AGU PUBLICATIONS

Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1002/2015GC006097

Algorithms and software for U-Pb geochronology by LA-ICPMS

Noah M. McLean[1,2], James F. Bowring[3], and George Gehrels[4]

[1]Department of Earth Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA, [2]Now at Department of Geology, University of Kansas, Lawrence, Kansas, USA, [3]Department of Computer Science, College of Charleston, Charleston, South Carolina, USA, [4]Department of Geosciences, University of Arizona, Tucson, Arizona, USA

Key Points:
- We present new software to review, reduce, interpret, and archive U-Pb LA-ICPMS data, *ET_Redux*
- *ET_Redux* uses matrix-based full uncertainty propagation for "intercept" and "downhole" methods
- We introduce statistical techniques to handle ratios, systematic uncertainties, and overdispersion

Correspondence to:
N. McLean,
noahmc@ku.edu

Citation:
McLean, N. M., J. F. Bowring, and G. Gehrels (2016), Algorithms and software for U-Pb geochronology by LA-ICPMS, *Geochem. Geophys. Geosyst.*, *17*, 2480–2496, doi:10.1002/2015GC006097.

Received 15 SEP 2015
Accepted 28 FEB 2016
Accepted article online 3 MAR 2016
Published online 10 JUL 2016

Abstract The past 15 years have produced numerous innovations in geochronology, including experimental methods, instrumentation, and software that are revolutionizing the acquisition and application of geochronological data. For example, exciting advances are being driven by Laser-Ablation ICP Mass Spectrometry (LA-ICPMS), which allows for rapid determination of U-Th-Pb ages with 10s of micrometer-scale spatial resolution. This method has become the most commonly applied tool for dating zircons, constraining a host of geological problems. The LA-ICPMS community is now faced with archiving these data with associated analytical results and, more importantly, ensuring that data meet the highest standards for precision and accuracy and that interlaboratory biases are minimized. However, there is little consensus with regard to analytical strategies and data reduction protocols for LA-ICPMS geochronology. The result is systematic interlaboratory bias and both underestimation and overestimation of uncertainties on calculated dates that, in turn, decrease the value of data in repositories such as EarthChem, which archives data and analytical results from participating laboratories. We present free open-source software that implements new algorithms for evaluating and resolving many of these discrepancies. This solution is the result of a collaborative effort to extend the *U-Pb_Redux* software for the ID-TIMS community to the LA-ICPMS community. Now named *ET_Redux*, our new software automates the analytical and scientific workflows of data acquisition, statistical filtering, data analysis and interpretation, publication, community-based archiving, and the compilation and comparison of data from different laboratories to support collaborative science.

1. Introduction

Time is a key aspect of any historical science and serves as a backbone of our understanding of the origin and evolution of our solar system. The geological record allows us to explore, in detail, the governing processes and interactions among the solid earth, biosphere, hydrosphere, and atmosphere in deep time. A highly resolved timescale allows us to answer new questions about the duration and rates of processes like climate change, extinction, evolution, geodynamics, magmatism, and orogenic processes. The past 15 years have seen explosive interest in geochronology, including the development of techniques and applications, major steps in instrument design and data acquisition, and the training of a new generation of students who use geochronology as part of their toolkit to understand Earth history.

Innovations in experimental methods and instrumentation are revolutionizing the acquisition and application of geochronological data in Earth Science research. In general, the number and capabilities of new instruments have led to dramatic growth in the size and number of available data sets. Laser-Ablation ICP Mass Spectrometry (LA-ICPMS), with its high spatial resolution and high sample throughput, has been at the forefront of this new wave of data: U-Th-Pb geochronology by laser ablation has become the most commonly applied tool for dating zircons, producing tens to hundreds of thousands of analyses per year and constraining a host of geological problems. The LA-ICPMS community is now faced with archiving these data and, more importantly, ensuring that data meet the highest standards for precision and accuracy and that interlaboratory biases are minimized. However, there is little consensus with regard to analytical strategies and data reduction protocols for U-Th-Pb laser ablation geochronology. The result is systematic interlaboratory bias and both underestimation and overestimation of uncertainties on calculated dates that, in turn, decrease the value of data repositories such as EarthChem hosting data from various laboratories. We are trying to solve the bias issue in order to make such community data repositories more valuable for scientific inquiry.

We seek to remedy this situation by developing cyberinfrastructure tools in support of LA-ICPMS U-Pb geochronology. Our strategy involves open-source, free software systems developed with input from the international Earth Science community. These software packages treat with statistical rigor all aspects of data reduction, from calculation of corrected isotopic ratios to correction for interferences and drift. Our approach follows the successful strategy used in developing similar software as part of the EARTHTIME project for U-Pb ID-TIMS data reduction [*Bowring et al*., 2011; *McLean et al*., 2011].

In the EARTHTIME ID-TIMS case, this collaborative software development and utilization resulted in dramatic improvements in the ability to precisely and accurately quantify geological time due to advances in radioisotopic dating techniques and integration with cyclostratigraphic-based approaches [*Schoene*, 2014]. These improvements have highlighted subtle biases between geochronological techniques (e.g., Ar-Ar and U-Pb) [*Wotzlaw et al*., 2013] and between laboratories using the same techniques [*Blackburn et al*., 2013]. This successful, exemplar interdisciplinary collaboration produced the free open-source program *ET_Redux* (previously *U-Pb_Redux*) as part of the EARTHTIME initiative. *ET_Redux* is cyberinfrastructure that provides a one-stop destination for reducing, examining, archiving, reporting, and compiling conventional ID-TIMS U-Pb data. The *ET_Redux* system implements techniques that support the seamless federation of analytical results from many labs into a public archival database using standard and open techniques. These techniques include the universal identification of geological samples, transparent data reduction algorithms, robust data formats, Internet-based storage and retrieval, and transparent compilation and aggregation techniques for analysis of records retrieved from the database. To translate this success to LA-ICPMS, a renewed evaluation and quantification of systematic uncertainties of a magnitude similar to analytical precision is required.

This contribution documents the extension of *ET_Redux* to support the end-to-end processing of LA-ICPMS U-Th-Pb geochronologic data. *ET_Redux* ingests raw data files produced by mass spectrometer software packages and provides interactive visualizations and statistical filtering of background and on-peak measured counts and ratios, including interactive fitting of parametric functions to measured ratios of both individual reference material and unknown sample acquisitions. The software produces both a downhole [*Paton et al*., 2010] and intercept-based visualization and treatment of the reference material for the session and provides for fitting parametric or spline functions to the session-scale drift in elemental and isotopic fractionation. *ET_Redux* corrects for common lead with model isotopic compositions selected by the user, and then generates a customizable data table and visualizations such as concordia and probability density/kernel density plots, with support for sample date interpretations using weighed means and upper/lower intercepts. As in the case of ID-TIMS analyses, the user can interact with Geochron.org to archive data and results.

As an example of how *ET_Redux* supports day-to-day research, consider the following scenario: a variety of researchers individually processing samples with either ID-TIMS or LA- ICPMS techniques use *ET_Redux* to analyze and interpret results for samples from somewhere in Arizona. With unique sample identifiers supplied by the System for Earth Sample Registration (SESAR), the researchers upload the data and analytical interpretations to the public portion of the EarthChem Geochron database using *ET_Redux*. Sometime later, an individual researcher queries SESAR or Geochron for a set of registered samples for the same general geographic area in Arizona. The researcher then uses this set of identifiers with the *ET_Redux* system to retrieve data sets from the Geochron database and to compile results from both ID-TIMS and LA-ICPMS analyses in a common analytical and visualization interface. This compilation mode is one of the major contributions of *ET_Redux*, because it provides for user-friendly and robust visualization tools that support the integration of data from diverse sources. We anticipate this will allow researchers to pose and answer new questions about geochronology.

2. Principles of LA-ICPMS Data Handling

Despite the popularity and appeal of LA-ICPMS U-Pb dating, there is no broadly adopted data reduction and uncertainty propagation protocol in use today. Largely, each laboratory has created its own data handling protocols, specific to its instrumentation and scientific goals. An outline of the data handling process has been discussed at several recent meetings [e.g., *Horstwood et al*., 2010, and recent meetings of the LA-ICPMS Working Group at Goldschmidt in 2013 and 2015] and in a recent summary article by *Horstwood*

et al. [2016]. Here we detail several new contributions to the statistical analysis of LA-ICPMS U-Pb data, provide examples of how they contribute to more accurate and precise U-Pb measurements, and explain their implementation in *ET_Redux*.

2.1. Compositional Data Analysis

One notable recent debate in the U-Pb geochronology literature is about how to calculate the average value of several measurements of the relative abundances of multiple isotopes. This problem has some unique characteristics: we are interested in the relative abundances of the isotopes present, usually expressed as ratios, and rarely require or have information on their absolute abundance to the same precision.

In one school of thought, the average isotopic composition is the arithmetic mean of the measured ratios [*Fisher et al*., 2010]. This calculation occurs automatically on most mass spectrometers, which report "grand mean" averages of measured isotope ratios at the end of an analysis. For instance, using the variable z to represent the average value of n measurements of the ratio of isotopes x and y,

$$z = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{y_i} \tag{1}$$

Other workers have noticed that the resulting means may comprise a skewed distribution and contain a number of outliers, especially for very large or small isotope ratios, and/or imprecise data. To address this issue, it has been suggested [e.g., *Ulianov et al*., 2012; *Ulianov and Müntener*, 2014] that a more accurate measure of the average isotopic composition is instead the ratio of the integrated ion beam intensities. For instance, the same average z above could be expressed as

$$z = \frac{1/n\sum_{i=1}^{n} x_i}{1/n\sum_{i=1}^{n} y_i} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} \tag{2}$$

The "mean of the ratios" versus "ratio of the means" debate has provoked considerable interest and analysis [*Fisher et al*., 2010], with no apparent solution.

To resolve this issue, it helps to clarify what an "average" isotopic composition is and to enumerate several properties that a suitable average must possess. This is the subject of a series of related papers by *Aitchison* [1984, 1986, 1992], who defined a "composition" as a measure of the relative abundance of several components. Here the components are isotopes, and the pertinent information is their proportions, which could be expressed for instance as ratios or relative percentages, regardless of their absolute abundances. In his seminal 1984 publication, Aitchison showed that neither the "mean of the ratios" nor the "ratio of the means" passed certain common-sense tests of self-consistency.

For instance, the "mean of the ratios" approach, as utilized in all modern mass spectrometry software packages, yields a different "average" isotopic composition depending on the choice of numerator and denominator. This is illustrated in Table 1. The first three columns represent randomly generated intensity data for isotopes A, B, and C, and the next four columns their isotope ratios, along with their mean and relative standard deviation. Note that the same data is used to calculate the fourth and sixth columns, expressed as B/A and A/B. If these two mean compositions are the same, then the mean of column six should be the same as the reciprocal of the mean of column four. A simple check shows that this is not the case, and that the two means (and for instance, any ages calculated or conclusions drawn from them) are different. The difference between these two measures becomes larger as scatter increases in ratio-to-ratio data but remains nonzero for all data sets. We therefore consider the mean of the ratios to be inappropriate, even for precise data sets.

The "ratio of the means" approach does not involve averaging across multiple ratios, but it has two other significant shortcomings related to assigning an uncertainty to the estimated ratio. The first derives from calculating a standard deviation of the measured intensities of the numerator and denominator isotopes. Measured intensities in mass spectrometry are very rarely stable, and laser ablation data are no exception, with a very fast rise time as the first pulse of laser-ablated material reaches the mass spectrometer, steady decay during the ablation process, and then a rapid fall after the final ablated material passes through the system. Integrating the measured intensities (e.g., measuring the total counts) for each of the isotopes and

Table 1. Randomly Generated Intensity, Ratio, and Log-Ratio Data for Isotopes A, B, and C[a]

A	B	C		B/A	C/A	A/B	B/C		log(B/A)	log(C/A)	log(A/B)	log(B/C)
3.8816	2.2237	3.5034		0.5729	0.9026	1.7456	0.6347		−0.5571	−0.1025	0.5571	−0.4546
3.4189	3.6334	4.4136		1.0627	1.2909	0.9410	0.8232		0.0608	0.2554	−0.0608	−0.1945
1.8736	3.4878	6.2357		1.8615	3.3282	0.5372	0.5593		0.6214	1.2024	−0.6214	−0.5810
2.7661	8.5963	4.0573		3.1077	1.4668	0.3218	2.1187		1.1339	0.3831	−1.1339	0.7508
2.7887	3.5317	7.1290		1.2664	2.5564	0.7896	0.4954		0.2362	0.9386	−0.2362	−0.7024
2.2993	2.3495	6.8411		1.0218	2.9753	0.9786	0.3434		0.0216	1.0903	−0.0216	−1.0687
6.9564	8.9176	1.8384		1.2819	0.2643	0.7801	4.8507		0.2484	−1.3308	−0.2484	1.5791
1.9362	7.4160	2.3421		3.8302	1.2096	0.2611	3.1664		1.3429	0.1903	−1.3429	1.1526
2.3554	3.5661	2.7637		1.5140	1.1733	0.6605	1.2903		0.4148	0.1599	−0.4148	0.2549
2.1014	1.8837	3.3307		0.8964	1.5850	1.1156	0.5656		−0.1094	0.4606	0.1094	−0.5699
			Mean:	1.6416	1.6752	0.8131	1.4848	Mean:	0.3414	0.3247	−0.3414	0.0167
			1/(B/A)			0.6092		1/log(B/A)			−0.3414	
			(B/A)/(C/A)				0.9799	log(B/A)/log(C/A)				0.0167
								exp(mean):	1.4069	1.3837	0.7108	1.0168
								1/(B/A)			0.7108	
								(B/A)/(C/A)				1.0168

[a]Examining the arithmetic means of the ratio data reveals that the estimated isotopic composition depends on an arbitrary choice of numerator and denominator for the two isotopes of interest. The geometric mean, or exponentiated mean of log-ratios, does not share this problem and provides an internally consistent estimate of the isotopic composition of the sample.

then dividing to calculate average isotope ratios gives an accurate measure of their relative abundance. However, evaluating the standard error of the intensities and interpreting this as a measure of uncertainty about the mean (e.g., $\pm 2\sigma_{\bar{x}}$) violates many of the assumptions involved with calculation of the standard error.

Specifically, the data used to calculate a mean and its standard error should be randomly scattered, independent, and identically distributed. From the large, predictable trend of rise, slow decay, and fall in the data from an ablation, one can predict subsequent data points from the previous ones, which is not consistent with all three of these conditions. Additionally, both the "mean of the ratios" and "ratio of the means" approaches assign normal (Gaussian) distributions to the average isotope ratios calculated. This is a problem because the normal distribution is defined (and assigns finite probability) over all real numbers, both positive and negative. For instance, an imprecise measurement of a small quantity may result in a mean whose uncertainty overlaps with zero and extends to negative numbers, a clear impossibility.

Correcting these shortcomings and several others, *Aitchison* [1986] proposed instead to use the geometric mean of the relative abundances, all expressed to a common denominator component. This can be accomplished simply by evaluating the logarithm (with any consistent base) of each ratio, which *Aitchison* [1986] terms the additive log-ratio transform (*alr*), calculating the arithmetic mean of these log-ratios, and finally exponentiating the results.

$$\log(z) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{x_i}{y_i}\right)$$
$$z = \exp(\log(z)) \tag{3}$$

This produces self-consistent means: for instance, the geometric mean of the (A/B) terms in Table 1 is equal to the reciprocal of the geometric mean of the (B/A) terms. In the same way that it produces self-consistent means, the *alr* also produces self-consistent uncertainties and uncertainty correlations that can be used for uncertainty propagation. Although the log-ratios themselves can be positive or negative, exponentiating a negative number always produces a positive number. Therefore, assigning a normal distribution in "log-ratio space" results in a lognormal distribution in "ratio space," which is defined only over the positive real numbers.

ET_Redux utilizes Aitchison's *alr* starting at the data collection stage. Baseline and on-peak data cannot be directly transformed into log-ratios because the size of the ion beam of interest (and the relative abundance of a species in the sample) is defined as the difference between the on-peak and the baseline intensities. This is especially important for Faraday measurements, where the zero on an amplifier can be assigned to any positive or negative voltage. Therefore, (1) a normal distribution is assigned in *ET_Redux* to each of the

baseline and on-peak integrations (see section 2.2), a user-designated function, for instance the mean of a stable baseline or an exponential fit to a decaying washout curve, is fitted to the baseline data, and (2) its extrapolation is subtracted from the on-peak data. Notably, because the baseline extrapolation is subtracted from all on-peak integrations, the result is an uncertainty correlation between the on-peak integration uncertainties, since they now share a common uncertainty contribution. The baseline-corrected on-peak intensities, which are constrained to be positive, can now be used to calculate log-ratios.

2.2. Weighted Least Squares and Overdispersion

Although the rigorous (log-ratio) description of isotope ratio data is important in estimating an accurate and self-consistent description of its uncertainty, log-ratio analysis of raw intensity data alone does not sufficiently capture the observed variability of measured isotope log-ratios. It is possible to estimate the theoretical uncertainty expected for a given measured log-ratio, using its two ion beam intensities and the properties of the collectors used. The noise from random fluctuations in a steady ion beam is called "shot" noise and can be modeled as a Poisson process [*Schottky*, 1918], which is governed by the fact that ion generation is a random process, with each event independent from all others. Ion counters attempt to detect each ion arriving at the analyzer of the mass spectrometer, and if perfectly efficient, should reflect only the shot noise from ion generation. In contrast, Faraday collectors measure the intensity of ion beams using large operational amplifier circuits, which add additional noise to the signal. This noise is dominated by the Johnson-Nyquist noise, or thermal noise, which depends in large part on the resistance in its largest resistor and, to a lesser extent, its temperature [*Nyquist*, 1928].

ET_Redux utilizes user-input integration times (section 4.1) to calculate the theoretical uncertainties on each measured log-ratio. This gives higher statistical weight to more precise, higher-intensity measurements that often occur during the beginning of an ablation, and lower weight to less precise, lower intensity measurements that often occur at the end of the ablation. The anticipated uncertainty of each ion beam is the sum its relevant sources: shot noise applies to every ion beam, with the addition of a dead-time-related uncertainty for an ion counter or a Johnson-Nyquist noise term for a Faraday collector, along with uncertainties in the relative gains when multiple collectors are used.

Assuming that the measured relative abundances of the isotopes vary in a way that can be fit by a smooth function, the anticipated uncertainty does not often adequately explain the scatter in the data (e.g., Figure 1). This specifically hampers fitting a parametric curve to measured data to determine an "intercept," typically assumed to correspond to the time at which the laser fires or the sample arrives at the mass spectrometer source, resulting in consistently high MSWDs, better known as reduced chi-square values (χ^2_{red}). Because the measured data point uncertainties have been underestimated, a conventionally calculated intercept uncertainty will be underestimated as well. The solution used in the fission track [*Galbraith*, 2005] and (U-Th)/He [*Vermeesch*, 2010] communities is to include an extra source of uncertainty called an overdispersion, a variance term added to each measurement to account for the excess scatter in the data. Potential sources of overdispersion for LA-ICPMS measurements are discussed in more detail in *Ulianov et al*. [2015].

Including an overdispersion term preserves the preferential weighting of higher-intensity data while effectively striking a compromise between utilizing only theoretical measurement uncertainties, which underestimates total uncertainty budgets, and weighting all data points the same (e.g., a conventional least squares approach), which fails to exploit the higher precision of higher-intensity data. This principle is illustrated in Figure 1.

Parametric (e.g., mean, line, and exponential) fits are used in *ET_Redux* data processing in two stages: modeling downhole elemental and isotopic fractionation during the course of a single "spot" or analysis, and modeling the slow drift in these parameters over the course of an analytical session. For both, the end goal is predictive: to extrapolate from the given data back to "time zero," or to interpolate between the measured fractionation values of multiple reference materials and predict the value of that fractionation at the analysis time of an unknown. And for the session fit, the uncertainty in the interpolated value derives from both the overdispersion term and the uncertainty in the function fit to the data. *ET_Redux* automatically calculates an overdispersion for every parametric fit with $\chi^2_{red} > 1$, then incorporates this overdispersion into the uncertainties in interpolated session fits to estimate the uncertainties in unknowns. For to do this, *ET_Redux* first calculates the parametric fit without

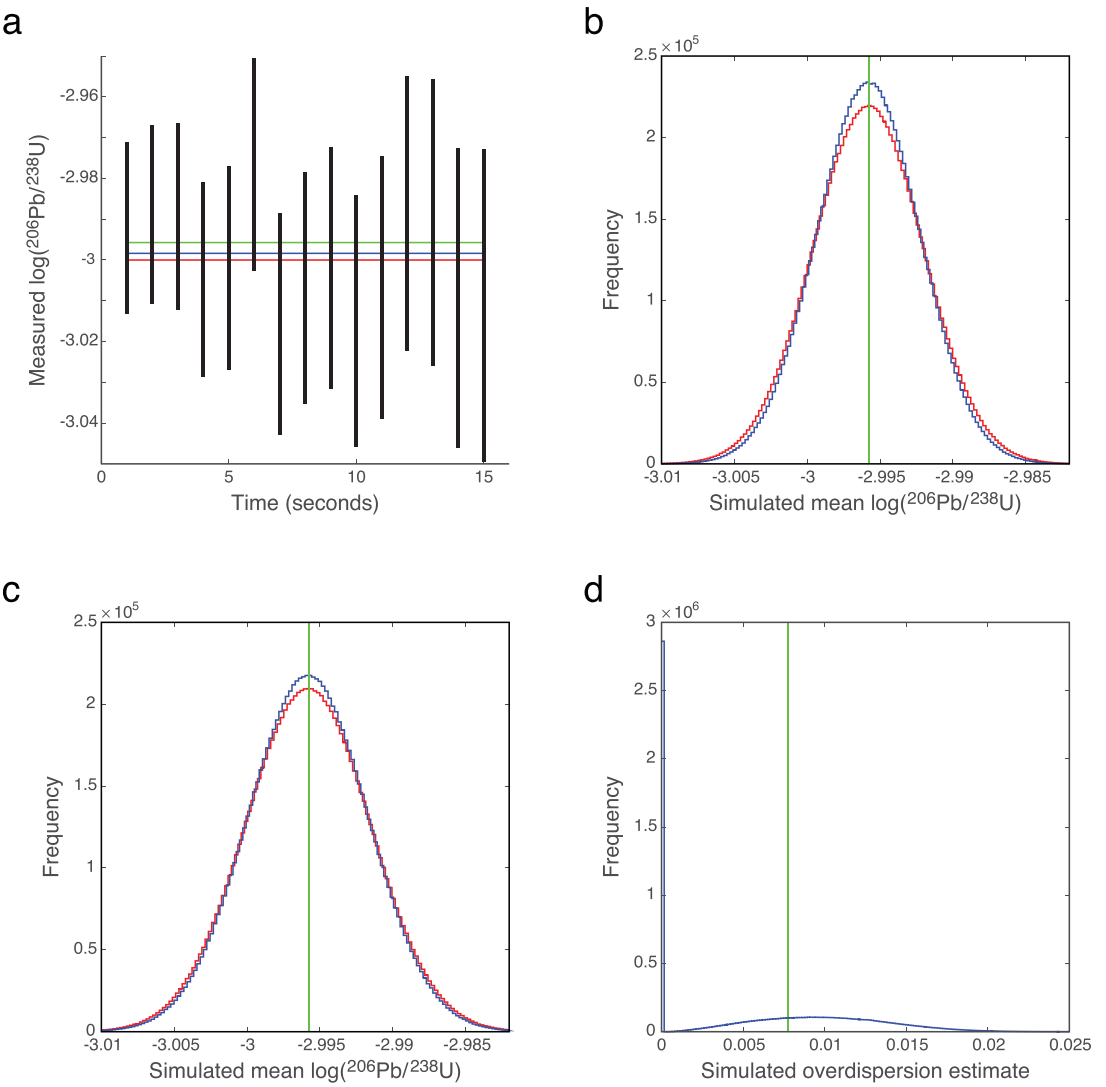

Figure 1. Results of Monte Carlo simulations for fitting heteroscedastic ablation data without and with overdispersion. (a) A single Monte Carlo simulation. ^{206}Pb and ^{238}U were simulated from Poisson distributions, with count rates consistent with a 10^6 cps ^{238}U beam decaying to 3×10^5 cps over 15 s, measuring ^{238}U for 10% and ^{206}Pb for 20% of each cycle. Vertical black bars are the expected 2σ uncertainties for each calculated log-ratio, based on counting statistics, and the horizontal red, blue, and green lines are the unweighed, uncertainty-weighted, and true means, respectively. (b) Histograms (bin-top lines only) showing 10^7 Monte Carlo simulations of the scenario described in Figure 1a. The red and blue histograms correspond to unweighed and weighted means, and the green vertical line is the true value. Weighting data points by their expected uncertainties (blue curve) yield modest but measurable improvement in the accuracy and precision of analyses. (c,d) Adding an overdispersion equivalent to 0.2% per analysis, or ~0.078% per integration, to the same 10^7 Monte Carlo trials evens out the difference in the magnitude of the uncertainty between the beginning and end integrations, so that the data becomes more homoscedastic and its uncertainty is better fit by the unweighted means (red trace of histogram bin-tops). However, solving separately for the mean and overdispersion, represented as blue trace of histogram bin-tops, still yields improvements in precision and accuracy, and additionally estimates the additional scatter outside of expected counting statistics, an important piece of information for assessing and improving instrument performance.

overdispersion, and then uses these estimated fit parameters to initialize a Levenberg-Marquardt solver tailored for the fit function. Pseudocode for the Levenberg-Marquardt algorithm is provided in Appendix A.

Nonparametric fits, such as the smoothing spline implemented only for the session fits in *ET_Redux*, use both the overdispersion and the uncertainties in the measured data set to uniquely determine their fit. For the smoothing spline, the piecewise cubic spline chosen minimizes S in equation (4), a sum of two terms [*Green and Silverman*, 1993]. For a smoothing spline $g(t)$ that produces a vector of residuals $\mathbf{r}$ for measured data with a covariance matrix $\boldsymbol{\Sigma}$,

$$S = \mathbf{r}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{r} + \lambda \int |g''(t)|\mathrm{d}t \quad (4)$$

The first term in equation (4) quantifies the misfit between the spline and the data, and it is the matrix analog of the sum of the squares of the weighted residuals [*McLean et al*., 2011]. The second term is the integral of the absolute value of the second derivative of the spline $g(t)$ itself, which is proportional to its "smoothness." The parameter λ governs a tradeoff between how closely the spline fits the data, which minimizes the first term, and how smooth the spline is, which minimizes the second term. For a given estimate of the uncertainties in the reference materials, *ET_Redux* solves for the unique value of λ that provides the smoothest spline such that $\chi^2_{red} = 1$.

A smoothing spline fit to the estimated analytical uncertainties alone will often appear "rough" or overfit, fluctuating faster than expected for a stable instrument setup, if the estimated uncertainties neglect an unrecognized uncertainty component (i.e., overdispersion). Adding an overdispersion term changes the λ parameter required to produce $\chi^2_{red} = 1$, and therefore the smoothness of the fit to the standards. Since the expected stability of the instrument is an arbitrary parameter, *ET_Redux* cannot solve for λ. Instead, it produces an interactive plot of the roughness of the spline, as given by the second term in equation (4), versus the overdispersion added. The user can then select a point on this plot to quickly estimate the overdispersion required to produce what they feel is a reasonably smooth fit to the session data.

2.3. Sample-Standard Bracketing

An additional concern when bracketing unknowns with reference material analyses regards the uncertainties propagated into each of the unknowns from the measured reference materials. Mass spectrometer measurements of elemental and isotopic ratios are always biased by the different volatilization, ionization, and detection efficiency of each species measured, and this bias may change with time during the analysis. To estimate this bias, reference materials of known elemental and isotopic composition are measured before, during, and after the unknowns of interest, and the calculated bias of the reference materials is then interpolated to correct the unknown measurements. If the bias, as measured by the reference materials during a session, does not change with time, then the best estimate of the average bias is the weighted mean of these measurements. If the bias changes with time, a parametric (e.g., line and exponential) or nonparametric (e.g., spline) fit should be used to model the time-varying behavior.

Because the reference material measurements themselves have measurement uncertainties, the resulting bias corrections make uncertainty contributions to each of the bias-corrected unknowns. When the same set or subset of reference materials is used to correct multiple unknowns, then this uncertainty contribution is shared among the bias-corrected unknowns, and their uncertainties are correlated. In other words, the bias-corrected unknown measurements are no longer independent and cannot be treated as such when their information is combined. This is an important consideration when multiple unknowns are used together in a calculation, for instance to estimate a weighted mean [*McLean et al*., 2011].

When evaluating weighted means of multiple unknowns run during the same session, neglecting these uncertainty correlations results in an underestimate of both the weighted mean uncertainty and the χ^2_{red}. However, the degree of underestimation depends on several factors, including the number and relative timing of samples and standards analyzed, the size of the uncertainty contribution from the bias correction, and the function used to fit the session data. Quantifying this underestimation requires evaluating the degree of statistical correlation among the interpolated function values throughout the session.

The plots in the top row of Figure 2 represent three different ways to fit a session of reference material data: a weighted mean (Figure 2a), weighted least squares line (Figure 2c), and a weighted least squares smoothing spline (Figure 2e). In each plot, black lines are weighted least squares fits to the data, with $\pm 2\sigma$ uncertainty envelopes in green. All data in Figures 2a and 2c are synthesized such that $\chi^2_{red} = 1$, and the data in Figure 2e are an actual suite of $^{208}Pb/^{232}Th$ measurements for a Sri Lanka zircon standard performed at the University of Arizona LaserChron facility.

Because the weighted mean in Figure 2a uses all the reference material measurements in the session, the interpolated value of the mean at all times t depends simultaneously on all the reference materials, meaning that all the interpolated values have correlated uncertainties. The magnitude of this correlation is illustrated in Figure 2b and is unity at all times t. In other words, the systematic uncertainty contribution from

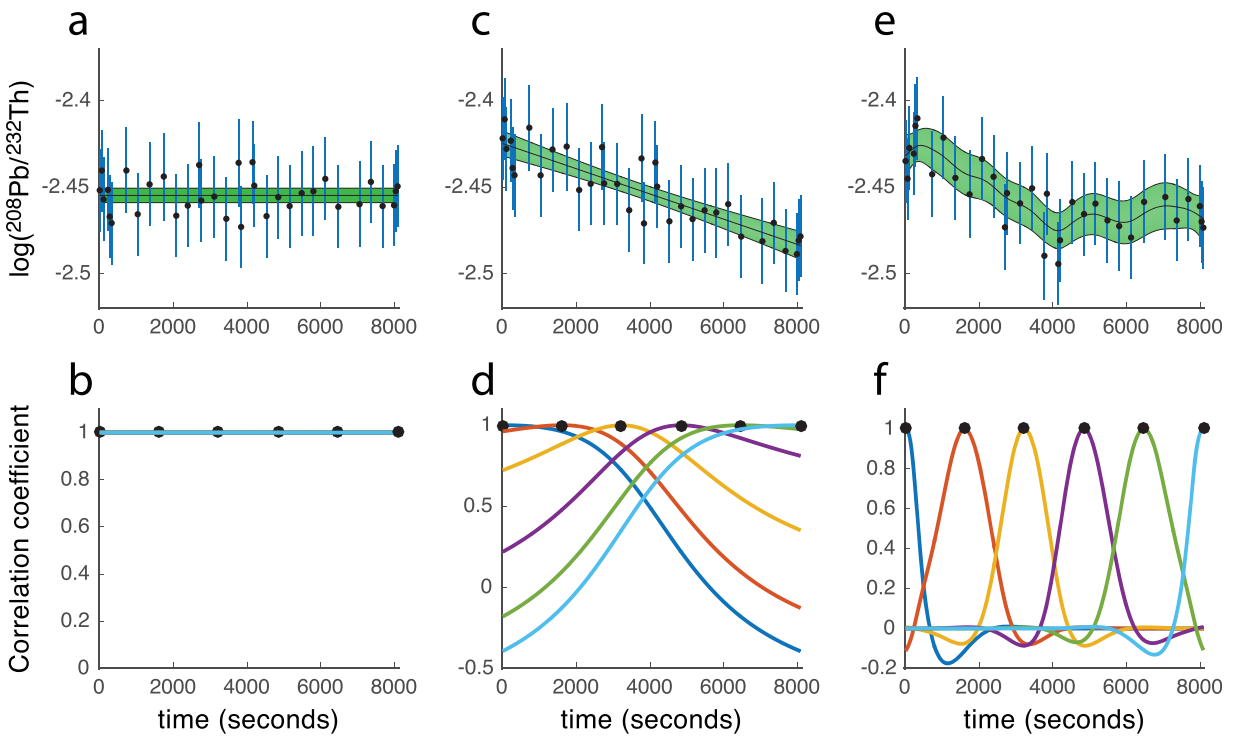

Figure 2. Sample-standard bracketing for a set of 33 standards. The top row of plots illustrates data sets fit by a mean, line, and spline, respectively. The bottom row illustrates the correlation structure among the uncertainties of unknowns from the same session, described in section 2.3.

interpolating between reference materials to determine the bias is the same at every point, and because each uncertainty derives from the same mean, it is perfectly correlated with all other uncertainties. When propagated into a weighted mean, the uncertainty contribution from sample-standard bracketing will not decrease by increasing the number of unknown data points n in the weighed mean, and this represents a limiting source of uncertainty.

For the weighted least squares line fit in Figure 2c, the uncertainty envelope is widest at the beginning and end of the analysis, and for approximately evenly spaced reference material analyses is narrowest at its centroid. Thus, the bias correction for any single unknown has its smallest uncertainty in the middle of the session. To calculate the weighted mean of two unknown date determinations, however, requires estimating the correlation coefficient between their two uncertainties, plotted in Figure 2d. For instance, the dark blue line shows the correlation coefficient between an unknown measured at time $t = 0$, marked with a black point, and another unknown measured later in the session, traced by the blue line. The unknown measured at the start of the session is positively correlated with other unknowns from the beginning of the session, has an uncertainty correlation near zero for an unknown measured near $t = 5500$ s, and has an uncertainty that is negatively correlated with any other unknown at the end of the session. The five other colored lines trace the correlation coefficients between times marked by black points, and another unknown from the rest of the session.

The correlation structure illustrated in Figure 2d for the weighted least squares linear fit guarantees that when evaluating the weighted mean of any two unknowns, the systematic uncertainty propagated from sample-standard bracketing is equal in magnitude to the uncertainty contribution from the weighted mean of the same data set. Qualitatively, this makes sense: two unknowns from near the middle of the session have smaller uncertainties in their bias correction factors, but those uncertainties are positively correlated, increasing the total uncertainty contribution to their weighted mean; an unknown from the beginning and another from the end of the session have larger uncertainties in their bias correction factors, but those uncertainties are negatively correlated, which decreases the total uncertainty contribution to their weighted mean. The two scenarios end up with equal uncertainty contributions from sample-standard bracketing.

In detail, the number of unknowns required to reach the minimum systematic uncertainty contribution from sample-standard bracketing depends on the number of parameters used in the function to fit the data. A weighted mean has one parameter (the mean), a line has two (a slope and y intercept), the shifted exponential of the form $y=a\exp(b\,x)+c$ used in *ET_Redux* has three parameters. A smoothing spline can have a varying and noninteger number of effective parameters, depending on the level of smoothing. The smoothest spline, with a near-infinite value of λ, is linear and has two effective parameters; an interpolating spline that has maximum roughness and connects every data point has n effective parameters, where n is the number of standards measured.

All realistic smoothing splines, which take into account measured uncertainties and estimated overdispersion (see section 2.2), will have an intermediate number of effective parameters, which can be estimated using a formula. The number of data degrees of freedom for a model, like a parametric or spline fit, is equal to the number of measurements made minus the number of parameters used in the model. The Welch-Shatterthwaite formula estimates the number of effective data degrees of freedom, df for a complex model such as a spline,

$$df=\mathrm{trace}(H^{\mathsf{T}}H)/\mathrm{trace}(H^{\mathsf{T}}H\,H^{\mathsf{T}}H) \tag{5}$$

where H is the hat matrix for the model, or linear approximation of the function that turns its measured values, y, into its model fit values $\hat{y}$ in the equation $\hat{y}=Hy$. The hat matrix for a spline depends on the measured data and the smoothing parameter λ chosen to fit the data [*Green and Silverman*, 1993], and the effective number of model parameters then becomes the number of measured values minus df in equation (5).

The spline plotted for the 33 standards in Figure 2e therefore has ~23.8 data degrees of freedom, and $33-23.8=\sim 9.2$ effective model parameters. The correlation structure illustrated in Figure 2f shows that the high number of effective parameters decreases the uncertainty correlation between unknowns measured at nearby times in the session, so that unknowns measured ~3000 s apart are effectively independent. Therefore, for spline fits, minimizing the uncertainty contribution of sample-standard bracketing is best accomplished by spacing unknowns as far apart as possible. The green uncertainty envelope for the spline fit in Figure 2e also narrows as the data density of the standards increases. If multiple reference materials are analyzed back-to-back due to time considerations in moving the laser stage to a separate mount, then unknowns are also optimally placed immediately before or after these clusters, in close temporal proximity to multiple reference material analyses.

3. Data Reduction and Uncertainty Propagation Algorithm Outline

ET_Redux propagates uncertainties from the measured intensities of ion beams through to radio-isotopic U-Th-Pb dates, corrected for initial common Pb where appropriate, and quantitative interpretations based on those dates, such as weighted means, regressions through discordant arrays of data, and kernel density estimates. All data reduction protocols generally follow the same established pattern, starting with a background or gas blank subtraction from an on-peak measurement, correction for any isobaric interferences (e.g., ^{204}Hg on ^{204}Pb), calculation of the relative abundances of U, Th, and Pb isotopes throughout the ablation, utilizing a set of bracketing reference materials to quantify laser and mass-spectrometer-induced elemental and isotopic fractionation, applying these average or time-dependent corrections to simultaneously measured unknowns, performing a common Pb correction if necessary, and then often interpreting multiple unknown analyses together, for instance in a weighted mean or kernel density function. Data reduction and uncertainty propagation in *ET_Redux* follows this blueprint but utilizes the innovative mathematical approaches described in section 2, employs the matrix-based full uncertainty propagation protocols described in *McLean et al.* [2011], and treats downhole laser-induced fractionation using both the "intercept" [e.g., *Sylvester and Ghaderi*, 1997; *Košler and Sylvester*, 2003; *Gehrels et al.*, 2008] and "downhole" [*Paton et al.*, 2010] methods. The data reduction and uncertainty propagation algorithms presented here broadly parallel the scientific workflow presented in section 4.

3.1. Baseline Subtraction

Raw data from mass spectrometer run files are read in as ion beam intensities, and uncertainties are assigned to each integration in the baseline based on its measured intensity (see section 2.2) and, if an ion counter is used, its dead time. A parametric regression is then fit to the baseline (e.g., a mean for a stable

baseline or a decaying exponential for a washout curve), and extrapolated under the on-peak integrations. The uncertainty in the fit parameters of this regression is then propagated into the on-peak, baseline-subtracted measurements. Because the uncertainty in the extrapolation of the baseline measurements affects every baseline-corrected on-peak intensity, it acts as a systematic uncertainty, increasing both the uncertainties of the baseline-corrected on-peak intensities and the uncertainty correlations between them.

3.2. Calculating Log-Ratios

Baseline-corrected on-peak data are then converted into the log-ratio compositional data structure of Aitchison (section 2.1). *ET_Redux* calculates $\log\left(^{206}\mathrm{Pb}/^{207}\mathrm{Pb}\right)$, $\log\left(^{206}\mathrm{Pb}/^{238}\mathrm{U}\right)$, $\log\left(^{208}\mathrm{Pb}/^{232}\mathrm{Th}\right)$ ratios, as well as $\log\left(^{206}\mathrm{Pb}/^{204}\mathrm{Pb}\right)$, $\log\left(^{207}\mathrm{Pb}/^{204}\mathrm{Pb}\right)$, and $\log\left(^{208}\mathrm{Pb}/^{204}\mathrm{Pb}\right)$ when $^{204}\mathrm{Pb}$ is measured and above the detection limit.

3.3. Correction for Laser and Mass-Spectrometer-Induced Elemental and Isotopic Fractionation

Mineral and/or glass reference material measurements during a session are used to correct for laser and mass-spectrometer-induced elemental and isotopic fractionation (LMSIEIF) by comparing the best estimate of the true value of the isotope ratio of interest to the measured isotope ratio. We term the multiplicative estimate of fractionation ϕ, so that

$$(X/Y)_{true} = \phi\,(X/Y)_{meas} \tag{6}$$

and for the log-ratios used to ensure proper compositional data handling (see section 2.1),

$$\log(X/Y)_{true} = \Phi + \log(X/Y)_{meas} \tag{7}$$

The value of $\Phi = \log(\phi)$ estimated by repeated measurement of the reference material is then used to transform the measured value of an unknown isotope ratio into its estimated true value.

There are two popular conceptual models that correct for LMSIEIF during the laser ablation process, known as "intercept" and "downhole" corrections. The "downhole" model [*Paton et al*., 2010] is more restrictive, with fewer model parameters and more stringent assumptions about the physical behavior of the laser ablation system. If these assumptions are met, it is therefore capable of calculating smaller uncertainties for the unknowns. The intercept method allows for variation in the downhole fractionation, and therefore requires more parameters to describe the physical behavior of the system. Because it makes less stringent assumptions on fractionation behavior, the intercept method is more flexible, but produces larger uncertainties in the unknowns. *ET_Redux* allows users to employ both methods, evaluate which set of assumptions best fits a data set, and then records these choices alongside the reduced data for full traceability.

3.3.1. Intercept Method

The "intercept" method assumes that the value of Φ at the beginning of the ablation is the same for reference materials and unknowns, though the downhole fractionation behavior may vary from reference material to reference material and unknown to unknown [e.g., *Sylvester and Ghaderi*, 1997; *Košler and Sylvester*, 2003; *Gehrels et al*., 2008]. The change in the $\Phi(t)$ for time-zero intercepts over time t during the session can be estimated by evaluating the time-zero intercept of parametric models fit to downhole fractionation trends. Because estimating time-zero intercepts involves extrapolation, only parametric functions can be used to fit downhole fractionation trends; splines are more appropriately used to interpolate between measurements.

3.3.2. Downhole/Iolite

Unlike the intercept model, the downhole correction model advocated by [*Paton et al*., 2010] and used in the software package Iolite assumes that the laser-induced downhole fractionation trend is the same for each reference material and unknown, though the average of this trend may shift toward higher or lower fractionations during the session. Therefore, LMSIEIF for each integration i of a reference material during the session time t can be modeled as the sum of two functions, $\Phi(i,t) = \Phi_1(i) + \Phi_2(t)$, where $\Phi_1(i)$ represents the average downhole fractionation trend of all the reference materials, and $\Phi_2(t)$ represents the shifts in its average value during the session.

3.4. Correcting Unknowns

For the intercept method, a parametric fit to the measured log-ratios of the unknowns can be corrected using equation (7) estimated from the standard values. The uncertainty in the unknown log-ratio then

derives from both terms in that equation: the uncertainty in the intercept of the measured log-ratios and the uncertainty in $\Phi(t)$. For the downhole method, each integration of an unknown measurement may be corrected for LMSIEIF, using the fit functions to $\Phi_1(i)$ and $\Phi_2(t)$. The uncertainty in the unknown log-ratio is then taken as the uncertainty in the weighted mean of these integrations, plus the uncertainty in $\Phi_1(i)$ and $\Phi_2(t)$. For both methods, use of a common set of reference materials to correct for LMSIEIF, embodied in the function Φ means that the uncertainties in the unknowns are correlated, as discussed in section 2.3.

3.5. Common Pb Correction

ET_Redux supports two different common Pb (Pbc) correction protocols. For samples where ^{204}Pb is both measured and above the detection limit, defined as $^{204}\text{Pb} - 2\sigma \geq 0$, a ^{204}Pb-based correction is made based on either a user-input Pbc isotopic composition (IC) or an IC whose Stacey-Kramers age [*Stacey and Kramers*, 1975] agrees with the ^{206}Pb/^{207}Pb, ^{206}Pb/^{238}U, or ^{208}Pb/^{232}Th date, as chosen by the user. For the Stacey-Kramers option, this agreement is reached quickly via an iterative Newton-Raphson solver. If there is no usable ^{204}Pb data, then a ^{207}Pb-based correction can be made using a user-input or Stacey-Kramers model ^{206}Pb/^{207}Pb ratio. This makes the assumption that the unknown analyses are perfectly concordant, so *ET_Redux* will not plot them on a concordia diagram and will only give the resulting corrected dates in the data table.

Two different types of uncertainty in the Pbc IC are required when evaluating the weighted mean of multiple analyses. The first is an estimate of the grain-to-grain variability in the Pbc IC and is propagated as a random contribution to a weighted mean uncertainty and can be reduced as the number n of analyses included in the weighted mean increases. The second estimated uncertainty is the systematic uncertainty in the mean of the Pbc IC, which is propagated into weighted means and does not decrease with n.

3.6. Data Interpretation: Weighted Means, Concordia Plots, Upper, and Lower Intercepts

For each selected date interpretation, the results include the calculated date with the uncertainty shown in the form X/Y/Z/W, calculated using the generalized weighted mean algorithm presented in *McLean et al.* [2011]. Here X refers to the analytical uncertainty alone, which takes into account the uncertainty correlations from sample-standard bracketing, Y also includes the lab's observed variability among multiple standards of the same mineral, Z additionally includes the systematic uncertainty in the mineral standard isotopic composition, and W adds in the decay constant uncertainties. Also shown are the χ^2_{red} and the count of fractions used.

Uncertainty ellipses on concordia plots and the upper and lower intercepts of lines fit through linear arrays of U-Pb data are calculated using the analytical uncertainty propagated through calculations for the unknowns, including the measurement and sample-standard bracketing uncertainties included in the X term of the weighted mean. Uncertainty correlations are not assumed to be zero, even in the Tera-Wasserburg plot of ^{238}U/^{206}Pb versus ^{207}Pb/^{206}Pb. These two ratios, for instance, share a common uncertainty: they both rely on the relative abundance of ^{206}Pb. The uncertainty correlation between the two is propagated fully through the matrix form of the uncertainty propagation equation from the measured intensities of each isotope.

4. Cyberinfrastructure to Support Workflow From Measurement to Database

The development process for *ET_Redux* for LA-ICPMS U-Pb geochronology began with understanding and defining the scientific workflows—including every input, parameter, procedure, algorithm, visualization, report, and potential use of partial and completed results. This most often involves populating a grain mount with known reference materials and unknown samples, and then sequentially ablating a series of these reference materials and unknowns while measuring the relative abundances of the isotopes of mercury, lead, thorium, and uranium. *ET_Redux* uses templates representing different mass spectrometer models to parse measured data from accepted data-output formats. In order to properly correct for elemental and isotopic fractionation and to propagate uncertainties, the entire session is recorded and modeled as a time series inside the software. The following sections summarize and illustrate the key elements of the resulting scientific workflow.

4.1. Initialization

The user first chooses a mass spectrometer and file protocol, then *ET_Redux* scans the raw data file and prompts for, depending on the context, the cycle duration and dead time, relative gain, integration time, and resistor values for the collector(s). The user then allocates analyses as primary and secondary reference materials and unknowns and chooses the mineral reference material models, which contain elemental and isotopic ratios and concentrations, used in the acquisition. *ET_Redux* supports the creation and archiving of new named and versioned parameter models, including mineral reference material models, to encode the TIMS-generated ratios and uncertainties, a reference for relevant source publications, and the calculated apparent dates.

4.2. Acquisition of Data

The user now chooses whether to deploy the "live workflow" mode to monitor each ablation analysis in real time, or to load and process raw intensity data from a completed file. After loading the data, *ET_Redux* attempts to parse the acquisitions into standards and unknowns, requiring the user to complete the task with an intuitive drag-and-drop interface. The user then establishes the parameters used for any required common Pb correction (see section 3.5).

4.3. Preparing the Raw Data

ET_Redux provides a raw data management window with visualizations and interpretive tools. The user can choose to view and filter the measured intensities including the baseline and on-peak intensities, baseline-corrected intensities, and raw ratios. *ET_Redux* presents the raw data as a grid of interactive and resizable graphs. The columns represent the reference material and unknown acquisitions arranged in temporal order, and the rows represent the data of interest, such as intensities or ratios. The user can show or hide a manually scalable local *y* axis and explore each graph individually. The user can also reject outliers or unwanted portions of the analysis. This rejection is synchronized across the analysis, so that, for instance, if the first ratio is rejected from the $^{206}Pb/^{238}U$, the first ratio is also rejected from the other measured ratios. A complete reference material or unknown analysis can also be rejected.

To prepare the raw data, the user is first presented with a grid of graphs showing the measured baseline and on-peak intensities for the reference materials, including a fit function for the baseline. The unknowns can also be viewed and managed in this grid. Figure 3 shows a portion of this screen for an example set of standards. Another view shows the same grid with the baseline-corrected and isobaric interference-corrected intensities, including the subtraction of ^{204}Hg from ^{204}Pb using ^{202}Hg (see Figure 3).

The third view in the workflow is of ratios of interest including $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, and $^{208}Pb/^{204}Pb$. In all three cases, if ^{204}Pb is below detection, the graph displays "below detection" and does not show any points for ratios involving ^{204}Pb.

In all views displaying ratios, the user may choose to view the ratios as real numbers, as log-ratios, or as a fractionation-factor denoted by alpha. See section 2.1 for a discussion of log-ratios and their use in this context.

4.4. Fitting the Reference Materials

The next step in the workflow is to fit the ratios of the reference materials with functions to support fractionation correction either by the downhole [*Paton et al*., 2010] or by the intercept technique.

For the downhole technique, *ET_Redux* presents an overlay graph for each ratio with individual lines plotting over time the acquisitions for all the reference materials, their mean, and a choice of a calculated fit function for that mean from a choice of mean, line, or exponential with and without over dispersion (see section 2.2). Then, the user can view an individual graph of each reference material with a line plotted through the pointwise difference between each acquisition's log-ratio and the corresponding log-ratio of the fit function. The user may then select to fit a mean or a mean with overdispersion to this difference for each standard. These means and their uncertainties are used to populate the session view with all the reference material plotted on a time axis, and which also provides a choice of fit functions for the session, including a configurable smoothing spline fit.

For the intercept technique, *ET_Redux* presents an individual plot of each standard's ratios and a choice of precalculated fit functions for each, fit with and without overdispersion. The intercepts of these fits at time

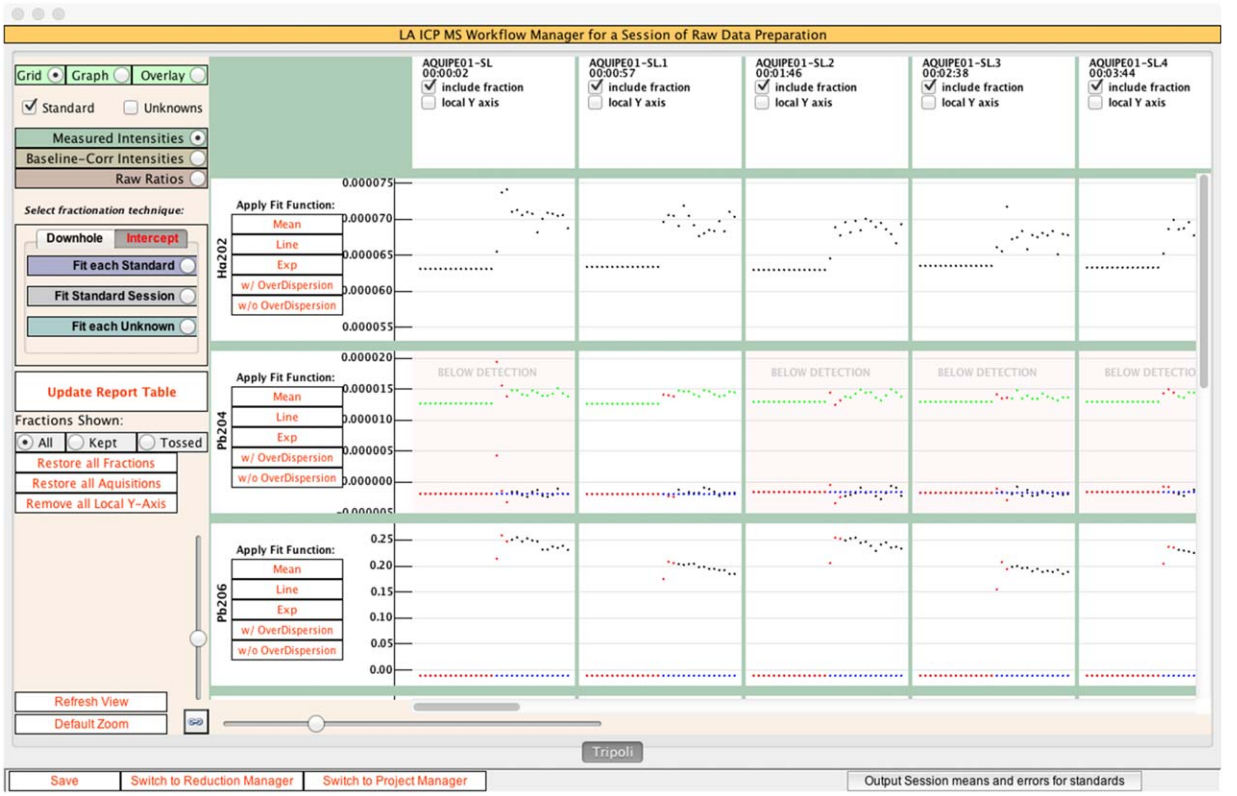

Figure 3. Screen shot of the raw data manager window of *ET_Redux* showing the measured intensity data for five reference material analyses. Each small graph shows the baseline in the left and the on-peak values on the right half of the plot. The graphs for ^{204}Pb show the estimated ^{204}Hg isobaric interference from ^{202}Hg data as a green overlay. In all but one case shown, the Hg-corrected ^{204}Pb is below the detection limit. A function, in this case a mean, is fit to the baseline and shown in red projected in blue under the on-peak readings to illustrate the magnitude of the baseline correction.

zero on the *y* axis and their associated uncertainties are the values used to populate the session fit described in the previous paragraph.

In all cases, *ET_Redux* provides the user with a wide array of choices, ranging from selecting one type of function (e.g., line) for each ratio for all reference materials, to choosing a different function for each ratio for each analysis (e.g., mean for ^{207}Pb/^{206}Pb, line for ^{206}Pb/^{238}U). Any changes made in views of the reference materials are reflected immediately in the session view. Figure 4 shows a view of the session with fit functions for each ratio.

4.5. Correcting Unknowns

The user manages the unknowns employing the same views of baseline, baseline-corrected intensities, and raw ratios used for the reference materials. Again, outliers and bad analyses can be excluded. Changes to unknowns are reflected immediately in the data table in the main window.

For the downhole technique, *ET_Redux* provides the user with plots and the choice of fitting a mean function with or without overdispersion to each ratio for each unknown, and the offsets from the bracketing reference material and their uncertainties are used with the values provided by the fitting function of the session plot to correct the unknowns.

For the intercept technique, *ET_Redux* provides the user with plots and function choices for each ratio for each unknown and the *y* intercepts and their uncertainties are used with the values provided by the fitting function of the session plot to correct the unknowns.

4.6. Final Calculations

Because the unknowns are correlated with each other and with the reference materials (see section 2.3), the final calculation of the correlation factors and the common lead corrections must wait until the user has

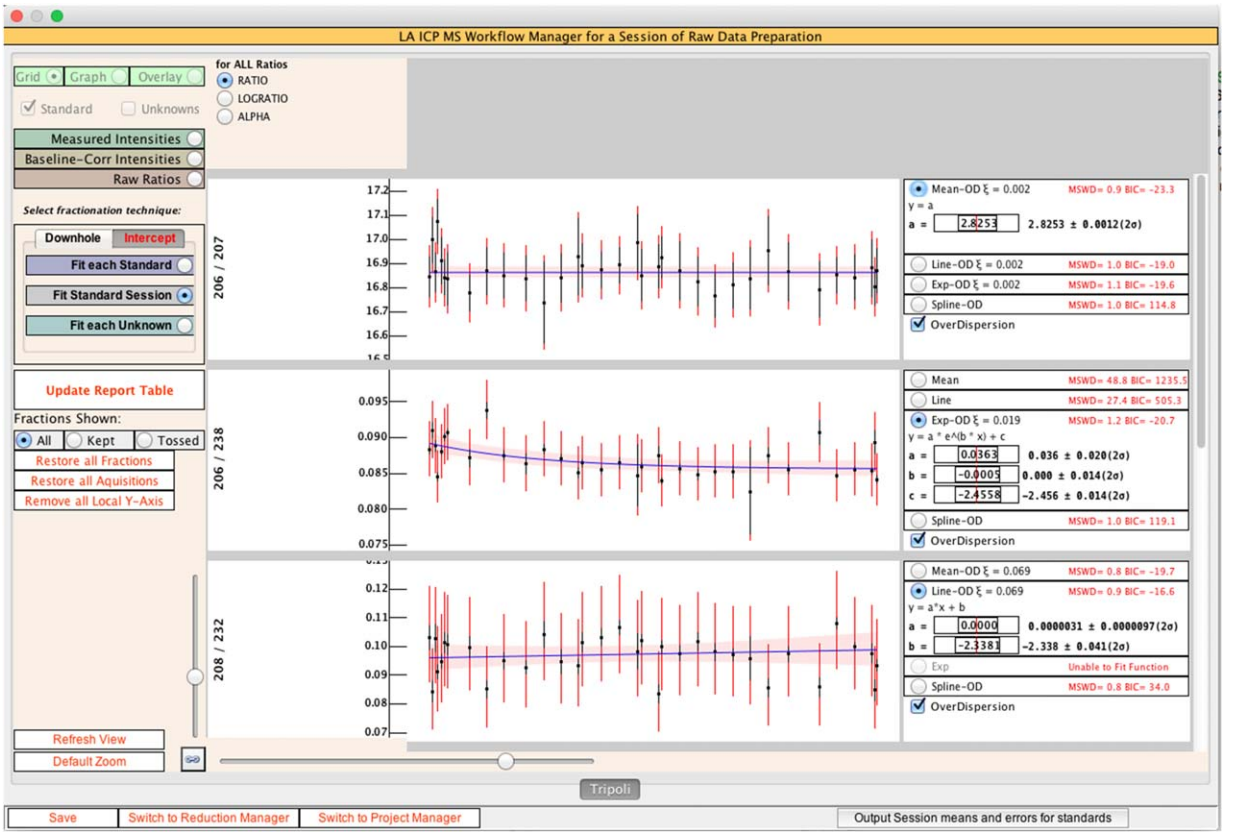

Figure 4. Screen shot of the raw data manager window of *ET_Redux* showing the session fits for 206/207, 206/238, and 208/232. The whisker plots show the mean for each included standard with the black whisker denoting the regular uncertainty and the red portion of the whisker denoting the added uncertainty due to overdispersion in the session fit function. In the right-hand boxes are the fit function choosers, with the selection shown and the pertinent parameter values displayed.

made the decisions outlined above. The user then clicks the "update data table" button to calculate uncertainty correlations and common lead corrections. These results are propagated to the data table and made available for plotting.

4.7. Data Table

The interactive data table is always visible and responsive to changes that the user makes in preparing the raw data as a way to provide real-time feedback and sensitivity analysis. The data table is fully configurable as described in *Bowring et al*. [2011] and easily exported for publication. The user can exclude or include analyses and sort the table on any column.

4.8. Plots

ET_Redux extends the original interactive date interpretations manager developed for ID-TIMS with the addition of probability density plot (PDP) and kernel density estimator (KDE) plots [*Ludwig*, 2003; *Vermeesch*, 2012] for detrital samples in addition to improved concordia and weighted mean plots.

For single age analyses, the analyst can choose one or more defined date interpretations for LA-ICPMS data based on a subset of analyses. These include single grain, weighted means, and upper/lower intercept dates. The user selects one of the interpretations as "preferred" and all activated interpretations become part of the archived analysis.

For detrital analyses, *ET_Redux* provides both probability density (PDP) plots and kernel density (KDE) plots. These plots automatically detect and label peaks, provide configurable histograms on demand, and are zoomable on the *x* axis. The user can also operate sliders that constrain the allowable negative and positive percent discordance and the uncertainty. A special slider on the concordia also allows the user to specify the numerical transition for the "best date" between the $^{206}Pb/^{238}U$ for younger and the $^{207}Pb/^{206}Pb$ for older dates.

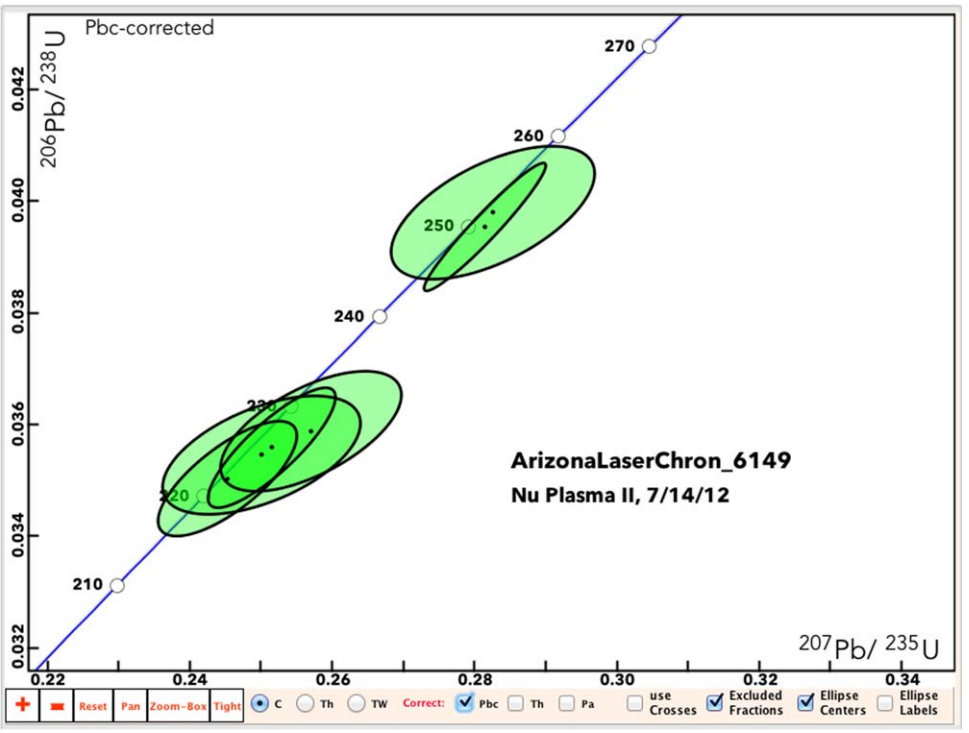

Figure 5. Screen shot of an example interactive concordia plot produced by *ET_Redux*. This image can be saved in vector format as.pdf or.svg files for easy incorporation in posters and publications.

4.9. Publication-Ready Documents

ET_Redux produces publication-ready documents, including data tables, concordia, weighted mean, and PDP/KDE plots. Output formats include scalable vector graphics (SVG), portable document format (PDF), text (TXT) files, or Microsoft Excel spreadsheets (XLS), as required. An example concordia plot, as shown in *ET_Redux*, ready for export, is shown in Figure 5.

4.10. Archiving and Compiling

ET_Redux provides archiving and retrieval services, including compilation, of analysis details and visualizations. The archived results are stored in a self-contained extensible markup language (XML) document. *ET_Redux* interacts with the NSF-sponsored community Geochron.org archival database at http://geochronportal.org to enable optional archiving and retrieval of these files. In compilation mode, several archived aliquots from different labs are compiled into a "project" that allows the analyst to interpret large data sets by viewing multiple samples side-by-side.

The key to *ET_Redux*'s archiving and compilation capabilities is the use of International Geo Sample Numbers (IGSNs) to provide unique identifiers for each sample and analysis and the use of the System for Earth Sample Registration (SESAR) to store and manage meta data about each sample and analysis. More information is available at www.geosamples.org.

5. Conclusions and Plans

ET_Redux is becoming the blueprint for a new generation of cyberinfrastructure tools to support the workflows and data analysis of geochronology with statistical rigor in all aspects of data reduction, from calculation of corrected isotopic ratios to correction for interferences and drift. Its evolution from ID-TIMS U-Pb geochronology to LA-ICPMS U-Pb geochronology is an important step in standardizing approaches to analysis and in establishing the foundations for long-term sustainability for scientific software because it is free and open source and developed and improved continuously with input from the international earth sciences community. We intend to further publicize our online community including the source code for our projects based at http://github.com/cirdles and to encourage users of the software to participate. As part of

our development-as-evolution strategy, we will make continuous evaluations and improvements to the software.

Appendix A: Levenberg-Marquardt Pseudocode

```
ET_Redux Levenberg-Marquardt fit using Covariance Matrix
  given column vectors of n measured intensities, y, and n associated times, t,
  a covariance matrix for the n intensities, Si, and
  the column vector pod with initial values for parameters a, [b, c] and ξ².
  Note: pod is sized m according to meanOD(2), lineOD(3), exponentialOD(4).
  Note: pod is sized m=3 for exponentialFast and exponentialMatrix, with no ξ².
begin
  maxIterations: = 100
  lambda: = 1                                   scalar damping factor
  onesv(n,1): = 1                               column vector of n ones
                                                Initialization
  yhat(n,1): = calc_yhat(pod, t, onesv)         y values at times t predicted by pod
  r(n,1): = y – yhat                            residuals
  Sod(n,n): = calcSod(S, pod)                   total cov. matrix is meas. unct. plus overdispersion
  L: = calcL(r,Sod)                             log likelihood
  G(m,1): = 0.0                                 initialize gradient: column vector of m entries
  H(m,m): = 0.0                                 initialize Hessian: square matrix of m×m entries
  calcGH(t, r, Sod, pod, onesv, G, H)           calculate G and H
  BIC: = 0.0                                    Bayesian Information Criterion, becomes nonzero with solution
for i:=1 to maxIterations do
    if BIC <> 0 then exit fi                    exit loop when solved
    podNew(m,1): = update_pod(pod)              calculate updated parameters
    yhat(n,1): = calc_yhat(podNew, t, onesv)    update predicted values
    SodNew(n,n): = calcSod(S, podNew)           update total covariance matrix
    rNew(n,1): = y – yhat                       update residuals
    LNew: = calcL(rNew, SodNew)                 scalar per function below
if (LNew > L) & |1 - Lnew/L| >= chiTolerance    solution worse than default
  then
    lambda: = lambda * 10.0
  else if |1 - Lnew/L| >= chiTolerance          solution improving
      then
        pod: = podNew                           adopt new calculated parameters
        lambda: = lambda/10.0
        L: = LNew
        r: = rNew
        Sod: = SodNew
        calcGH(t, r, Sod, pod, onesv, G, H)     calculate G and H
      else                                      solution found
        params: = podNew                        save calculated parameters
        L: = LNew
        MSWD: = calcMSWD(rNew, S)               calculate MSWD at solution
        Vp(m,m): = H⁻¹                          covariance matrix for a, [b, c], ξ²
        BIC: = 2 * L + m*log(n);                causes loop to terminate on BIC <> 0
    fi
  fi
 od
end
```

Acknowledgments
Funding for this work was provided through NSF awards #0930223 and #0929746 and NERC grant NE/I013814/1. The authors offer special thanks to Doug Walker, Sam Bowring, Jeff Vervoort, and Dan Condon for their support and encouragement. We acknowledge the support of the College of Charleston, MIT, the British Geological Survey, the University of Kansas, and the University of Arizona in making this work possible. This paper is theoretical in nature and all the numerical information provided in the figures is for illustration only. No new data are presented here.

References

Aitchison, J. (1984), The statistical analysis of geochemical compositions, *Math. Geol.*, *16*, 531–564, doi:10.1007/BF01029316.
Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, 416 pp., Chapman and Hall, London, U. K.
Aitchison, J. (1992), On criteria for measures of compositional difference, *Math. Geol.*, *24*, 365–379, doi:10.1007/BF00891269.
Blackburn, T. J., P. E. Olsen, S. A. Bowring, N. M. McLean, D. V. Kent, J. Puffer, G. McHone, E. T. Rasbury, and M. Et-Touhami (2013), Zircon U-Pb geochronology links the end-triassic extinction with the central Atlantic magmatic province, *Science*, *340*(6135), 941–945, doi:10.1126/science.1234204.
Bowring, J. F., N. M. McLean, and S. A. Bowring (2011), Engineering cyber infrastructure for U-Pb geochronology: Tripoli and U-Pb_Redux, *Geochem. Geophys. Geosyst.*, *12*, Q0AA19, doi:10.1029/2010GC003479.
Fisher, C. M., H. P. Longerich, S. E. Jackson, and J. M. Hanchar (2010), Data acquisition and calculation of U-Pb isotopic analyses using laser ablation (single collector) inductively coupled plasma mass spectrometry, *J. Anal. At. Spectrom.*, *25*(12), 1905–1920, doi:10.1039/C004955G.
Galbraith, R. (2005), *Statistics for Fission Track Analysis*, Chapman and Hall, Boca Raton, Fla.

Gehrels, G. E., V. A. Valencia, and J. Ruiz (2008), Enhanced precision, accuracy, efficiency, and spatial resolution of U-Pb ages by laser ablation-multicollector-inductively coupled plasma-mass spectrometry, *Geochem. Geophys. Geosyst.*, *9*, Q03017, doi:10.1029/2007GC001805.

Green, P., and B. Silverman (1993), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Monogr. Stat. Appl. Probab.*, vol. 58, Chapman and Hall, London, U. K.

Horstwood, M., G. Gehrels, and J. Bowring (2010), Improving consistency in laser ablation geochronology; workshop on data handling in LA-ICP-MS U-Th-Pb geochronology; San Francisco, California, 12–13 December 2009, *Eos Trans. AGU*, *91*(28), 247–247, doi:10.1029/2010EO280003.

Horstwood, M. S. A., et al. (2016), Community-derived standards for LA-ICP-MS U-Th-Pb geochronology—Uncertainty propagation, age interpretation and data reporting, *Geostand. Geoanal. Res.*, doi:10.1111/j.1751-908X.2016.00379.x.

Košler, J., and P. J. Sylvester (2003), Present trends and the future of zircon in geochronology: Laser ablation ICPMS, *Rev. Mineral. Geochem.*, *53*(1), 243–275.

Ludwig, K. R. (2003), *Isoplot/Ex Version 3.00: A Geochronological Toolkit for Microsoft Excel, Spec. Publ. 4*, 73 pp., Berkeley Geochronol. Cent., Calif.

McLean, N. M., J. F. Bowring, and S. A. Bowring (2011), An algorithm for U-Pb isotope dilution data reduction and uncertainty propagation, *Geochem. Geophys. Geosyst.*, *12*, Q0AA18, doi:10.1029/2010GC003478.

Nyquist, H. (1928), Thermal agitation of electric charge in conductors, *Phys. Rev.*, *32*(1), 110–113, doi:10.1103/PhysRev.32.110.

Paton, C., J. D. Woodhead, J. C. Hellstrom, J. M. Hergt, A. Greig, and R. Maas (2010), Improved laser ablation U-Pb zircon geochronology through robust downhole fractionation correction, *Geochem. Geophys. Geosyst.*, *11*, 1525–2027, doi:10.1029/2009GC002618.

Schoene, B. (2014), U–Th–Pb geochronology, in *Treatise on Geochemistry, vol. 4, The Crust*, 2nd ed., edited by H. D. Turekian and K. K. Holland, pp. 341–378, Elsevier, Oxford, U. K., doi:10.1016/B978-0-08-095975-7.00310-7.

Schottky, W. (1918), Über spontane stromschwankungen in verschiedenen elektrizitätsleitern, *Ann. Phys.*, *362*(23), 541–567, doi:10.1002/andp.19183622304.

Stacey, J., and J. Kramers (1975), Approximation of terrestrial lead isotope evolution by a two-stage model, *Earth Planet. Sci. Lett.*, *26*(2), 207–221, doi:10.1016/0012-821X(75)90088-6.

Sylvester, P. J., and M. Ghaderi (1997), Trace element analysis of scheelite by excimer laser ablation-inductively coupled plasma-mass spectrometry (ELA-ICP-MS) using a synthetic silicate glass standard, *Chem. Geol.*, *141*(1–2), 49–65, doi:10.1016/S0009-2541(97)00057-0.

Ulianov, A., and O. Müntener (2014), Differencing as a method to estimate the uncertainty of a transient LA-ICPMS signal, *J. Anal. At. Spectrom.*, *29*(5), 934–940, doi:10.1039/C4JA00004H.

Ulianov, A., O. Müntener, U. Schaltegger, and F. Bussy (2012), The data treatment dependent variability of U-Pb zircon ages obtained using mono-collector, sector field, laser ablation ICPMS, *J. Anal. At. Spectrom.*, *27*(4), 663–676, doi:10.1039/C2JA10358C.

Ulianov, A., O. Müntener, and U. Schaltegger (2015), The ICPMS signal as a Poisson process: A review of basic concepts, *J. Anal. At. Spectrom.*, *30*, 1297–1321, doi:10.1039/C4JA00319E.

Vermeesch, P. (2010), Helioplot, and the treatment of overdispersed (U-Th-Sm)/He data, *Chem. Geol.*, *271*(3-4), 108–111, doi:10.1016/j.chemgeo.2010.01.002.

Vermeesch, P. (2012), On the visualisation of detrital age distributions, *Chem. Geol.*, *312–313*, 190–194, doi:10.1016/j.chemgeo.2012.04.021.

Wotzlaw, J.-F., U. Schaltegger, D. A. Frick, M. A. Dungan, A. Gerdes, and D. Günther (2013), Tracking the evolution of large-volume silicic magma reservoirs from assembly to supereruption, *Geology*, *41*(8), 867–870, doi:10.1130/G34366.1.