$  require solving systems of equations with increasing complexity as follows, with a line break between polynomial orders:

$$
\sum (x_i - \gamma_1)(x_i - \gamma_2) = 0
$$
  
390 
$$
\sum x_i(x_i - \gamma_1)(x_i - \gamma_2) = 0
$$

$$
\sum (x_i - \delta_1)(x_i - \delta_2)(x_i - \delta_3) = 0
$$
  

$$
\sum x_i(x_i - \delta_1)(x_i - \delta_2)(x_i - \delta_3) = 0
$$
  

$$
\sum x_i^2(x_i - \delta_1)(x_i - \delta_2)(x_i - \delta_3) = 0
$$

395

$$
\sum (x_i - \varepsilon_1)(x_i - \varepsilon_2)(x_i - \varepsilon_3)(x_i - \varepsilon_4) = 0
$$
  

$$
\sum x_i(x_i - \varepsilon_1)(x_i - \varepsilon_2)(x_i - \varepsilon_3)(x_i - \varepsilon_4) = 0
$$
  

$$
\sum x_i^2(x_i - \varepsilon_1)(x_i - \varepsilon_2)(x_i - \varepsilon_3)(x_i - \varepsilon_4) = 0
$$
  

$$
\sum x_i^3(x_i - \varepsilon_1)(x_i - \varepsilon_2)(x_i - \varepsilon_3)(x_i - \varepsilon_4) = 0
$$

400 This system of equations can be solved for numerically using an optimisation algorithm, or as stated in Anenburg & Williams (2022) we can utilise Vieta's formulas to rearrange the complex system of equations to achieve an analytical solution. The application of Vieta's formula allows the conversion of the above complex systems to a simple polynomial whose real roots are the  $\gamma_{1...2}$ ,  $\delta_{1...3}$ ,  $\epsilon_{1...4}$  values. Defining  $\gamma_1 + \gamma_2 = a$  and  $\gamma_1 \gamma_2 = b$ , and through simplification we obtain the matrix form of the following problem:





$$
405\qquad \qquad \left[-\sum x_i \qquad N \atop -\sum x_i^2 \quad \sum x_i\right][_b^a] = \left[-\sum x_i^2 \atop -\sum x_i^3\right]
$$

Once solved, a and b are the used as coefficients in the quadratic polynomial [Eq. (A11)] whose real roots are the two parameters  $γ_1$  and  $γ_2$ .

$$
\gamma^2 - a\gamma + b = 0
$$

(A11)

410 Following the same process,  $\delta_{1...3}$  can be solved using Eq. (A12) and Eq. (A13), while  $\epsilon_{1...4}$  can be solved using Eq. (A14) and Eq. (A15).

$$
\begin{bmatrix} -\sum x_i^2 & \sum x_i & -N \\ -\sum x_i^3 & \sum x_i^2 & -\sum x_i \\ -\sum x_i^4 & \sum x_i^3 & -\sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -\sum x_i^3 \\ -\sum x_i^4 \\ -\sum x_i^5 \end{bmatrix}
$$
  

$$
\delta^3 - a\delta^2 + b\delta - c = 0
$$
 (A12)

(413)  
\n
$$
\begin{bmatrix}\n-\sum x_i^3 & \sum x_i^2 & -\sum x_i & N \\
-\sum x_i^4 & \sum x_i^3 & -\sum x_i^2 & \sum x_i \\
-\sum x_i^5 & \sum x_i^4 & -\sum x_i^3 & \sum x_i^2 \\
-\sum x_i^6 & \sum x_i^5 & -\sum x_i^4 & \sum x_i^3\n\end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} -\sum x_i^4 \\ -\sum x_i^5 \\ -\sum x_i^6 \\ -\sum x_i^7 \end{bmatrix}
$$
\n
$$
\epsilon^4 - a\epsilon^3 + b\epsilon^2 - c\epsilon + d = 0
$$
\n(A14)

(A15)

420 To account for errors in y (i.e. analyte ratios) we use generalised least squares [Eq. A16] to fit the model. let  $\Sigma = (X^T \Omega X)^{-1}$  and  $\Lambda = \Sigma X^T \Omega y$ , then  $\hat{y} = X\Lambda + \Sigma$ 

(A16)

To retain the orthogonal nature of the fit, the design matrix, **X**, uses the  $\varphi_{1...4}$  functions from above and is as follows:

$$
\begin{bmatrix} 1 & \varphi_1(x_1) & \varphi_2(x_1) & \varphi_3(x_1) & \varphi_4(x_1) \\ 1 & \varphi_1(x_i) & \varphi_2(x_i) & \varphi_3(x_i) & \varphi_4(x_i) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \varphi_1(x_n) & \varphi_2(x_n) & \varphi_3(x_n) & \varphi_4(x_n) \end{bmatrix}
$$

425 Let the individual weights,  $\omega_i$ , be the relative error of  $y_i$ , then the weight matrix,  $\Omega$ , is as follows:

⎣ ⎢ ⎢ ⎢ ⎢ ⎢ ⎢

$$
\begin{bmatrix} \frac{1}{\omega_1/_{\overline{\omega}}}& & & \\ & \frac{1}{\omega_{i/_{\overline{\omega}}}} & & \\ & & \frac{1}{\omega_{i/_{\overline{\omega}}}} & \\ & & & \frac{1}{\omega_n/_{\overline{\omega}}}\end{bmatrix}
$$

The system can then be solved for the vector of coefficients,  $\Lambda$ , using Eq. (A16).

In the following equations,  $y_i$  is the i-th value of the observed dependant variable (e.g. analyte ratio),  $\hat{y}_i$  is i-th value of the predicted dependant variable, n is the total number of values and k is the polynomial order. To assess the quality of fitted





430 models, and to assist with choosing an optimal model, we implemented several measures of fit. Standard measures of fit calculated are the residual sum of squares [rss, Eq. (A17)],

> $\text{rss}_k = \sum_{y_i} (y_i - y_{i_k})$ n  $i=1$ (A17)

435 the mean square error or reduced chi-squared value in multiple regression [mse or  $\chi^2_r$ , Eq. (A18)],

$$
\chi_{r_k}^2 = \text{mse}_{k} = \frac{\text{rss}}{n - k + 1}
$$

(A18)

the root mean square error [rmse, Eq. (A19)],

$$
rmse_k = \sqrt{mse_k}
$$

440 (A19)

and the multiple regression coefficient  $[\rho^2$ or  $R^2$ , Eq. (A20)]

$$
\rho^2 = 1 \frac{r s s_{\scriptscriptstyle k}}{[y - \bar{y}]^T \Omega [y - \bar{y}]}
$$

(A20)

Additionally, we implement two measures based on Bayesian reasoning and information theory. The first of these is the 445 corrected Bayesian (or Schwarz) information criterion [BICc, Eq. (A21)] and corresponding BICc weights [BICcW, Eq. (A22)] (Burnham and Anderson, 2002; Schwarz, 1978).

$$
BICc_k = n \log\left(\frac{rss}{n}\right) + (k+2)\log(n) + \log(2\pi) + n
$$
\n
$$
BICcW_k = \frac{\exp(-0.5BICc_k - \min[BICc])}{\sum_{0}^{max(k)} \exp(-0.5BICc_k - \min[BICc])}
$$
\n(A21)

450 (A22)

The second is the corrected Akaike information criterion [AICc, Eq. (A23)] and corresponding AICc weights [AICcW, Eq. (A24)] (Akaike, 1974; Burnham and Anderson, 2002).

AICc<sub>k</sub> = n log  $\left(\frac{rss}{n}\right)$  + 2(k + 2) +  $\frac{2(k+2)([k+2]+1)}{n-(k+2)-1}$  $n - (k + 2) - 1$  $455$  (A23)  $exp(-0.5AIC_{k} - min[AIC_{c}])$ 

$$
AICCW_k = \frac{1}{\sum_{0}^{max(k)} \exp(-0.5AICc_k - \min[AICc])}
$$
\n(A24)

Finally, we implement an Olkin-Pratt adjusted multiple correlation coefficient  $[\rho_{0P}^2, Eq. (A25)]$  which is the optimal unbiased estimator of  $\rho^2$  (Karch, 2020; Olkin and Pratt, 1958).

460 
$$
\hat{\rho}_{OP}^2(\rho^2) = 1 - \frac{n-3}{n - (k+1) - 1} (1 - \rho^2)_2 F_1\left(1, 1; \frac{n - (k+1) - 1}{2}; 1 - \rho^2\right)
$$
(A25)

The Olkin-Pratt adjusted multiple correlation coefficient requires computation of the Gauss hypergeometric function, which is computationally non-trivial, however, Karch (2020) outlines the process to do this, and this is implemented as Eq. (A26) using the Taylor series expansion [Eq. (A27)] of the Gauss hypergeometric function (Pearson et al., 2017).





465

$$
{}_{2}F_{1}(a,b;c;z) = \begin{cases} 0 & \text{if } z = 0 \\ \frac{c-1}{c-2} & \text{if } z = 1 \\ \sum_{j=0}^{\infty} \frac{(a)_{j}(b)_{j}}{(c)_{j}} \frac{1}{j!} z^{j} & \text{if } 0 < z < 1 \end{cases}
$$

(A26)

For the case,  $0 < z < 1$ , the Taylor series is truncated when either the ratio of the value for next term in the series and the current sum of the series are less than or equal to the machine epsilon value for a Float64 type, or when the number of iterations (and thus terms) reaches 1000 [Eq. (A27)]. This will effectively truncate the series when the machine cannot resolve the 470 difference between the change in successive terms.

Let 
$$
T_0 = 1, S_0 = 1, j = 1
$$
 then for  $1 < j < 1000$   $\wedge \frac{T_{j+1}}{S_j} > \epsilon$   
  

$$
T_{j+1} = T_j \left[ \frac{(a+j)(b+j)}{(c+j)z} \right]
$$
  

$$
S_j = \sum_{j=1}^{1000} S_0 + T_{j+1}
$$
 (A27)

475

### **Outlier Detection**

We implement automated outlier removal based on the studentised residual. An outlier is considered to have a studentised residual  $\geq$  3 from the model with the polynomial order (k) that minimises the AICc. This outlier removal process is computationally intensive as it requires calculation of leverages  $(h_{ii})$  which are the diagonal values of the projection matrix.

480 The individual leverages are calculated using Eq. (A28).

$$
h_{ii} = \sum \mathbf{X}_{i,1:k} \big[ \mathbf{\Sigma}_{1:k,1:k} \mathbf{X}_{1:k,1:n}^{\mathrm{T}} \mathbf{\Omega} \big]
$$

(A28)

From the non-studentised residuals [Eq. (A29)] and the mean square error [mse, Eq. (A18)], the studentised residuals are calculated using Eq. (A30).

$$
r_i = y_i - \hat{y}_i
$$
\n
$$
s_i = \sqrt{\text{mse}(1 - h_{ij})}
$$
\n(A29)

(A30

If the user chooses this automated outlier removal, the algorithm will loop until either a) no studentised residuals are  $\geq 3$ , or b) 490 the loop has performed ten iterations.





# **Appendix B: Laser Conditions**



# **Table B2: Summary of materials analysed in each session, and laser conditions**















#### **Data and Code Availability**

495 The Julia package which implements the above algorithms is in early development, however it is available to all via GitHub at: https://github.com/jarredclloyd/GeochemistryTools.jl

The raw and derived data is available from figshare at this link: https://doi.org/10.25909/26778298

The code used to compile the raw data, fit the data, and generate the figures in this manuscript are available from figshare at this link: https://doi.org/10.25909/26779255

### 500 **Supplementary Figures**

Supplementary figures S01 and S02 detailing the automatic signal times algorithm employed in this manuscript, and example data it was tested on are available at: https://doi.org/10.25909/27041821 Additional supplementary figures showing the individual analysis signals are also available from figshare at: https://doi.org/10.25909/26778592. A plot exists for each sample for each session, with the arbitrary colours of each plot

505 representing individual analyses.

# **CRediT Author Statement**

Jarred C. Lloyd: Conceptualisation, data curation, formal analysis, investigation, methodology, resources, software, validation, visualisation, writing – original, writing – review  $\&$  editing

**Carl Spandler**: Conceptualisation, resources, supervision, writing – review & editing

510 **Sarah E. Gilbert**: Formal analysis, methodology, writing – review & editing **Derrick Hasterok**: Methodology, validation, writing – review & editing

### **Competing Interests**

The authors declare that they have no conflicts of interest.





#### **Acknowledgements**

515 We acknowledge and pay respects to the Kaurna People, the traditional custodians whose ancestral lands the University of Adelaide is built on and we work on. We acknowledge the deep feelings of attachment and relationship of the Kaurna people to Country, and we respect and value their past, present and ongoing connection to the land and cultural beliefs. We acknowledge the Tate Museum and South Australian Museum for the provision of some sample material used in this study.

# **Funding**

520 This research was co-funded by the Australian Critical Minerals Research Centre at the University of Adelaide, and the Department for Energy and Mining, South Australia.

### **References**

Agatemor, C. and Beauchemin, D.: Matrix effects in inductively coupled plasma mass spectrometry: A review, Analytica Chimica Acta, 706, 66–83, https://doi.org/10.1016/j.aca.2011.08.027, 2011.

- 525 Akaike, H.: A new look at the statistical model identification, IEEE Trans. Automat. Contr., 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.
	- Allen, C. M. and Campbell, I. H.: Identification and elimination of a matrix-induced systematic error in LA–ICP–MS 206Pb/238U dating of zircon, Chem. Geol., 332–333, 157–165, https://doi.org/10.1016/j.chemgeo.2012.09.038, 2012.
- Anenburg, M. and Williams, M. J.: Quantifying the Tetrad Effect, Shape Components, and Ce–Eu–Gd Anomalies in Rare 530 Earth Element Patterns, Math Geosci, 54, 47–70, https://doi.org/10.1007/s11004-021-09959-5, 2022.
	- Bevington, P. R. and Robinson, D. K.: Data reduction and error analysis for the physical sciences, 3rd ed., McGraw-Hill, Boston, 320 pp., 2003.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B.: Julia: A Fresh Approach to Numerical Computing, SIAM Rev., 59, 65–98, https://doi.org/10.1137/141000671, 2017.

- 535 Bockmann, M. J., Hand, M., Morrissey, L. J., Payne, J. L., Hasterok, D., Teale, G., and Conor, C.: Punctuated geochronology within a sustained high-temperature thermal regime in the southeastern Gawler Craton, Lithos, 430–431, 106860, https://doi.org/10.1016/j.lithos.2022.106860, 2022.
	- Burnham, K. P. and Anderson, D. R.: Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed., Springer, New York, 488 pp., 2002.
- 540 Chew, D. M., Drost, K., and Petrus, J. A.: Ultrafast, > 50 Hz LA-ICP-MS Spot Analysis Applied to U–Pb Dating of Zircon and other U‐Bearing Minerals, Geostand. Geoanal. Res., 43, 39–60, https://doi.org/10.1111/ggr.12257, 2019.
	- Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nature Communications, 11, 5444, https://doi.org/10.1038/s41467-020-19160-7, 2020.
- Danisch, S. and Krumbiegel, J.: Makie.jl: Flexible high-performance data visualization for Julia, Journal of Open Source 545 Software, 6, https://doi.org/10.21105/joss.03349, 2021.
	- Fernández, J.-J., García, I., and Garzón, E. M.: Floating point arithmetic teaching for computational science, Future Generation Computer Systems, 19, 1321–1334, https://doi.org/10.1016/S0167-739X(03)00090-6, 2003.
- Fletcher, I. R., McNaughton, N. J., Aleinikoff, J. A., Rasmussen, B., and Kamo, S. L.: Improved calibration procedures and new standards for U–Pb and Th–Pb dating of Phanerozoic xenotime by ion microprobe, Chemical Geology, 209, 295– 550 314, https://doi.org/10.1016/j.chemgeo.2004.06.015, 2004.
	- Gain, S. E. M., Gréau, Y., Henry, H., Belousova, E., Dainis, I., Griffin, W. L., and O'Reilly, S. Y.: Mud Tank Zircon: Long‐ Term Evaluation of a Reference Material for U‐Pb Dating, Hf‐Isotope Analysis and Trace Element Analysis, Geostand. Geoanal. Res., 43, 339–354, https://doi.org/10.1111/ggr.12265, 2019.
- Gehrels, G. E., Valencia, V. A., and Ruiz, J.: Enhanced precision, accuracy, efficiency, and spatial resolution of U-Pb ages by 555 laser ablation-multicollector-inductively coupled plasma-mass spectrometry, Geochem. Geophys. Geosyst., 9, https://doi.org/10.1029/2007gc001805, 2008.
	- Gilbert, S. E., Olin, P., Thompson, J., Lounejeva, E., and Danyushevsky, L.: Matrix dependency for oxide production rates by LA-ICP-MS, J. Anal. At. Spectrom., 32, 638–646, https://doi.org/10.1039/C6JA00395H, 2017.
- Glorie, S., March, S., Nixon, A., Meeuws, F., O`Sullivan, G. J., Chew, D. M., Kirkland, C. L., Konopelko, D., and De Grave, 560 J.: Apatite U–Pb dating and geochemistry of the Kyrgyz South Tian Shan (Central Asia): establishing an apatite fingerprint for provenance studies, Geoscience Frontiers, https://doi.org/10.1016/j.gsf.2020.06.003, 2020.
	- Glorie, S., Mulder, J., Hand, M., Fabris, A., Simpson, A., and Gilbert, S.: Laser ablation (in situ) Lu-Hf dating of magmatic fluorite and hydrothermal fluorite-bearing veins, Geoscience Frontiers, 101629, https://doi.org/10.1016/j.gsf.2023.101629, 2023.





- 565 Guillong, M. and Günther, D.: Effect of particle size distribution on ICP-induced elemental fractionation in laser ablationinductively coupled plasma-mass spectrometry, J. Anal. At. Spectrom., 17, 831–837, https://doi.org/10.1039/B202988J, 2002.
- Günther, D., v. Quadt, A., Wirz, R., Cousin, H., and Dietrich, V. J.: Elemental Analyses Using Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry (LA-ICP-MS) of Geological Samples Fused with Li2B4O7 and Calibrated Without 570 Matrix-Matched Standards, Microchimica Acta, 136, 101–107, https://doi.org/10.1007/s006040170038, 2001.
	- Habib, E. A. E.: Geometric Mean for Negative and Zero Values, IJRRAS, 11, 2012.
	- Hall, J. W., Glorie, S., Reid, A. J., Boone, S. C., Collins, A. S., and Gleadow, A.: An apatite U–Pb thermal history map for the northern Gawler Craton, South Australia, Geoscience Frontiers, 9, 1293–1308, https://doi.org/10.1016/j.gsf.2017.12.010, 2018.
- 575 Hergenröder, R.: A model of non-congruent laser ablation as a source of fractionation effects in LA-ICP-MS, J. Anal. At. Spectrom., 21, 505–516, https://doi.org/10.1039/B600698A, 2006.
	- Hirata, T. and Nesbitt, R. W.: U-Pb isotope geochronology of zircon: evaluation of the laser probe-inductively coupled plasma mass spectrometry technique, Geochimica et Cosmochimica Acta, 59, 2491–2500, https://doi.org/10.1016/0016- 7037(95)00144-1, 1995.
- 580 Hogmalm, K. J., Zack, T., Karlsson, A. K. O., Sjöqvist, A. S. L., and Garbe-Schönberg, D.: In situ Rb–Sr and K–Ca dating by LA-ICP-MS/MS: an evaluation of N2O and SF6 as reaction gases, J. Anal. At. Spectrom., 32, 305–313, https://doi.org/10.1039/c6ja00362a, 2017.
- Horn, I., Rudnick, R. L., and McDonough, W. F.: Precise elemental and isotope ratio determination by simultaneous solution nebulization and laser ablation-ICP-MS: application to U–Pb geochronology, Chemical Geology, 164, 281–301, 585 https://doi.org/10.1016/S0009-2541(99)00168-0, 2000.
	- Horstwood, M. S. A., Foster, G. L., Parrish, R. R., Noble, S. R., and Nowell, G. M.: Common-Pb corrected in situ U–Pb accessory mineral geochronology by LA-MC-ICP-MS, J. Anal. At. Spectrom., 18, 837–846, https://doi.org/10.1039/B304365G, 2003.
- Horstwood, M. S. A., Košler, J., Gehrels, G. E., Jackson, S. E., McLean, N. M., Paton, C., Pearson, N. J., Sircombe, K. N., 590 Sylvester, P., Vermeesch, P., Bowring, J. F., Condon, D. J., and Schoene, B.: Community-Derived Standards for LA-ICP-MS U-(Th-)Pb Geochronology - Uncertainty Propagation, Age Interpretation and Data Reporting, Geostand. Geoanal. Res., 40, 311–332, https://doi.org/10.1111/j.1751-908X.2016.00379.x, 2016.
	- Karch, J.: Improving on Adjusted R-Squared, Collabra: Psychology, 6, 45, https://doi.org/10.1525/collabra.343, 2020.
- Kendall-Langley, L. A., Kemp, A. I. S., Grigson, J. L., and Hammerli, J.: U-Pb and reconnaissance Lu-Hf isotope analysis of 595 cassiterite and columbite group minerals from Archean Li-Cs-Ta type pegmatites of Western Australia, Lithos, 352–353, 105231, https://doi.org/10.1016/j.lithos.2019.105231, 2020.

Kirkwood, T. B. L.: Geometric Means and Measures of Dispersion, Biometrics, 35, 908–909, 1979.

- Košler, J., Wiedenbeck, M., Wirth, R., Hovorka, J., Sylvester, P., and Míková, J.: Chemical and phase composition of particles produced by laser ablation of silicate glass and zircon—implications for elemental fractionation during ICP-MS analysis, 600 J. Anal. At. Spectrom., 20, 402–409, https://doi.org/10.1039/B416269B, 2005.
	- Larson, K. P., Dyck, B., Shrestha, S., Button, M., and Najman, Y.: On the viability of detrital biotite Rb–Sr geochronology, Geochronology, 6, 303–312, https://doi.org/10.5194/gchron-6-303-2024, 2024.
		- Liu, Z., Wu, F., Guo, C., Zhao, Z., Yang, J., and Sun, J.: In situ U-Pb dating of xenotime by laser ablation (LA)-ICP-MS, Chin. Sci. Bull., 56, 2948–2956, https://doi.org/10.1007/s11434-011-4657-y, 2011.
- 605 Lloyd, J. C., Blades, M. L., Counts, J. W., Collins, A. S., Amos, K. J., Wade, B. P., Hall, J. W., Hore, S., Ball, A. L., Shahin, S., and Drabsch, M.: Neoproterozoic geochronology and provenance of the Adelaide Superbasin, Precambrian Res., 350, 105849, https://doi.org/10.1016/j.precamres.2020.105849, 2020.
	- Lloyd, J. C., Collins, A. S., Blades, M. L., Gilbert, S. E., and Amos, K. J.: Early evolution of the Adelaide Superbasin, Geosciences, 12, 154, https://doi.org/10.3390/geosciences12040154, 2022.
- 610 Lloyd, J. C., Preiss, W. V., Collins, A. S., Virgo, G. M., Blades, M. L., Gilbert, S. E., Subarkah, D., Krapf, C. B. E., and Amos, K. J.: Geochronology and formal stratigraphy of the Sturtian Glaciation in the Adelaide Superbasin, Geological Magazine, 1–24, https://doi.org/10.1017/S0016756823000390, 2023.
- Lloyd, J. C., Collins, A. S., Blades, M. L., Gilbert, S. E., Mulder, J. A., and Amos, K. J.: Mid- to Late Neoproterozoic Development and Provenance of the Adelaide Superbasin, τeκτoniκa, 2, 85–111, 615 https://doi.org/10.55575/tektonika2024.2.2.65, 2024.
	- Mank, A. J. G. and Mason, P. R. D.: A critical assessment of laser ablation ICP-MS as an analytical tool for depth analysis in silica-based glass samples, J. Anal. At. Spectrom., 14, 1143–1153, https://doi.org/10.1039/A903304A, 1999.
- Marillo-Sialer, E., Woodhead, J., Hergt, J., Greig, A., Guillong, M., Gleadow, A., Evans, N., and Paton, C.: The zircon 'matrix effect': evidence for an ablation rate control on the accuracy of U–Pb age determinations by LA-ICP-MS, J. Anal. At. 620 Spectrom., 29, 981–989, https://doi.org/10.1039/c4ja00008k, 2014.
	- McFarlane, C. R. M.: Allanite UPb geochronology by 193 nm LA ICP-MS using NIST610 glass for external calibration., Chemical Geology, 438, 91–102, https://doi.org/10.1016/j.chemgeo.2016.05.026, 2016.
	- McInnes, L., Healy, J., and Melville, J.: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, https://doi.org/10.48550/arXiv.1802.03426, 17 September 2020.
- 625 McLean, N. M., Bowring, J. F., and Gehrels, G.: Algorithms and software for U-Pb geochronology by LA-ICPMS, Geochemistry, Geophysics, Geosystems, 17, 2480–2496, https://doi.org/10.1002/2015GC006097, 2016.





- Mohammadi, N., Lentz, D. R., Thorne, K. G., Walker, J., Rogers, N., Cousens, B., and McFarlane, C. R. M.: U-Pb and Re-Os geochronology and lithogeochemistry of granitoid rocks from the Burnthill Brook area in central New Brunswick, Canada: Implications for critical mineral exploration, Geochemistry, 126087, https://doi.org/10.1016/j.chemer.2024.126087, 2024.
- 630 Norris, A. and Danyushevsky, L.: Towards Estimating the Complete Uncertainty Budget of Quantified Results Measured by LA-ICP-MS, 2018.
	- Olkin, I. and Pratt, J. W.: Unbiased Estimation of Certain Correlation Coefficients, The Annals of Mathematical Statistics, 29, 201–211, https://doi.org/10.1214/aoms/1177706717, 1958.
- O'Neill, H. S. C.: The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts, J. Petrol., 57, 635 1463–1508, https://doi.org/10.1093/petrology/egw047, 2016.
	- Paton, C., Woodhead, J. D., Hellstrom, J. C., Hergt, J. M., Greig, A., and Maas, R.: Improved laser ablation U-Pb zircon geochronology through robust downhole fractionation correction, Geochem. Geophys. Geosyst., https://doi.org/10.1029/2009GC002618, 2010.
- Paton, C., Hellstrom, J., Paul, B., Woodhead, J., and Hergt, J.: Iolite: Freeware for the visualisation and processing of mass 640 spectrometric data, J. Anal. At. Spectrom., 26, 2508–2518, https://doi.org/10.1039/C1JA10172B, 2011.
	- Payne, J. L., Hand, M., Barovich, K. M., and Wade, B. P.: Temporal constraints on the timing of high-grade metamorphism in the northern Gawler Craton: implications for assembly of the Australian Proterozoic, Aust. J. Earth Sci., 55, 623–640, https://doi.org/10.1080/08120090801982595, 2008.
- Pearson, J. W., Olver, S., and Porter, M. A.: Numerical methods for the computation of the confluent and Gauss hypergeometric 645 functions, Numer Algor, 74, 821–866, https://doi.org/10.1007/s11075-016-0173-0, 2017.
	- Polhill, J. G., Izquierdo, L. R., and Gotts, N. M.: What every agent-based modeller should know about floating point arithmetic, Environmental Modelling & Software, 21, 283–309, https://doi.org/10.1016/j.envsoft.2004.10.011, 2006.
- Roberts, N. M. W., Drost, K., Horstwood, M. S. A., Condon, D. J., Chew, D., Drake, H., Milodowski, A. E., McLean, N. M., Smye, A. J., Walker, R. J., Haslam, R., Hodson, K., Imber, J., Beaudoin, N., and Lee, J. K.: Laser ablation inductively 650 coupled plasma mass spectrometry (LA-ICP-MS) U–Pb carbonate geochronology: strategies, progress, and limitations, Geochronology, 2, 33–61, https://doi.org/10.5194/gchron-2-33-2020, 2020.
	- Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., 6, 461–464, https://doi.org/10.1214/aos/1176344136, 1978.
- Simpson, A., Gilbert, S., Tamblyn, R., Hand, M., Spandler, C., Gillespie, J., Nixon, A., and Glorie, S.: In-situ Lu Hf geochronology of garnet, apatite and xenotime by LA ICP MS/MS, Chem. Geol., 577, 655 https://doi.org/10.1016/j.chemgeo.2021.120299, 2021.
	- Sláma, J., Košler, J., Condon, D. J., Crowley, J. L., Gerdes, A., Hanchar, J. M., Horstwood, M. S. A., Morris, G. A., Nasdala, L., Norberg, N., Schaltegger, U., Schoene, B., Tubrett, M. N., and Whitehouse, M. J.: Plešovice zircon — A new natural reference material for U–Pb and Hf isotopic microanalysis, Chem. Geol., 249, 1–35, https://doi.org/10.1016/j.chemgeo.2007.11.005, 2008.
- 660 Solari, L. A., Ortega-Obregón, C., and Bernal, J. P.: U–Pb zircon geochronology by LAICPMS combined with thermal annealing: Achievements in precision and accuracy on dating standard and unknown samples, Chem. Geol., 414, 109– 123, https://doi.org/10.1016/j.chemgeo.2015.09.008, 2015.
- Spandler, C., Hammerli, J., Sha, P., Hilbert-Wolf, H., Hu, Y., Roberts, E., and Schmitz, M.: MKED1: A new titanite standard for in situ analysis of Sm–Nd isotopes and U–Pb geochronology, Chemical Geology, 425, 110–126, 665 https://doi.org/10.1016/j.chemgeo.2016.01.002, 2016.
	- Subarkah, D., Blades, M. L., Collins, A. S., Farkaš, J., Gilbert, S., Löhr, S. C., Redaa, A., Cassidy, E., and Zack, T.: Unraveling the histories of Proterozoic shales through in situ Rb-Sr dating and trace element laser ablation analysis, Geology, https://doi.org/10.1130/G49187.1, 2021.
- Tamblyn, R., Gilbert, S., Glorie, S., Spandler, C., Simpson, A., Hand, M., Hasterok, D., Ware, B., and Tessalina, S.: 670 Molybdenite Reference Materials for In Situ LA-ICP-MS/MS Re-Os Geochronology, Geostandards and Geoanalytical Research, n/a, https://doi.org/10.1111/ggr.12550, 2024.
	- Thompson, J. M., Meffre, S., Maas, R., Kamenetsky, V., Kamenetsky, M., Goemann, K., Ehrig, K., and Danyushevsky, L.: Matrix effects in Pb/U measurements during LA-ICP-MS analysis of the mineral apatite, J. Anal. At. Spectrom., 31, 1206– 1215, https://doi.org/10.1039/C6JA00048G, 2016.
- 675 Thompson, J. M., Meffre, S., and Danyushevsky, L.: Impact of air, laser pulse width and fluence on U–Pb dating of zircons by LA-ICPMS, J. Anal. At. Spectrom., 33, 221–230, https://doi.org/10.1039/C7JA00357A, 2018.
	- Ver Hoeve, T. J., Scoates, J. S., Wall, C. J., Weis, D., and Amini, M.: Evaluating downhole fractionation corrections in LA-ICP-MS U-Pb zircon geochronology, Chem. Geol., 483, 201–217, https://doi.org/10.1016/j.chemgeo.2017.12.014, 2018. Wiedenbeck, M., Hanchar, J. M., Peck, W. H., Sylvester, P., Valley, J., Whitehouse, M., Kronz, A., Morishita, Y., Nasdala,
- 680 L., Fiebig, J., Franchi, I., Girard, J. P., Greenwood, R. C., Hinton, R., Kita, N., Mason, P. R. D., Norman, M., Ogasawara, M., Piccoli, P. M., Rhede, D., Satoh, H., Schulz-Dobrick, B., Skår, O., Spicuzza, M. J., Terada, K., Tindle, A., Togashi, S., Vennemann, T., Xie, Q., and Zheng, Y. F.: Further Characterisation of the 91500 Zircon Crystal, Geostand. Geoanal. Res., 28, 9–39, https://doi.org/10.1111/j.1751-908X.2004.tb01041.x, 2004.
- van der Wolff, E. J.: Detrital Provenance and Geochronology of the Burra, Umberatana and Wilpena Groups in the Mount 685 Lofty Ranges, Honours Thesis, The University of Adelaide, Adelaide, South Australia, 2020.
	- Yang, C.-X., Santosh, M., Lloyd, J., Glorie, S., Anilkumar, Y., Anoop, K. S., Gao, P., and Kim, S.-W.: Breakup of the Neoarchean supercontinent Kenorland: Evidence from zircon and baddeleyite U-Pb ages of LIP-related mafic dykes in the Coorg Block, southern India, Geoscience Frontiers, 101804, https://doi.org/10.1016/j.gsf.2024.101804, 2024.





Zack, T., Stockli, D. F., Luvizotto, G. L., Barth, M. G., Belousova, E., Wolfe, M. R., and Hinton, R. W.: In situ U–Pb rutile 690 dating by LA-ICP-MS: 208Pb correction and prospects for geological applications, Contrib. Mineral. Petrol., 162, 515– 530, https://doi.org/10.1007/s00410-011-0609-4, 2011. Zhang, D. K.: dzhang314/MultiFloats.jl, 2024.