An algorithm for U–Pb geochronology by secondary ion mass spectrometry

Pieter Vermeesch[1]

[1]London Geochronology Centre, Department of Earth Sciences, University College London, London WC1E 6BT, United Kingdom

Correspondence: Pieter Vermeesch (p.vermeesch@ucl.ac.uk)

Abstract. Secondary Ion Mass Spectrometry (SIMS) is a widely used technique for in-situ U–Pb geochronology of accessory minerals. Existing algorithms for SIMS data reduction and error propagation make a number of simplifying assumptions that degrade the precision and accuracy of the resulting U–Pb dates. This paper uses an entirely new approach to SIMS data processing that introduces the following improvements over previous algorithms. First, it treats SIMS measurements as compositional data, using logratio statistics. This means that, unlike existing algorithms: (a) its isotopic ratio estimates are guaranteed to be strictly positive numbers; (b) identical results are obtained regardless of whether data are processed as normal ratios (e.g. ^{206}Pb/^{238}U) or reciprocal ratios (e.g. ^{238}U/^{206}Pb); and (c) its uncertainty estimates account for the positive skewness of measured isotopic ratio distributions. Second, the new algorithm accounts for the Poisson noise that characterises Secondary Electron Multiplier (SEM) detectors. By fitting the SEM signals using the method of maximum likelihood, it naturally handles low intensity ion beams, in which zero-count signals are common. Third, the new algorithm casts the data reduction process in a matrix format, and thereby captures all sources of systematic uncertainty. These include significant inter-spot error correlations that arise from the Pb/U–UO$_{(2)}$/U calibration curve. The new algorithm has been implemented in a new software package called `simplex`. `simplex` was written in `R` and can be used either online, offline or from the command line. The program can handle SIMS data from both Cameca and SHRIMP instruments.

1 Introduction

Secondary Ion Mass Spectrometry (SIMS) combines high sensitivity with high mass resolution (Williams, 1998). This allows the technique to obtain precise U–Pb dates on ng-sized samples whilst resolving isobaric interferences on ^{204}Pb to a degree that is currently unachievable by other techniques such as Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LAICPMS). There are some other differences between LAICPMS and SIMS as well. LAICPMS instrumentation is built by numerous manufacturers, whose data files are compatible with a rich ecosystem of data reduction codes such as `Iolite`, `Glitter` and `LADR`. This facilitates the intercomparison of different laboratories, different instrument designs and so forth. In contrast, the SIMS U–Pb world is dominated by just two manufacturers, whose SHRIMP (Sensitive High Resolution Ion Micro-Probe) and Cameca instruments use completely separate data reduction protocols. Most SHRIMP laboratories use `Squid`

(Ludwig, 2000; Bodorkos et al., 2020), which is incompatible with Cameca data. In contrast, Cameca data tends to be processed by in-house software such as M. J. Whitehouse's NordSIM spreadsheet, which are incompatible with SHRIMP data.

This paper introduces a unified algorithm for SIMS U–Pb data reduction that aims to address five problems with existing data reduction methods. The first three of these problems are:

1. Accuracy: existing algorithms give (slightly) different results depending on whether the raw data are processed as ^{206}Pb/^{238}U-ratios or as ^{238}U/^{206}Pb-ratios, say.

2. Precision: current data reduction routines produce symmetric confidence intervals, which are unrealistic for low intensity ion beams.

3. Systematic uncertainties: current data reduction protocols use a hierarchical error propagation approach, in which random uncertainties and systematic uncertainties are propagated separately. However such a clean separation is not always possible, and this can complicate higher order data processing steps such as isochron regression and averaging.

Sections 2 – 4 will provide further details about these problems, using synthetic examples. Section 5 will show that problems 1, 2 and 3 can be solved by treating the U–Pb system as a *compositional* data space, using logratio statistics. However logratio statistics does not solve the remaining two problems:

4. Backgrounds: background correction of low intensity signals such as ^{204}Pb sometimes exceeds 100%, producing physically impossible negative isotope ratios.

5. Zeros: Pb-isotopes are usually measured by Secondary Electron Multipliers (SEMs), which record ions as counts. For ^{204}Pb and other low intensity ion species, it is not uncommon to register zero counts during any given analytical cycle. This causes problems if the zero count appears in the denominator of an isotopic ratio.

These problems are discussed in more detail in Section 6. Section 7 addresses them by incorporating multinomial counting statistics into the compositional data framework. Section 12 applies the new data reduction paradigm to two datasets produced by Cameca and SHRIMP instruments, and Section 13 introduces a computer code called `simplex` that generated these results. Finally, further details about the implementation of the new algorithm are reported in an appendix to this paper.

2 Accuracy

The common Pb corrected ^{206}Pb/^{238}U-age (t) depends on the relative abundances of three nuclides, ^{204}Pb, ^{206}Pb and ^{238}U:

$$t = \frac{1}{\lambda_{238}} \ln \left(1 + \frac{\left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right] - \left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}}\right]_c}{\left[\frac{^{238}\text{U}}{^{204}\text{Pb}}\right]} \right) \tag{1}$$

where the subscript c marks the common lead composition, and λ_{238} is the decay constant of ^{238}U. Equation 1 does not depend on the absolute amounts of ^{204}Pb, ^{206}Pb and ^{238}U, but only on their ratios. Unfortunately, the statistical analysis of the ratios of

strictly positive numbers is full of potential pitfalls, as will be illustrated with an example that was inspired by McLean et al. (2016). Consider a simple dataset of ten synthetic U–Pb measurements:

^{238}U	215.9	208.9	212.4	186.3	217.8	196.7	216.4	171.8	216.0	200.1
^{206}Pb	18.45	12.40	21.35	62.22	21.35	45.08	26.65	75.88	29.02	11.40
^{204}Pb	0.1570	0.2870	0.1627	0.01425	0.1092	0.08175	0.06900	0.02250	0.04975	0.3850

Table 1. A synthetic U-Pb dataset, in arbitrary units (e.g, fmol, mV, pA or kHz).

Let us calculate the $^{206}Pb/^{238}U$-ratios for these data, which are needed to solve Equation 1. Comparing these ratios with their reciprocals yields two new sets of ten numbers:

$^{206}Pb/^{238}U$	0.085	0.059	0.101	0.334	0.098	0.229	0.123	0.442	0.134	0.057
$^{238}U/^{206}Pb$	11.70	16.85	9.95	2.99	10.20	4.36	8.12	2.26	7.44	17.55

The elementary rules of mathematics dictate that $1/(y/x) = x/y$ for any two real numbers x and y. In other words, the reciprocal of the reciprocal ratio ought to equal that ratio. Indeed, for our example it is easy to see that $1/0.085 = 11.70$ and so forth. However, when we take the arithmetic means of the (reciprocal) ratios:

$$\left(\overline{^{206}Pb/^{238}U}\right)_a = \sum_{i=1}^{10} \left(^{206}Pb/^{238}U\right)_i /10 = 0.166, \text{ and}$$
$$\left(\overline{^{238}U/^{206}Pb}\right)_a = \sum_{i=1}^{10} \left(^{238}U/^{206}Pb\right)_i /10 = 9.14$$

then we find that

$$\frac{1}{\left(\overline{^{206}Pb/^{238}U}\right)_a} = \frac{1}{0.166} = 6.01 \neq 9.14 = \left(\overline{^{238}U/^{206}Pb}\right)_a$$

So the reciprocal of the mean reciprocal ratio does *not* equal the mean of that ratio! This is a counter-intuitive and clearly wrong result. Unfortunately, current algorithms for SIMS data reduction average ratios using the arithmetic mean, or perform (linear) regression through ratio data, which causes similar problems (Ogliore et al., 2011). Inaccurate $^{206}Pb/^{238}U$-ratios inevitably result in inaccurate U–Pb dates, with the degree of inaccuracy scaling with the relative standard deviation of the measured data. Therefore, the numerical example shown in this section is deeply troubling for isotope geochemistry in general and SIMS U–Pb geochronology in particular.

3 Precision

Traditionally, the precision of isotopic data used in U–Pb geochronology has been calculated as symmetric confidence intervals. Unfortunately, this is fraught with similar problems as those discussed in Section 2. For example, take the arithmetic mean ($\bar{x}$) and standard deviation (s_x) of the $^{206}Pb/^{204}Pb$ ratios in Table 1, and construct a studentised 95% confidence interval for $\bar{x}$:

x	$\bar{x}$	s_x	$\bar{x} + \frac{s_x}{\sqrt{10}} t_9^{2.5}$	$\bar{x} + \frac{s_x}{\sqrt{10}} t_9^{97.5}$
^{206}Pb/^{204}Pb	978	1554	-134	2090

where t_{df}^{α} is the α-percentile of a t-distribution with df degrees of freedom (e.g., $t_9^{2.5} = -2.262$). Then the lower limit of the confidence interval is negative. This nonsensical result is yet another indication that there are some fundamental problems with the application of 'conventional' statistical operations to isotopic data. Negative isotope ratios can also arise from background corrections, but this will be further discussed in Section 6.

4   Systematic errors

The statistical uncertainty of analytical data can be classified into two components (Renne et al., 1998):

1. Random (or internal) errors are caused by electronic noise in the ion detectors, counting statistics, temporal variability of the background as a result of changes in the lab environment, etc. They are *independent* for different aliquots of the same sample, and can be quantified by taking replicate measurements. The standard error of these measurements ($\sigma/\sqrt{N}$ where σ is the standard deviation of N replicate measurements) is a measure of their *precision*. The standard error can be reduced to arbitrarily low levels by simply averaging more measurements (i.e., by increasing N). For example, the precision of SIMS ^{206}Pb/^{204}Pb-ratio measurements can be increased by simply extending the duration of the primary ion bombardment.

2. Systematic (or external) errors include the effects of decay constant uncertainty, the ^{206}Pb/^{238}U-ratio of reference materials etc. Getting these constants wrong causes *bias* in some or all of the measurements and thus affects the *accuracy* of the age determinations. As their name suggests, the systematic uncertainties are not independent but *correlated* between different aliquots and samples. They cannot be reduced by simple averaging.

Great care must be taken in choosing which sources of uncertainty should or should not be included in the error propagation. In some cases, inter-sample comparisons of SIMS U–Pb data may legitimately ignore systematic uncertainties. However, when comparing data acquired by different methods, both random and some systematic uncertainties must be accounted for. For example, when comparing U–Pb and ^{40}Ar/^{39}Ar data, the decay constant uncertainty must be propagated; and when comparing SIMS and TIMS U–Pb data, SIMS primary standard age uncertainty and the TIMS tracer uncertainty must be propagated. The conventional way to tackle inter-sample comparisons is called 'hierarchical' error propagation (Renne et al., 1998; Min et al., 2000; Horstwood et al., 2016). Under this paradigm, the random uncertainties are processed first, and the systematic uncertainties afterwards.

Hierarchical error propagation is straightforward in principle but not always in practice. Some processing steps are of a hybrid nature, including both systematic and random uncertainties. ^{206}Pb/^{238}U calibration for SIMS is a good example of this. ^{206}Pb/^{238}U-ratios are sensitive to elemental fractionation in SIMS analysis (see Section 10 for further details). These
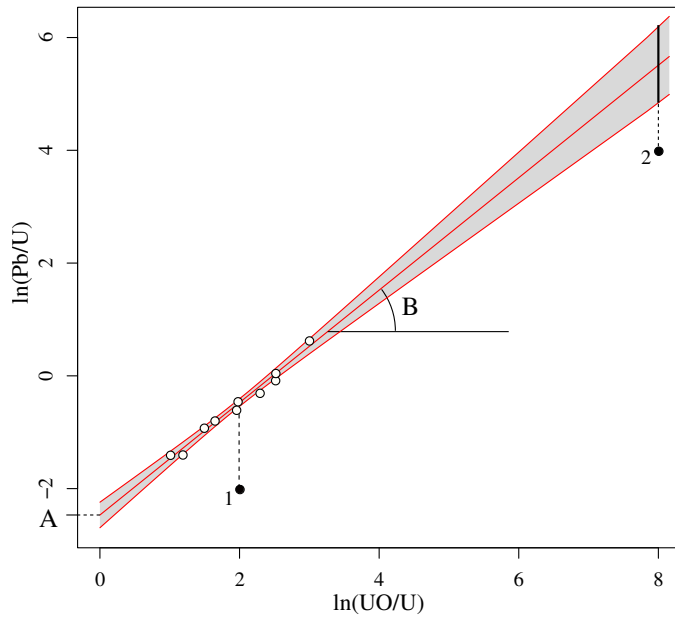
Figure 1. SIMS U–Pb calibration curve. White circles mark the isotopic measurements of the reference material, black circles those of two aliquots of the same sample. The uncertainty of the linear fit is shown as a 95% confidence interval (grey area). This uncertainty can be propagated into the Pb/U-composition of the sample. It is a systematic uncertainty in the sense that it affects both aliquots. But it does not do so to the same degree. The calibration uncertainty of aliquot 2 is greater than that of aliquot 1, due to its horizontal offset relative to the calibration data.

fractionation effects are captured by the following power law (Williams, 1998; Jeon and Whitehouse, 2015):

$$\ln\left[\frac{^{206}\mathrm{Pb}^+}{^{238}\mathrm{U}^+}\right]_m = A + B\ln\left[\frac{^{238}\mathrm{U}^{16}\mathrm{O}^+_{(2)}}{^{238}\mathrm{U}^+}\right]_m \tag{2}$$

where the subscript m stands for the measured signal ratio, which is generally different from the atomic ratio; and the subscript (2) refers to the fact that the calibration normally uses uranium oxide for SHRIMP and uranium dioxide for Cameca instruments. The atomic $^{206}\mathrm{Pb}/^{238}\mathrm{U}$-logratio of a sample is determined by (1) determining the intercept (A) and slope (B) of a reference material (r) of known Pb/U-ratio[1], and (2) using this calibration curve to estimate the equivalent Pb/U-logratio of the reference material corresponding to the UO$_{(2)}$/U-logratio of the sample (s). Then:

$$\ln\left[\frac{^{206}\mathrm{Pb}}{^{238}\mathrm{U}}\right]^s_a = \ln\left[\frac{^{206}\mathrm{Pb}}{^{238}\mathrm{U}}\right]^r_a + \ln\left[\frac{^{206}\mathrm{Pb}^+}{^{238}\mathrm{U}^+}\right]^s_m - A - B\ln\left[\frac{^{238}\mathrm{U}^{16}\mathrm{O}^+_{(2)}}{^{238}\mathrm{U}^+}\right]^s_m \tag{3}$$

where the subscript a stands for the estimated (for the sample) or known (for the reference material) atomic logratios. The analytical uncertainty of $\ln\left[^{206}\mathrm{Pb}/^{238}\mathrm{U}\right]^s_a$ depends on the analytical uncertainties of both the intercept (A) and slope (B) of the reference material. But it does not necessarily do so to the same degree for all aliquots. Analytical spots that have similar UO$_{(2)}$/U-ratios as the reference material will be less affected by the uncertainty of the reference material fit than spots that have very different UO$_{(2)}$/U-ratios (Figure 1). Hence it is not possible to make a clean separation between random and systematic uncertainties.

---

[1]If the reference material contains variable amounts of common lead, then the left hand side of Equation 2 needs to be corrected for that before applying the calibration.

5 U–Pb geochronology as a compositional data problem

Section 2 pointed out that the U–Pb age equation (Equation 1) does not depend on the absolute amounts of ^{204}Pb, ^{206}Pb and ^{238}U, but only on their relative abundances. Thus we could normalise the ^{204}Pb, ^{206}Pb and ^{238}U measurements of Table 1 to unity and plot them on a ternary diagram. The same is true for other geochronometers such as U–Th–He and ^{40}Ar/^{39}Ar
120 (Vermeesch, 2010, 2015). In mathematics, the ternary sample space is known as a two-dimensional *simplex*. Data that live within this type of space are called *compositional* data.

Ternary systems are common in igneous petrology (e.g., the A–F–M diagram) and sedimentary petrography (e.g., the Q–F–L diagram). Geologists have long been aware of the problems associated with averages, confidence regions, and linear regression in these closed dataspaces (Chayes, 1949, 1960). But a general solution to this conundrum was not found until the 1980s, when
125 the Scottish statistician John Aitchison published a landmark paper and book on the subject (Aitchison, 1982, 1986).

In this work, Aitchison proved that all the problems associated with the statistical analysis of compositional data can be solved by mapping those data from the simplex to a Euclidean space by means of an additive logratio (ALR) transformation. For example, given the ternary system $\{x, y, z\}$, we can define two new variables $\{u, v\}$ so that:

$$u = \ln(x/z) \text{ and } v = \ln(y/z) \tag{4}$$

130 In this space, Aitchison showed, one can safely calculate averages and confidence limits. Once the statistical analysis of the transformed data has been completed, the results can then be mapped back to the simplex by means of an inverse logratio transformation:

$$x = \frac{e^u}{1 + e^u + e^v}, y = \frac{e^v}{1 + e^u + e^v} \text{ and } z = \frac{1}{1 + e^u + e^v} \tag{5}$$

For example, the ^{204,6}Pb–^{238}U system of Table 1 can be mapped from the ternary diagram to a bivariate $\ln(^{204}$U/^{238}U)–
135 $\ln(^{206}$U/^{238}U)-space:

$u = \ln\left[\frac{^{206}\text{Pb}}{^{238}\text{U}}\right]$	-2.46	-2.82	-2.30	-1.10	-2.32	-1.47	-2.09	-0.82	-2.01	-2.87
$v = \ln\left[\frac{^{204}\text{Pb}}{^{238}\text{U}}\right]$	-7.23	-6.59	-7.17	-9.48	-7.60	-7.79	-8.05	-8.94	-8.38	-6.25

Alternatively, we could also use ^{206}Pb as the denominator isotope:

$u = \ln\left[\frac{^{238}\text{U}}{^{206}\text{Pb}}\right]$	2.46	2.82	2.30	1.10	2.32	1.47	2.09	0.82	2.01	2.87
$v = \ln\left[\frac{^{204}\text{Pb}}{^{206}\text{Pb}}\right]$	-4.77	-3.77	-4.88	-8.38	-5.28	-6.31	-5.96	-8.12	-6.37	-3.39

Calculating the average of the transformed data and mapping the results back to the simplex using the inverse logratio
140 transformation yields the *geometric* mean of the ratios:

$$\left(\overline{^{238}\text{U}/^{206}\text{Pb}}\right)_g = 7.58 = \frac{1}{0.13} = \frac{1}{\left(\overline{^{206}\text{Pb}/^{238}\text{U}}\right)_g}$$

which is an altogether more satisfying result than in Section 2. Moving on to the 95% confidence intervals of the ^{206}Pb/^{204}Pb-ratios, we first determine the conventional confidence limits for the logratios:

	u	$\bar{u}$	s_u	$\bar{u} + \frac{s_u}{\sqrt{10}}t_9^{2.5}$	$\bar{u} + \frac{s_u}{\sqrt{10}}t_9^{97.5}$
$\log(^{206}\text{Pb}/^{204}\text{Pb})$	5.72	1.66	4.53	6.91

After the inverse-logratio transformation, these values produce an asymmetric 95% confidence interval for the geometric mean $^{206}\text{Pb}/^{204}\text{Pb}$-ratio of 305^{+695}_{-212}. This interval contains only strictly positive values, solving the problem of Section 3. The ALR transformation of Equation 4 can easily be generalised to more than three components. For example, if ^{207}Pb is added to the mix, then the four-component $^{204|6|7}\text{Pb}$–^{238}U-system can be mapped to the three component $\ln(^{204|6|7}\text{Pb}/^{238}\text{U})$-space (Figure 2).

The compositional nature of isotopic data embeds a covariant structure into the very DNA of geochronology: in a K-component system, increasing the absolute amount of one of the components automatically lowers the relative amount of the remaining (K-1) components (Chayes, 1960). To deal with this phenomenon, it is customary in compositional data analysis to process data in matrix form, using the full covariance matrix. For example, four components yield three logratios, which require a 3×3 covariance matrix to describe the uncertainties and uncertainty correlations.

This approach is now widely used in sedimentary geology, geochemistry and ecology (e.g., Weltje, 2002; Vermeesch, 2006; Pawlowsky-Glahn and Buccianti, 2011), and has recently been adopted for geochronological applications as well (Vermeesch, 2010, 2015; McLean et al., 2016). The logratio covariance matrix approach is also uniquely suited to capture the systematic uncertainties (i.e. the inter-spot error correlations) that are produced by the SIMS U–Pb calibration procedure (Section 4). In that case, the covariance matrix must be expanded to accommodate multiple spots (e.g., Vermeesch, 2015). For example, the covariance structure of three spots in a four component system can be captured in a 9×9 matrix.

In conclusion, the logratio transformation solves the statistical woes described in Sections 2, 3 and 4 of this paper. However there are two additional problems that require further remediation.

6   Background correction

Many mathematical operations are easier in logarithmic space than in linear space: multiplication becomes addition, division becomes subtraction, and exponentiation becomes multiplication. These mathematical operations are very common in mass spectrometer data processing chains (Vermeesch, 2015). However there are exceptions. For example, background correction does not involve the multiplication but subtraction of two signals. For low intensity ion beams such as ^{204}Pb it is possible, by chance, that the background exceeds the signal. This results in negative values of which one cannot take the logarithm.

The subtraction problem can be solved by using a different logarithmic change of variables:

$$\beta_x^y \equiv \ln\left(\frac{y-b}{x-b}\right) \tag{6}$$

where x and y are the signals and b is the background. The infinite space of β_x^y covers all possible values of x, y and b for which $x > b$ and $y > b$. Thus, background correction should be done in β-space, given an appropriate error model as described in Section 7.
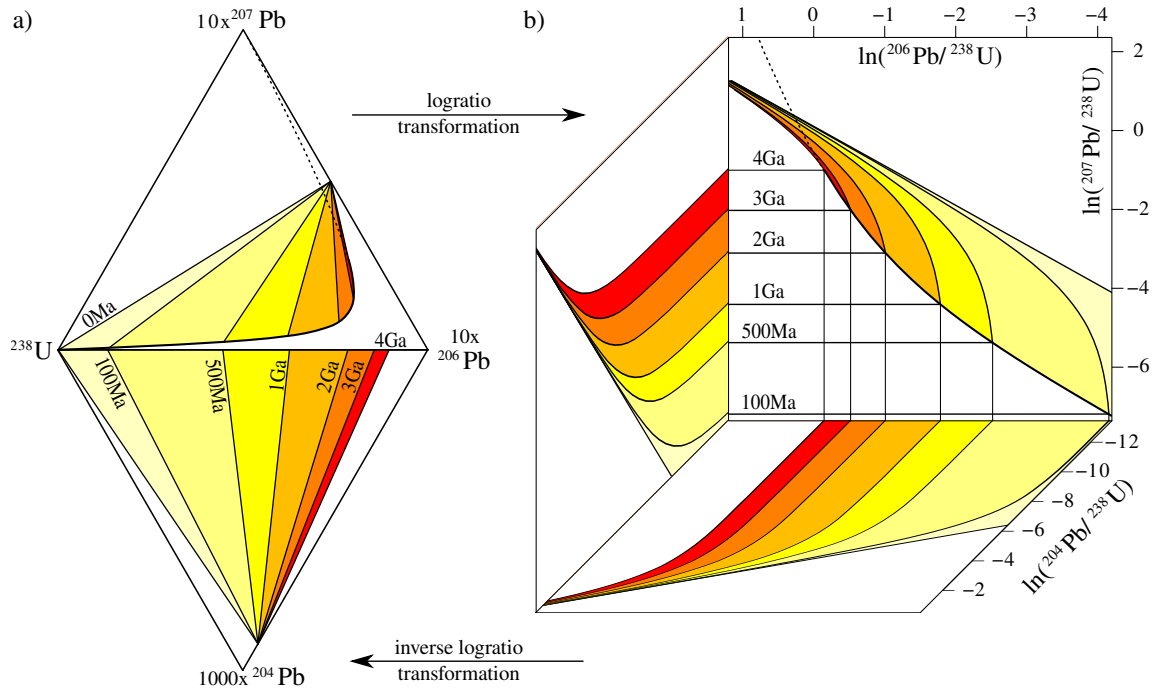
Figure 2. The U–Pb age equation (a) projected onto the four-component simplex, and (b) mapped to a three-dimensional Euclidean logratio space. ^{235}U is omitted from the diagrams because it exists in a constant ratio to ^{238}U (^{238}U/^{235}U=137.818, Hiess et al., 2012). The boundaries between the coloured regions mark mixing lines between radiogenic and inherited end-member compositions, assuming common Pb ratios of $[^{207}Pb/^{204}Pb]_c = 10$ and $[^{206}Pb/^{204}Pb]_c=9$. The mixing lines define isochron surfaces that are linear in the simplex and curved in logratio space. Rotating the logratio plot 90° clockwise produces a logarithmic version of the familiar Wetherill concordia diagram. Concordant $^{206}Pb/^{238}U$ and $^{207}Pb/^{235}U$ compositions are marked by a thick black line from 0 to 4Ga, and by a dotted line beyond 4Ga. This paper makes the case that U-Pb data processing is best done in logratio space, because this 'liberates' the data from the geometric constraints of the simplex, producing symmetric uncertainty distributions and more accurate results.

One caveat to this background correction method is that it does not account for isobaric interferences, which may result
175   in 'overcounted' signals. The high mass resolution of SIMS instruments removes most but not all isobaric interferences. For example, spurious HfSi, REE dioxide, or long-chain hydrocarbon ions can interfere with ^{204}Pb, which is generally the rarest isotopic species detected. If unaccounted for, these interferences lead to the proportion of non-radiogenic Pb being overestimated (and the proportion of radiogenic Pb underestimated), resulting in excessive common Pb corrections, and underestimated dates. The accuracy of the background measurements can be monitored via the use of isotopically homogeneous reference ma-
180   terials (Black, 2005). A correction can then be applied by choosing the 'session blank' that brings the common-Pb corrected $^{207}Pb/^{206}Pb$-ratios in alignment with the reference values.

Another issue arises when SIMS signals are registered by Secondary Electron Multiplier (SEM) detectors. These record ion beams as discrete counts, i.e. as integers, which are incompatible with the logratio transformation. For example, it is not uncommon for SEM detectors to register zero counts for low intensity ion beams such as ^{204}Pb. These zero counts blow up the logratio transformation, because $\ln(0) = -\infty$.

This and other issues are diagnostic of a fundamental difference between compositional data and counting data that has been previously recognised and solved in fission track dating (Galbraith, 2005) and in sedimentary petrography (Vermeesch, 2018b). The same solutions can be applied to mass spectrometric count data in general, and to U–Pb geochronology in particular (Section 7).

7 Dealing with count data

Standard data reduction procedures for geochronology assume normally distributed residuals. In compositional data analysis, these are replaced by logistic normal distributions. However, neither the normal nor the logistic normal distribution are perfectly suited to dealing with discrete count data. The multinomial distribution is a simple alternative that seems more appropriate for the task at hand. Before we proceed, let us define the following variables:

- ϕ_x and ϕ_b: the normalised true ion beam intensities (in counts per second) of mass x (from a set of monitored masses $\mathtt{x}$) and the normalised background signal, respectively, so that

$$\phi_b + \sum_{x \in \mathtt{x}} \phi_x = 1 \tag{7}$$

- d_x, d_b: the dwell times of mass x and the background b

- θ_x and θ_b: the normalised expected beam *counts* of the ions and the background, so that

$$\theta_x = \frac{\phi_x d_x}{\phi_b d_b + \sum_{y \in \mathtt{x}} \phi_y d_y} \tag{8}$$

and

$$\theta_b + \sum_{x \in \mathtt{x}} \theta_x = 1 \tag{9}$$

Then the probability of observing n_4 counts at mass 204, n_6 counts at mass 206 and n_b counts of background is given by

$$p(n_4, n_6, n_b | \theta_4, \theta_6, \theta_b) = \frac{(n_4 + n_6 + n_b)!}{n_4! n_6! n_b!} \theta_4^{n_4} \theta_6^{n_6} \theta_b^{n_b} \tag{10}$$

Whereas the observations n_4, n_6 and n_b are integers, the parameters θ_4, θ_6 and θ_b are decimal numbers that are constrained to a constant sum. In other words, they belong to the simplex. Thus, we can map the three multinomial parameters to two

logratio parameters, thereby establishing a natural link between counting data and compositional data. For example:

$$\beta_6^4 \equiv \ln\left(\frac{\phi_4 - \phi_b}{\phi_6 - \phi_b}\right) = \ln\left(\frac{\theta_4/d_4 - \theta_b/d_b}{\theta_6/d_6 - \theta_b/d_b}\right) \text{ and} \tag{11}$$

$$\beta_6^b \equiv \ln\left(\frac{\phi_b}{\phi_6 - \phi_b}\right) = \ln\left(\frac{\theta_b/d_b}{\theta_6/d_6 - \theta_b/d_b}\right) \tag{12}$$

210   β_6^4 and β_6^b can be estimated from n_4, n_6 and n_b by maximising the likelihood defined by Equation 10.

The normal and logistic normal distributions are controlled by two sets of parameters: location parameters and shape parameters. In the case of the normal distribution, the location parameter is the mean and the shape parameter is the standard deviation (or covariance matrix). In contrast, the multinomial distribution has only one set of (θ) parameters. The precision of multinomial counts is governed by the number of observed counts ($\sigma[n] = \sqrt{n}$). More sophisticated models are possible when
215   the observed dispersion of the data exceeds that which is expected from the multinomial counting statistics (e.g., Galbraith and Laslett, 1993; Vermeesch, 2018b). However in this paper we will assume that such *overdispersion* is absent from the reference materials, and from the single-spot analyses.

8   Dead-time correction

It takes a few tens of nanoseconds for a secondary electron multiplier to record the arrival of an ion. During this 'dead-time',
220   the detector is unable to register the arrival of additional ions. This phenomenon can significantly bias isotope ratio estimates that include ion beams of contrasting intensity. Fortunately, the dead-time effect can be easily corrected. It suffices that the dwell times are adjusted by the cumulative amount of time that the detectors were incapacitated. Let d'_x be the 'effective dwell time' of ion beam x:

$$d'_x = d_x - n_x \delta_x \tag{13}$$

225   where δ_x is the (non extended) dead-time of the detector that measures x. Then the expected normalised beam counts can be re-defined as:

$$\theta'_x = \frac{\phi_x d'_x}{\phi_b d'_b + \sum_{y \in \mathsf{x}} \phi_y d'_y} \tag{14}$$

and the normalised beam intensities as

$$\phi'_x = \frac{\theta_x/d'_x}{\theta_b/d'_b + \sum_{y \in \mathsf{x}} \theta'_y/d_y} \tag{15}$$

230   The difference between this approach and existing SIMS data reduction approaches is that conventional data reduction algorithms apply the dead-time correction to the raw data, before calculating isotopic data; whereas the new method applies the dead-time correction to θ'_x and ϕ'_x, which are unknown parameters that must be estimated from the data.

9   Within-spot drift correction

Thus far we have assumed that all ions are measured synchronously, which is the case in multicollector configurations. However in single collector experiments, the measurements are made asynchronously. This can cause biased results if the signals drift over time. In SHRIMP data processing, it is customary to correct this drift by normalising to a secondary beam monitor (SBM) signal[2], followed by double interpolation of numerator and denominator isotopes (Bodorkos et al., 2020; Dodson, 1978). A unified data reduction algorithm for SHRIMP and Cameca instruments requires a different approach, in which the time dependency of the signals is parameterised using a log-linear model. For Faraday detectors:

$$n_x^i = \text{bkg} + \mathcal{N}(\exp[\alpha_x + \gamma_X \tau_x^i], \sigma^2) \tag{16}$$

where n_x^i is the ion beam intensity of the ith integration for mass x of element X evaluated at time τ_x^i, σ is the standard deviation of the normally distributed data scatter around the best fit line, and 'bkg' is the background signal. This is usually a nominal value for Cameca instruments and an actual set of measurements (n_b^i) for SHRIMP data. Note that the intercept parameters (α_x) are ionic mass-specific, whereas the slopes (γ_X) are element-specific. See the Appendix for an example. For SEM detectors, the scatter of the data around the log-linear fit is controlled by Poisson shot noise:

$$n_x^i \sim \text{bkg} + \text{Pois}\big(\exp[\alpha_x + \gamma_X \tau_x^i] d_x'\big) \tag{17}$$

The background-corrected signal ratio (evaluated at τ_x^i) for two ion beams x and y of different elements X and Y can then be drift corrected as follows:

$$^i\beta_x^y \equiv \ln\left(\frac{\phi_y^i - \phi_b^i}{\phi_x^i - \phi_b^i}\right) + \gamma_Y\left(\tau_x^i - \tau_y^i\right) \tag{18}$$

where ϕ_x^i and ϕ_y^i are the dead time corrected normalised beam intensities for the ith integration of masses x and y, respectively, and ϕ_b^i is the corresponding background value. See the Appendix for further details.

10   Fractionation

Mass spectrometer signals are recorded in Volt (for Faraday detectors) or Hertz (or secondary electron multipliers). The age equation, however, requires atomic ratios. In general, signal ratios do not equal atomic ratios, because they are affected by two types of fractionation:

1. Mass-dependent fractionation: The Pb-isotopes span a range of four mass units, with ^{208}Pb being 2% heavier than ^{204}Pb. Both the production and detection efficiency of secondary ions vary with atomic mass, and significant errors can potentially occur if the resulting mass fractionation is uncorrected for. Mass fractionation can be quantified by comparing the measured signal ratios of a reference material with its known isotopic ratio. This is easy to do in a log-ratio context (Vermeesch, 2015).

---

[2]SBM normalisation is not a magic bullet. For example, whereas the ^{206}Pb signal of SHRIMP spot `M127.1.2` rises with time (Figure 3b), its corresponding SBM signal drops.

2. Elemental fractionation: The fractionation between the Pb-isotopes is caused by (slight) differences in their physical properties, i.e. their mass. As briefly mentioned in Section 4, much stronger fractionation effects tend to occur between the isotopes and Pb and U, because they are not only physically, but also chemically different. These chemical differences affect the complex processes that occur when the primary ion beam interacts with the target material (Williams, 1998).

In the context of SIMS U–Pb geochronology, mass-dependent fractionation is commonly ignored (but not always, e.g., Stern et al., 2009), because the most important isochemical ratio is that between ^{206}Pb and ^{207}Pb, which lie within 0.5% mass units of each other. This is unresolvable given typical analytical uncertainties. The mass fractionation is greater for ^{204}Pb, but so it its analytical uncertainty. Therefore, the atomic ^{204}Pb/^{206}Pb, ^{207}Pb/^{206}Pb and ^{208}Pb/^{206}Pb ratios can be directly estimated from the (drift corrected) ^{204}Pb/^{206}Pb, ^{207}Pb/^{206}Pb and ^{208}Pb/^{206}Pb signal ratios. This is not the case for the ^{206}Pb/^{238}U and ^{208}Pb/^{232}Th-ratios, which are affected by strong elemental fractionation effects. This fractionation expresses itself in two ways:

1. Within-spot fractionation

Over the course of a SIMS spot analysis, the Pb/U and Pb/Th ratio changes as a function of time. This elemental fractionation can be modelled using a log-linear model that is similar to that used for the within-spot drift correction:

$$^{i}\beta_x^y = {}^{0}\beta_x^y + \gamma_X^Y \tau_x^i \tag{19}$$

where $^{0}\beta_x^y$ is the inferred logratio of the background-corrected signals at 'time zero', which can be found using the method of maximum likelihood (see Appendix). Note that $\gamma_X^Y = 0$ if $X = Y$ and, if the sensitivity drift of the data is smooth and closely follows the trends defined by Equations 16 and 17, then $^{0}\beta_x^y \approx \alpha_y - \alpha_x$ and $\gamma_X^Y \approx \gamma_Y - \gamma_X$.

Using Equation 19, the isotopic logratios can be interpolated (or extrapolated) to any point in time (τ):

$$^{\tau}\beta_x^y = {}^{0}\beta_x^y + \gamma_X^Y \tau \tag{20}$$

The most precise values of $^{\tau}\beta_x^y$ are obtained when τ is chosen in the middle of the analytical sequence. These values can be used for subsequent calculations. Alternatively, we can also use the time-zero intercepts $^{0}\beta_x^y$.

2. Between-spot fractionation

The Pb/U and Pb/Th signal ratios may vary between adjacent spots on the same isotopically homogenous reference material. This fractionation obeys the power law relationship given by Equation 2. Expressing this formula in terms of corrected signal ratios:

$$\ln\left[\frac{(\phi_6' - \phi_b') - (\phi_4' - \phi_b')(6/4)_c}{\phi_u' - \phi_b'}\right] = A + B \ln\left[\frac{\phi_o' - \phi_b'}{\phi_u' - \phi_b'}\right] \tag{21}$$

where $(6/4)_c$ stands for the ^{206}Pb/^{204}Pb-ratio of the common Pb (see the footnote of Section 4). Recasting Equation 21 in terms of the interpolated logratio estimates:

$$\ln\left[\exp({}^{\tau}\beta_u^6) - \exp({}^{\tau}\beta_u^4)(6/4)_c\right] = A + B \, {}^{\tau}\beta_u^o \tag{22}$$

where 'o' stands for the uranium oxide (^{238}U^{16}O$_2^+$ or ^{238}U^{16}O$^+$) and 'u' stands for ^{238}U$^+$.

11 U–Pb age calculation

Having applied Equation 22 to a reference material with known atomic $^{206}\mathrm{Pb}/^{238}\mathrm{U}$-ratio $\left[^{206}\mathrm{Pb}/^{238}\mathrm{U}\right]_a^r$, the atomic $^{206}\mathrm{Pb}/^{238}\mathrm{U}$-ratio of the sample is given by

$$\ln\left[\frac{^{206}\mathrm{Pb}}{^{238}\mathrm{U}}\right]_a^s = \ln\left[\frac{^{206}\mathrm{Pb}}{^{238}\mathrm{U}}\right]_a^r + {^\tau}\beta_u^6(sm) - A - B\,{^\tau}\beta_u^o(sm) \tag{23}$$

where ${^\tau}\beta_u^6(sm)$ and ${^\tau}\beta_u^o(sm)$ are the interpolated logratio estimates of the sample. The $^{206}\mathrm{Pb}/^{238}\mathrm{U}$-age is then obtained by plugging $\left[^{206}\mathrm{Pb}/^{238}\mathrm{U}\right]_a^s$ into the age equation. Uncertainties are obtained by standard error propagation (see Appendix).

12 Examples

The following paragraphs will illustrate the SIMS U–Pb data reduction process using two datasets:

1. Dataset 1 was acquired by Dr. Yang Li at IGG-CAS Beijing, using a Cameca 1280HR instrument with five SEMs and two Faraday detectors that was run in single collector mode, using an SEM for all mass-stations. The dataset uses Temora2 zircon (416.8±1.1 Ma, Black et al., 2004) as a reference material and 91500 zircon (1062.4±0.2 Ma, Wiedenbeck et al., 1995) as a sample. Measurements consist of seven cycles through a set of 11 mass-stations per single spot measurement, for $^{90}\mathrm{Zr}_2\mathrm{O}$ (0.48 second dwell time per cycle), $^{92}\mathrm{Zr}_2\mathrm{O}$ (0.08 s), mass 200.5 (background, 4.00 s), $^{94}\mathrm{Zr}_2\mathrm{O}$ (0.32 s), $^{204}\mathrm{Pb}$ (4.96 s), $^{206}\mathrm{Pb}$ (2.96 s), $^{207}\mathrm{Pb}$ (6.00 s), $^{208}\mathrm{Pb}$ (2.00 s), $^{238}\mathrm{U}$ (2.96 s), ThO_2 (2.96 s), and UO_2 (2.96 s).

2. Dataset 2 was acquired by Dr. Simon Bodorkos at Geoscience Australia using a single collector SHRIMP-II instrument, employing an SEM for all mass-stations. This dataset also uses Temora2 as a reference material, and 91500 zircon and OG.1 (3440.7±3.2 Ma, Stern et al., 2009) and M127 (524.36±0.16 Ma, Nasdala et al., 2016) as samples. Measurements consist of six cycles through a set of 10 mass-stations per single spot measurement, for $^{90}\mathrm{Zr}_2\mathrm{O}$ (2.0 second dwell time per cycle), $^{204}\mathrm{Pb}$ (20 s), mass 204.04091 (background, 20 s), $^{206}\mathrm{Pb}$ (15 s), $^{207}\mathrm{Pb}$ (40 s), $^{208}\mathrm{Pb}$ (5 s), $^{238}\mathrm{U}$ (5 s), ThO (2 s), UO (2 s), and UO_2 (2 s).

Figure 3 shows the time resolved SEM counts of one representative spot measurement for each dataset. Side-by-side comparison of these two datasets reveals some interesting similarities and differences. All of the high intensity signals exhibit clear transient behaviour, which is caused by changes in oxygen availability that occur during primary ion bombardment (Magee et al., 2017). The transience of the individual SEM signals biases the isotopic ratios. For example, there are 109 seconds between the $^{204}\mathrm{Pb}$ and $^{208}\mathrm{Pb}$ measurements in each SHRIMP cycle. During these 109 seconds, the $^{208}\mathrm{Pb}$ signal drops on average by 2%, resulting in an equivalent bias of the $^{204}\mathrm{Pb}/^{208}\mathrm{Pb}$ ratio (Section 9). A similar drop per cycle is observed for the Cameca data.

However, there is a key difference between the Cameca and SHRIMP datasets. For the Cameca data, the U and Pb signals both drift in the same direction, resulting in positive slopes for the example of Figure 3. But for the SHRIMP data, the U- and Pb-drifts act in opposite directions: the U-signal exhibits an increase in sensitivity with time, whereas the Pb-signals decrease
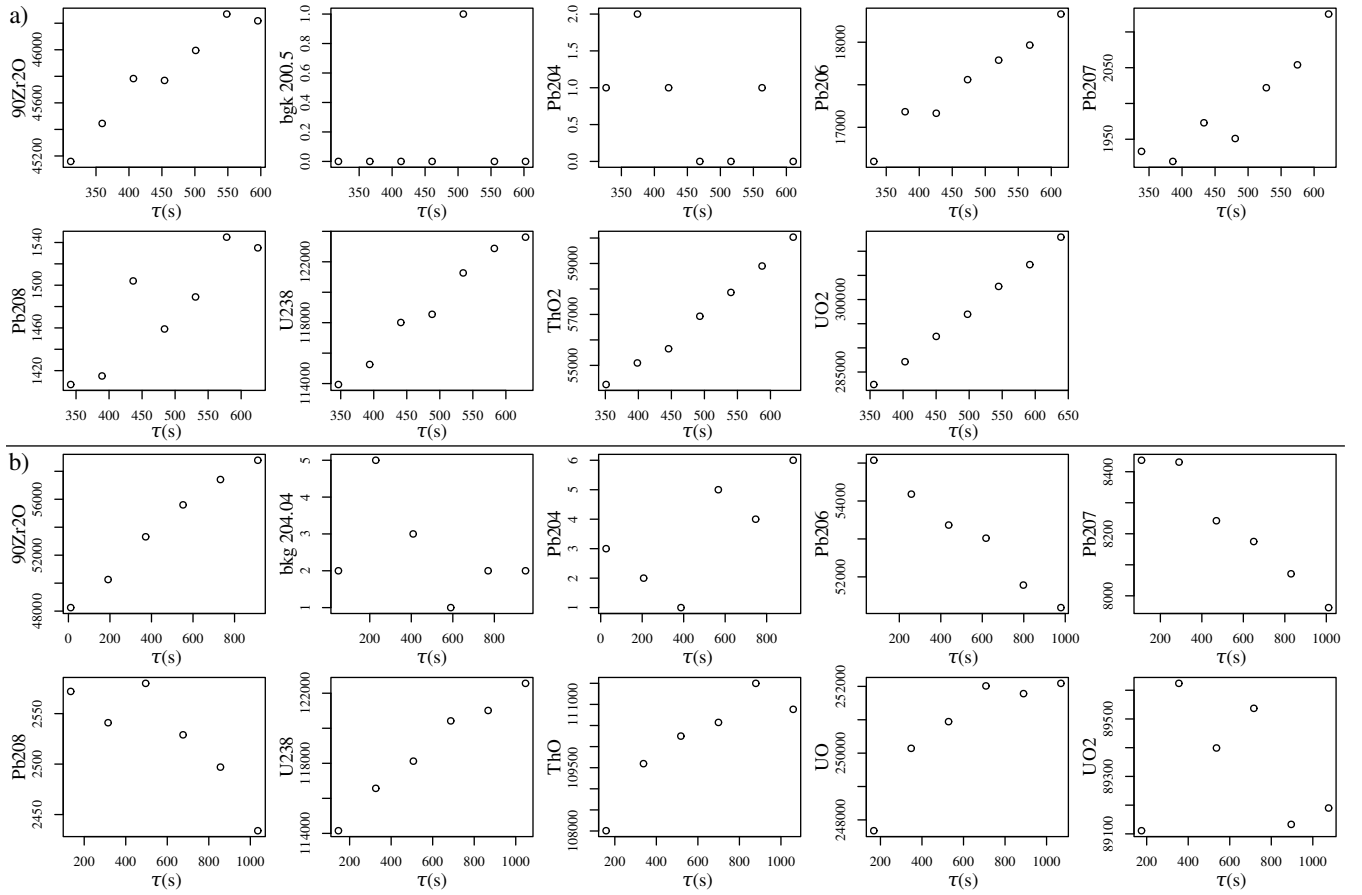
Figure 3. Time resolved signals (counts) of (a) Temora2 (spot `Tem@6`) analysed by a Cameca 1280HR instrument at IGG-CAS Beijing; (b) M127 zircon (spot `M127.1.2`) analysed by a SHRIMP-II instrument at Geoscience Australia.

in intensity with time. This marked difference in behaviour between the two instruments reflects a difference in their design, causing a difference in the energy window of the secondary ions analysed (Ireland and Williams, 2003).

For Faraday detectors, the within-spot drift correction uses a generalised linear model with a lognormal link function (Equation 16). For SEM data, it uses a Poisson link function (Equation 17). In either case, the model enforces strictly positive isotopic

325 abundances. Isotopes of the same element (such as $^{204|6|7|8}$Pb) have the same slope parameter, but different intercepts. Isotopes of different elements are free to have different slopes (Figure 4).

Figure 5 applies another log-linear function (Equation 19) to model the within-spot fractionation of Temora2 spot 11. This function models the drift-corrected logratios as a linear function of analysis time. The slopes of the log-linear functions are a function of the within-spot fractionation between the numerator and denominator elements. Because there is no fractionation

330 between two isotopes of the same element, the slope of the Pb/Pb ratios is zero.
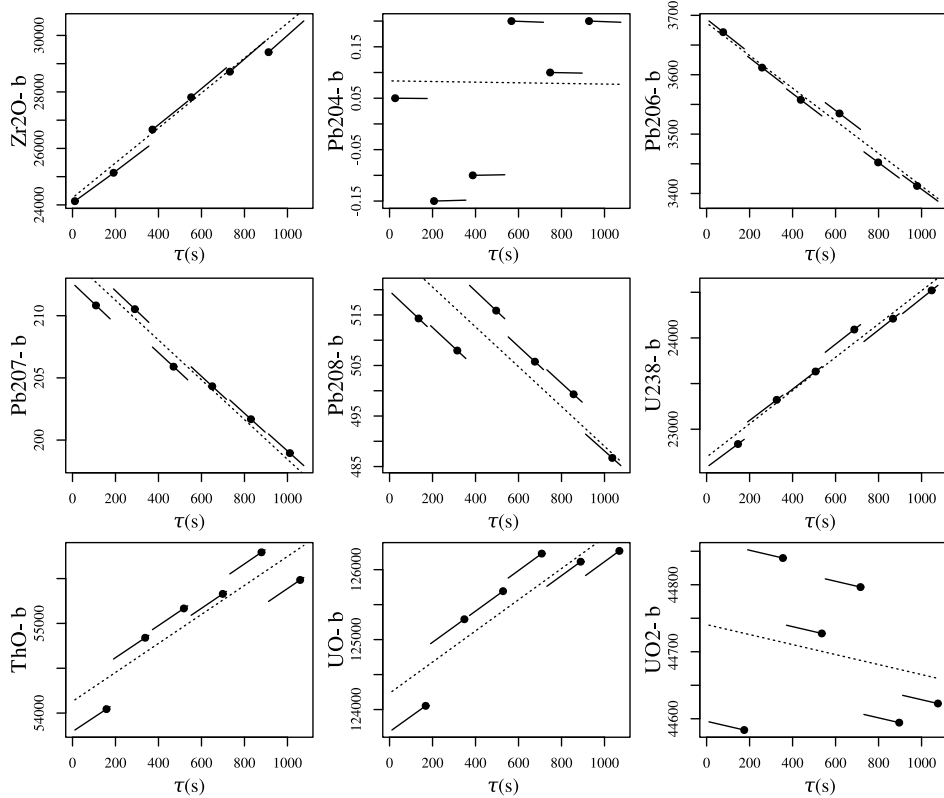
Figure 4. Within-spot drift correction of the background-corrected SHRIMP data from Figure 3. Dotted lines are log-linear functions whose element-specific slopes (γ_X in Equation 17) are used for the drift correction but for no other purpose. Solid lines mark the duration of each mass spectrometer cycle, with the black dots representing the starting point of each individual mass station within the cycles. The solid lines are parallel to the dotted lines (in log space) and express how the isotope ratio fits can be translated to match the asynchronous mass spectrometer signals. Vertical axes have units of counts per second.

The ability of logratio statistics to avoid negative ratios is apparent from the first panel of Figure 5. Even though some of the ratios of the background-corrected signals are zero or negative (because the background exceeded the signal), the generalised linear fit is strictly positive. The natural ability of compositional data analysis to rule out negative ratios avoids many problems further down the data processing chain.

335   The right hand side of Figure 5 maps the four (log)ratios back to five equivalent raw signals (one for each isotope). The last two panels of the figure show how the logratio approach manages to effectively capture subtle fluctuations of the U and UO signal intensities.
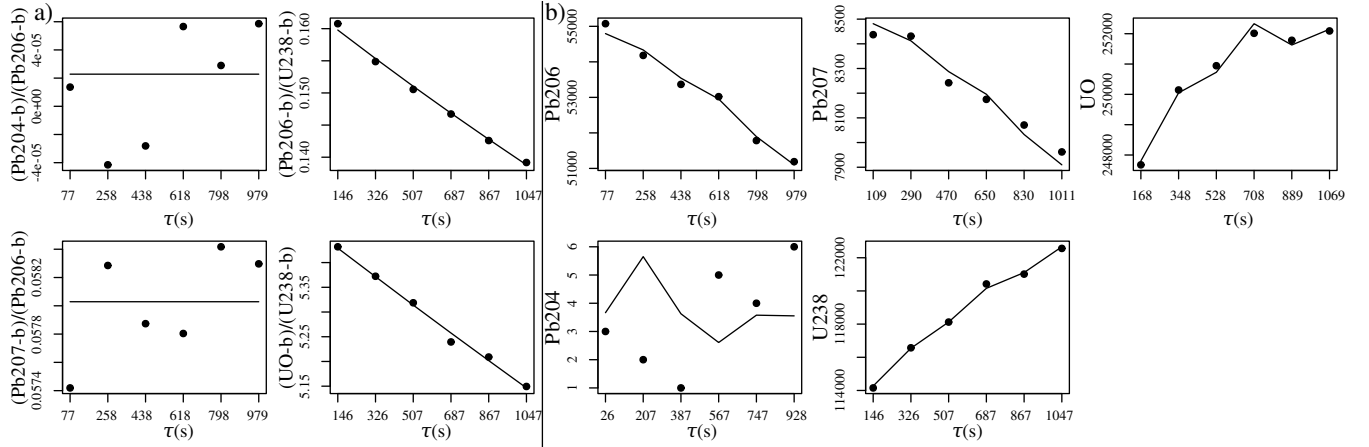
Figure 5. a) Background- and drift-corrected ratio fits to the SHRIMP data from Figure 3, obtained using the generalised linear model of Equation 19. The slope parameter (γ_X^Y in Equation 19) is zero for isotopes of the same element and non-zero for isotopes of different elements. Vertical axes are unitless. b) The four (log)ratio fits can be converted back to five isotope signal fits using the inverse logratio transformation. Multiplying the normalised ion beam intensity fits with the total number of counts per sweep allows direct comparison with the raw measurements, which are shown as filled circles. Vertical axes have units of counts. Note that the measurements have Poisson uncertainties, which scale with the square root of the values.

The logratio intercepts obtained by Equation 19 form a linear array of calibration data (Equation 22). Figure 6a fits a straight line through these points using the linear regression algorithm of York et al. (2004). Alternatively, instead of fitting a calibration

340   line through the logratio intercepts ($\tau = 0$, Figure 6a), it is also possible to interpolate or extrapolate the logratio composition to any other point in time. For example, the green ellipses in Figure 6b show the inferred logratio compositions at 544 seconds (i.e., $\tau = 544$), which represents the midpoint of the analyses. The slope of this calibration line is notably different than that obtained by fitting a line through the compositions at 0 seconds. This change in slope reflects the different mechanisms that are responsible for elemental fractionation within and between SIMS spots.

345   Figure 7 applies the Pb/U-calibration (Equation 23) to 91500 zircon, using the Temora2 data for calibration. It shows only the purely random errors, i.e., ignoring the uncertainty of the calibration. Including the calibration errors does not only inflate the uncertainties, but also causes inter-spot error correlations (Figure 1). To demonstrate this phenomenon, let us revisit the
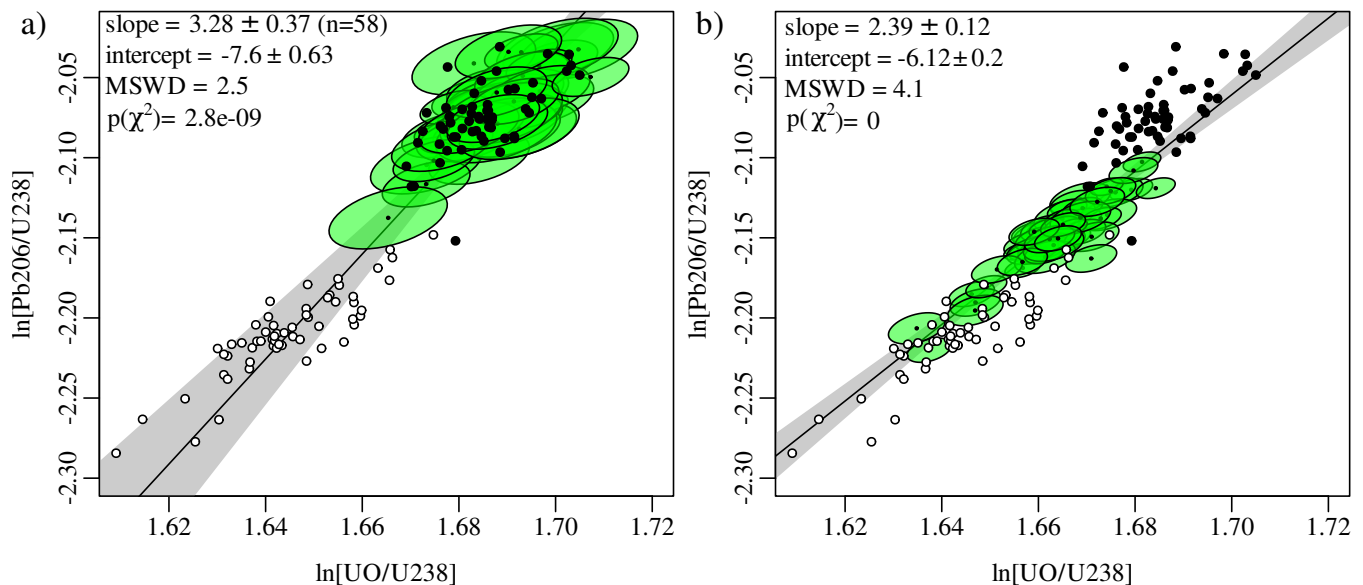
Figure 6. a) U–Pb calibration curve for the Temora2 SHRIMP data using the time-zero ($\tau = 0$) intercepts (green ellipses). Black and white dots mark the first and last cycles of each analysis, respectively. b) The logratio data and calibration fit for the same data, but evaluated at the midpoint ($\tau = 544$), resulting in a more precise calibration. All uncertainties are shown at 95% confidence.

Cameca data and swap the sample and reference materials around. Thus, we use 91500 for the calibration curve, and Temora2 as a sample (Figure 8). Table 2 shows the uncertainty budget of four selected aliquots from this sample.

	$\ln[^{206}\text{Pb}/^{238}\text{U}]$	int. unc. (%)	tot. unc. (%)	$r[*,b]$	$r[*,c]$	$r[*,d]$
a	-2.692	0.32	0.70	0.62	0.36	-0.34
b	-2.685	0.31	0.44		0.36	-0.17
c	-2.693	0.30	0.36			0.019
d	-2.719	0.45	0.56			

Table 2. Uncertainty budget of the four Temora2 zircon analyses highlighted in Figure 8. The first two data columns show the calibrated $^{206}\text{Pb}/^{238}\text{U}$ logratios and their standard errors ignoring the uncertainty of the calibration fit (i.e., using internal uncertainties only). The third column shows the total error including the external uncertainty associated with the calibration fit. The upper triangular matrix shown in the remaining three columns contain the (total) error correlations of the four aliquots.

350 Propagating the calibration curve uncertainty increases the error estimates (Table 2) by different amounts for different spots. For aliquot c, which is located immediately below the mean of the 91500 data, the calibration uncertainty only mildly increases the standard error from 0.30% to 0.36%. However, for aliquot a, which is horizontally offset from the mean of the calibration data, the systematic calibration uncertainty more than doubles the standard error from 0.32% to 0.70%. The calibration error
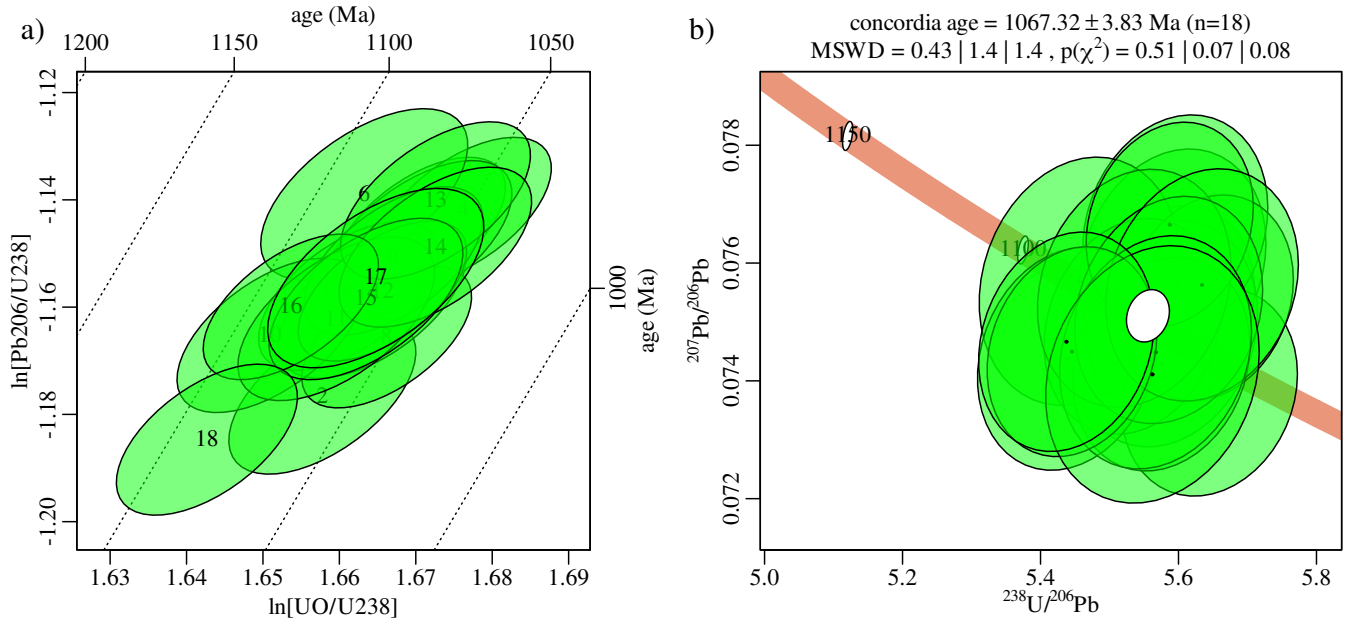
Figure 7. a) Calibration plot of the SHRIMP 91500 zircon data, using Temora2 as a reference material. Dotted lines are parallel to the best fit line of Figure 6b. b) The same data shown on a Tera-Wasserburg concordia diagram, which was obtained with `IsoplotR` and does not take into account systematic uncertainties associated with the calibration fit. All uncertainties are shown at 95% confidence. The white ellipse marks the weighted mean composition, with MSWD and p-values representing the goodness of fit for equivalence, concordance, and equivalence + concordance.

also causes the standard errors of the various aliquots to be correlated with each other. For example, the total uncertainties

355 of aliquots a and b are positively correlated ($r[a,b] = 0.62$) because their UO$_2$/U-ratios are both offset from the mean of the calibration in the same direction. In contrast, the uncertainties of aliquots a and d are negatively correlated ($r[a,d] = -0.34$) because they are offset in opposite directions from the mean of the calibration data.

The inter-spot error correlations are important when calculating weighted means (Vermeesch, 2015) and isochrons. Taking into account the full covariance structure of the data affects both the accuracy and the precision of any derived age information.

360 For example, the conventional weighted mean can be replaced with a matrix expression that accounts for correlated uncertainties (Equation 92 of Vermeesch, 2015). Applying this modified algorithm to the positively correlated aliquots a and b of Table 2 changes the weighted mean from -2.6922 to -2.6854 and increases the standard error of that mean from 0.37% to 0.44%. Applying the same calculation to the negatively correlated aliquots a and d changes their weighted mean from -2.7083 to -2.7076 and reduces its uncertainty from 0.43% to 0.35%. To take full advantage of the covariance matrix will require the de-

365 velopment of a new generation of high level data reduction software. For example, future versions of `IsoplotR` (Vermeesch, 2018a) will be modified to handle these rich data structures. A comprehensive discussion of this topic falls outside the scope of this paper.
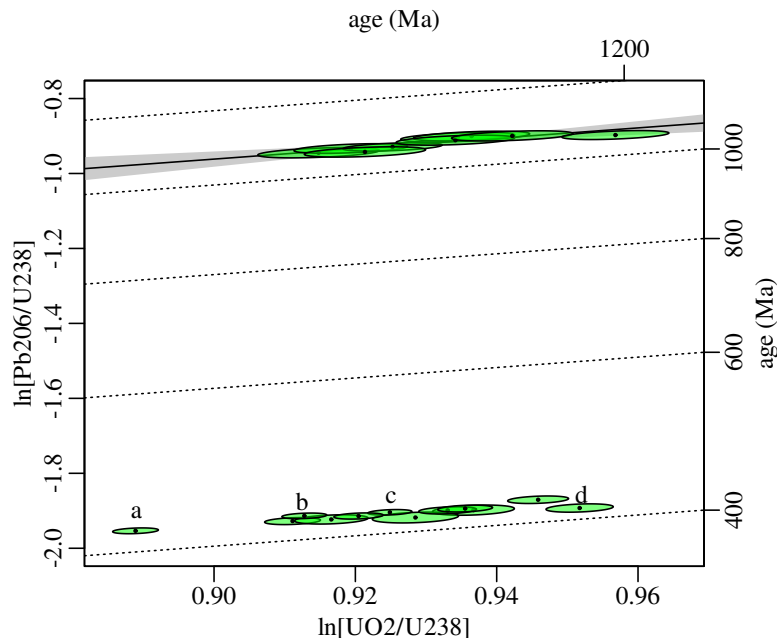
Figure 8. Calibration curve of the Cameca data, using 91500 zircon as a reference material (top half of the plot), and Temora2 zircon as a sample (bottom half). *a–d* mark four Temora2 aliquots whose uncertainty budget is explored in Table 2.

13   The R package `simplex`

`simplex` is an R package for SIMS data processing that implements the algorithm presented in this paper. The program can
370   be run in three modes: online, offline, and from the command line. The online version can be accessed at https://isoplotr.es.ucl.
ac.uk/simplex/. It contains three example U–Pb datasets, including the two datasets used in this paper, plus a Cameca monazite
U–Th–Pb dataset.

`simplex` currently accepts raw data as Cameca `.asc` and SHRIMP `.op` and `.pd` files. Support for SHRIMP `.xml` files
will be added later. The online version is a good place to try the look and feel of the software. However, it is probably not
375   the most practical way to process lots of large data files. For a more responsive user experience, `simplex` can also be run
natively on any operating system (Windows, Mac OS or Linux). To this end, the user needs to install R on their system (see
https://r-project.org/ for details). Within R, the `simplex` package can be installed from the `GitHub` code-sharing platform
using the `remotes` package, by entering the following commands at the console:

```
install.packages("remotes")
remotes::install_github("pvermees/simplex")
```

Once installed, `simplex`' graphical user interface (GUI) can be started by entering the following command at the console:

```
simplex::simplex()
```

A third and final way to use `simplex` is from the command line. This allows advanced users to create automation scripts and extend the functionality of the package. `simplex` comes with an extensive API (Application Programming Interface) of

385 fully documented user functions. An overview of all these functions can be obtained by typing the following command at the console:

```
help(package="simplex")
```

14   Discussion

This paper introduced a new algorithm for SIMS U–Pb geochronology, in which raw mass spectrometer signals are processed

390 using a combination of logratio analysis and Poisson counting statistics. In contrast with existing data reduction protocols, which handle each aliquot of an analytical sequence separately, the new algorithm simultaneously processes all of them in parallel. It thereby produces an internally consistent set of isotopic ratios and their associated covariance matrix. This covariance matrix is a rich source of information that captures both random and systematic uncertainties, including inter-spot error correlations that have hitherto been ignored in geochronology.

395   The example data of Section 12 showed that these inter-spot error correlation can be either positive or negative (see also Vermeesch, 2015; McLean et al., 2016). Ignoring them affects both the accuracy and precision of high end data processing steps such as isochron regression and concordia age calculation. Unfortunately, existing postprocessing software such as `Isoplot` (Ludwig, 2003) and `IsoplotR` (Vermeesch, 2018a) does not yet handle inter-spot error correlations. Further work is needed to extend these codes and take full advantage of the new algorithm. `IsoplotR` was designed with such future upgrades in

400 mind: its input window contains a large number of spare columns that will accommodate covariance matrices in a future update. Once the aforementioned high end data reduction calculations have been updated, it will be possible to fully quantify the gain in precision and accuracy of the new algorithm compared to the previous generation of SIMS data reduction software.

The data reduction principles laid out in this paper are applicable not only to U–Pb geochronology, but also to other SIMS applications such as stable isotope analysis. In fact, `simplex` already handles such data for multicollector Cameca instru-

405 ments. It is worth mentioning that the stable isotope functionality can also be used to correct ^{207}Pb/^{206}Pb ratio measurements for mass-dependent isotope fractionation, as was briefly discussed in Section 10. Future updates of the mass-dependent fractionation correction will also addresses the overcounted background problem that was mentioned in Section 6.

Besides U–Pb geochronology and stable isotopes, the new data reduction paradigm can also be adapted to other chronometers and other mass spectrometer designs, such as Thermal Ionisation Mass Spectrometry (TIMS, Connelly et al., 2021), noble gas

410 mass spectrometry (Vermeesch, 2015), and LAICPMS (McLean et al., 2016). `simplex` already includes a function to export data to `IsoplotR`. Adding similar functionality to other data processing software will improve geochronologists' ability to integrate multiple datasets whilst keeping track of systematic uncertainties, including those associated with reference materials and decay constants.

Code and data availability. The source code, installation instructions and example datasets for `simplex` can be accessed at https://github.
415 com/pvermees/simplex/

Appendix A

This Section provides further algorithmic details for the new U–Pb data processing workflow. It assumes that ions are recorded on SEM detectors, which is by far the most common configuration. The case of Faraday collectors is similar and, in fact, simpler.

420 Within-spot drift is modelled using a log-linear function (Equation 17) with a distinct intercept (α_x) for each ion channel (x), and shared slopes (γ_X) between isotopes of the same element (X). To illustrate this concept, consider the case of $^{204|6|7}$Pb (inclusion of ^{208}Pb is a trivial extension). Let $\hat{n}_x^i$ be the time dependent parameter of the shot noise for ^{20x}Pb, where $x \in \{4, 6, 7\}$:

$$\hat{n}_x^i = \text{bkg} + \exp(\alpha_x + \gamma_{Pb} t_x^i) d_x' \tag{A1}$$

425 then the log-likelihood function of the parameters given the data is defined as:

$$\mathcal{LL}_d \left(\alpha_{\{x\}}, \gamma_{Pb} | n_{\text{x}}^i \right) = \text{Const.} + \sum_x \sum_{i=1}^N \left(n_x^i \ln\left[\hat{n}_x^i\right] - \hat{n}_x^i \right) \tag{A2}$$

where N represents the number of cycles. The parameters $\alpha_{\{x\}} = \{\alpha_4, \alpha_6, \alpha_7\}$ and γ_{Pb} are estimated by maximising $\mathcal{LL}_d$ with respect to them. Only γ_{Pb} is used in subsequent calculations. The intercepts $\alpha_{\{x\}}$ are discarded.

The next step of the data reduction extracts logratios from the raw data using a log-linear model that is similar to the within-
430 spot drift correction (Equation 19). Here, in contrast with the drift correction, the intercepts are just as important as the slopes. For isochemical ratios such as ^{207}Pb/^{206}Pb, the slope of the drift-corrected logratios is zero and we only need to estimate the intercept. For multichemical ratios such as ^{238}U/^{206}Pb, both the slope and the intercept are non-zero. In order to keep track of covariances, it is useful to process all the isotopes together, using a common denominator. For example, using ^{206}Pb ('6') as a common denominator and ^{204}Pb ('4'), ^{207}Pb ('7'), ^{238}U ('u') and UO ('o') as numerators:

435 $$^i\beta_6^x = {}^0\beta_6^x + \gamma_6^x \tau_6^i + \gamma_X \left(\tau_x^i - \tau_6^i \right) + \ln[d_6'] - \ln[d_x'] \tag{A3}$$

where 'X' stands for Pb if $x \in \{4, 6, 7\}$, for U if $x = u$, and for UO if $x = o$. Then the normalised ion counts are given by:

$$\theta_y^i = \exp\left[{}^i\beta_6^y\right] / D_i \quad \text{and} \quad \theta_6^i = 1/D_i \tag{A4}$$

for $y \in \{4, 7, u, o\}$, with

$$D_i = 1 + \sum_y \exp[{}^i\beta_6^y] + \frac{n_b^i}{\sum_z n_z^i} \tag{A5}$$

440 in which $z \in \{4, 6, 7, u, o, b\}$. Then the log-likelihood is calculated as:

$$\mathcal{LL}_l\left({}^0\beta_6^{\{x\}}, \gamma_6^{\{x\}} | n_{\mathbb{x}}^i\right) = \text{Const.} + \sum_{i=1}^{N} \sum_x n_x^i \ln[\theta_x^i] \tag{A6}$$

where $\gamma_6^4 = \gamma_6^7 = 0$ because there is no elemental fractionation between the Pb-isotopes. From the logratios with common denominator, it is easy to derive any other logratio:

$${}^\tau\beta_x^y = {}^\tau\beta_6^y - {}^\tau\beta_6^x \tag{A7}$$

445   One of the main advantages of the new data reduction method is its ability to keep track of the full covariance structure of the data, including inter-spot error correlations. This ability is derived from the fact that all parameters are derived by the method of maximum likelihood, which stipulates that the approximate covariance matrix of any set of estimated parameters can be obtained by inverting the negative matrix of second derivatives (i.e., the Hessian matrix) of the log-likelihood function with respect to said parameters:

450 $\Sigma \approx -\mathcal{H}^{-1}$ \hfill (A8)

For example, to estimate the covariance matrix of the logratio slopes and intercepts for a single spot analysis, the Hessian is a 6×6 matrix that includes the second derivatives of $\mathcal{LL}_l$ with respect to $\beta_6^4, \beta_6^7, \beta_6^u, \beta_6^o, \gamma_6^u$ and γ_6^o. Computing this matrix is tedious to do by hand but straightforward to do numerically.

Given the covariance matrix of the logratios, subsequent data reduction steps propagate the analytical uncertainties by
455 conventional first order Taylor approximation. Thus, if $y = f(x)$, then:

$$\Sigma_y \approx J_f \Sigma_x J_f^T \tag{A9}$$

where J_f is the Jacobian matrix (and J_f^T its transpose) of partial derivatives of f with respect to x. For example, to estimate the $m \times m$ covariance matrix of m fractionation-corrected ${}^{206}\text{Pb}/{}^{238}\text{U}$-ratios, error propagation of Equation 23 would involve an $m \times (2m+3)$ Jacobian matrix and an $(2m+3) \times (2m+3)$ covariance matrix containing the uncertainties of A, B, $\ln\left[\frac{{}^{206}\text{Pb}}{{}^{238}\text{U}}\right]_a^r$,
460 as well as ${}^\tau\beta_u^6(j)$ and ${}^\tau\beta_u^o(j)$ (for j from 1 to m).

Author contributions. PV is the sole author of this paper

Competing interests. The author is an Associate Editor of *Geochronology*

Acknowledgements. The idea for this work was born during an academic visit of the author to the Institute of Geology and Geophysics (IGG) at the Chinese Academy of Sciences (CAS) in Beijing. With no prior experience in handling SIMS data, the author relied on the

465   expertise of Dr. Yang Li and Dr. Li-Guang Wu (IGG-CAS) to introduce him to the world of Cameca data processing, and of Dr. Simon Bodorkos, Dr. Charles Magee and Dr. Andrew Cross (Geoscience Australia) for SHRIMP data processing. The test data were kindly shared by Dr. Li and Dr. Bodorkos, who also provided detailed feedback on the manuscript prior to submission. Software development for the `simplex` data reduction program was supported by NERC Standard Grant #NE/T001518/1 ('Beyond `Isoplot`'), which aims to develop an 'ecosystem' of inter-connected geochronological data processing software. The graphical user interface for `simplex` makes extensive

470   use of the `shinylight` and `dataentrygrid` packages developed by software engineer Tim Band, who is employed on this NERC grant.

References

Aitchison, J.: The Statistical Analysis of Compositional Data, Journal of the Royal Statistical Society, 44, 139–177, 1982.

Aitchison, J.: The statistical analysis of compositional data, London, Chapman and Hall, 1986.

Black, L. P.: The use of multiple reference samples for the monitoring of ion microprobe performance during zircon $^{207}Pb/^{206}Pb$ age determinations, Geostandards and Geoanalytical Research, 29, 169–182, 2005.

Black, L. P., Kamo, S. L., Allen, C. M., Davis, D. W., Aleinikoff, J. N., Valley, J. W., Mundil, R., Campbell, I. H., Korsch, R. J., Williams, I. S., et al.: Improved $^{206}Pb/^{238}U$ microprobe geochronology by the monitoring of a trace-element-related matrix effect; SHRIMP, ID–TIMS, ELA–ICP–MS and oxygen isotope documentation for a series of zircon standards, Chemical Geology, 205, 115–140, 2004.

Bodorkos, S., Bowring, J., and Rayner, N.: Squid3: Next-generation Data Processing Software for Sensitive High Resolution Ion Micro Probe (SHRIMP), Geoscience Australia, 2020.

Chayes, F.: On ratio correlation in petrography, The Journal of Geology, 57, 239–254, 1949.

Chayes, F.: On correlation between variables of constant sum, Journal of Geophysical research, 65, 4185–4193, 1960.

Connelly, J., Bollard, J., Costa, M., Vermeesch, P., and Bizzarro, M.: Improved methods for high-precision Pb–Pb dating of extra-terrestrial materials, Journal of Analytical Atomic Spectrometry, 36, 2579–2587, 2021.

Dodson, M.: A linear method for second-degree interpolation in cyclical data collection, Journal of Physics E: Scientific Instruments, 11, 296, 1978.

Galbraith, R. F.: Statistics for fission track analysis, CRC Press, 2005.

Galbraith, R. F. and Laslett, G. M.: Statistical models for mixed fission track ages, Nuclear tracks and radiation measurements, 21, 459–470, 1993.

Hiess, J., Condon, D. J., McLean, N., and Noble, S. R.: $^{238}U/^{235}U$ systematics in terrestrial uranium-bearing minerals, Science, 335, 1610–1614, 2012.

Horstwood, M. S., Košler, J., Gehrels, G., Jackson, S. E., McLean, N. M., Paton, C., Pearson, N. J., Sircombe, K., Sylvester, P., Vermeesch, P., et al.: Community-Derived Standards for LA-ICP-MS U-(Th-) Pb Geochronology–Uncertainty Propagation, Age Interpretation and Data Reporting, Geostandards and Geoanalytical Research, 40, 311–332, 2016.

Ireland, T. R. and Williams, I. S.: Considerations in zircon geochronology by SIMS, Reviews in mineralogy and geochemistry, 53, 215–241, 2003.

Jeon, H. and Whitehouse, M. J.: A critical evaluation of U–Pb calibration schemes used in SIMS zircon geochronology, Geostandards and Geoanalytical Research, 39, 443–452, 2015.

Ludwig, K. R.: Decay constant errors in U–Pb concordia-intercept ages, Chemical Geology, 166, 315–318, 2000.

Ludwig, K. R.: User's manual for Isoplot 3.00: a geochronological toolkit for Microsoft Excel, Berkeley Geochronology Center Special Publication, 4, 2003.

Magee, C., Danisik, M., and Mernagh, T.: Extreme isotopologue disequilibrium in molecular SIMS species during SHRIMP geochronology, Geoscientific Instrumentation, Methods and Data Systems, 6, 523–536, 2017.

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, 2016.

Min, K., Mundil, R., Renne, P. R., and Ludwig, K. R.: A test for systematic errors in $^{40}Ar/^{39}Ar$ geochronology through comparison with U/Pb analysis of a 1.1-Ga rhyolite, Geochimica et Cosmochimica Acta, 64, 73–98, 2000.

Nasdala, L., Corfu, F., Valley, J. W., Spicuzza, M. J., Wu, F.-Y., Li, Q.-L., Yang, Y.-H., Fisher, C., Münker, C., Kennedy, A. K., et al.: Zircon M127–A homogeneous reference material for SIMS U–Pb geochronology combined with hafnium, oxygen and, potentially, lithium isotope analysis, Geostandards and Geoanalytical Research, 40, 457–475, 2016.

Ogliore, R., Huss, G., and Nagashima, K.: Ratio estimation in SIMS analysis, Nuclear instruments and methods in physics research section B: beam interactions with materials and atoms, 269, 1910–1918, 2011.

Pawlowsky-Glahn, V. and Buccianti, A.: Compositional data analysis: Theory and applications, John Wiley & Sons, 2011.

Renne, P. R., Swisher, C. C., Deino, A. L., Karner, D. B., Owens, T. L., and DePaolo, D. J.: Intercalibration of standards, absolute ages and uncertainties in ^{40}Ar/^{39}Ar dating, Chemical Geology, 145, 117–152, 1998.

Stern, R. A., Bodorkos, S., Kamo, S. L., Hickman, A. H., and Corfu, F.: Measurement of SIMS instrumental mass fractionation of Pb isotopes during zircon dating, Geostandards and Geoanalytical Research, 33, 145–168, 2009.

Vermeesch, P.: Tectonic discrimination diagrams revisited, Geochemistry, Geophysics, Geosystems, 7, https://doi.org/10.1029/2005GC001092, 2006.

Vermeesch, P.: HelioPlot, and the treatment of overdispersed (U-Th-Sm)/He data, Chemical Geology, 271, 108 – 111, https://doi.org/10.1016/j.chemgeo.2010.01.002, 2010.

Vermeesch, P.: Revised error propagation of ^{40}Ar/^{39}Ar data, including covariances, Geochimica et Cosmochimica Acta, 171, 325–337, 2015.

Vermeesch, P.: IsoplotR: a free and open toolbox for geochronology, Geoscience Frontiers, 9, 1479–1493, 2018a.

Vermeesch, P.: Statistical models for point-counting data, Earth and Planetary Science Letters, 501, 1–7, 2018b.

Weltje, G.: Quantitative analysis of detrital modes: statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology, Earth-Science Reviews, 57, 211 – 253, https://doi.org/10.1016/S0012-8252(01)00076-9, 2002.

Wiedenbeck, M., Alle, P., Corfu, F., Griffin, W., Meier, M., Oberli, F. v., Quadt, A. v., Roddick, J., and Spiegel, W.: Three natural zircon standards for U-Th-Pb, Lu-Hf, trace element and REE analyses, Geostandards newsletter, 19, 1–23, 1995.

Williams, I. S.: U-Th-Pb geochronology by ion microprobe, Reviews in Economic Geology, 7, 1–35, 1998.

York, D., Evensen, N. M., Martínez, M. L., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the best straight line, American Journal of Physics, 72, 367–375, 2004.