



Quantifying Long-Term Reproducibility of Zircon Reference Materials by U-Pb LA-ICP-MS Dating

Jakub T. Sliwinski (1, 3)* , Marcel Guillong (1) , Matthew S. A. Horstwood (2)  and Olivier Bachmann (1) 

(1) Institute of Geochemistry and Petrology, ETH Zürich, Clausiusstrasse 25, 8092, Zürich, Switzerland

(2) National Environmental Isotope Facility, Geochronology and Tracers Facility, British Geological Survey, Nicker Hill, Keyworth, NG12 5GG, Nottingham, UK

(3) Present address: School of Earth and Environmental Sciences, University of St Andrews, Bute Building, Queen's Terrace, KY1 6 9TS, St Andrews, Scotland

* Corresponding author. e-mail: jts21@st-andrews.ac.uk

Inter-session excess variance of U/Pb and Pb/Pb ratios in LA-ICP-MS zircon dating is the largest contributor to systematic errors, which in turn limit the accuracy of age determinations. Quantifying long-term excess variance of reference materials allows for the estimation of excess variance in samples, but such compilations are not available in the literature. Here, we present the results of over 100 measurement sessions over 6 years for seven common zircon reference materials and calculate a characteristic excess variance for each. For $^{206}\text{Pb}/^{238}\text{U}$ ages, these values ($2s$) are as follows: AusZ7-1 = 1.7% (115 sessions), AusZ7-5 = 0% (74), OD-3 = 0.6% (19), Temora2 = 1.2% (86), Plešovice = 1.1% (100), 91500 = 1.0% (146) and Mud Tank = 3.0% (12). For $^{207}\text{Pb}/^{206}\text{Pb}$ ages, smaller excess variances are observed: Temora2 = 0.3%, Plešovice = 0.4%, Mud Tank = 1.7% and 91500 = 0.4%. These values are well-constrained estimates of inter-session excess variance for the ETH Zürich LA-ICP-MS laboratory using a sector-field ICP-MS, and may provide either a first-order estimate for other laboratories or a value to compare against their own reference material compilations.

Keywords: LA-ICP-MS, zircon, reproducibility, excess variance, reference materials.

Received 30 Nov 21 – Accepted 22 May 22

Due to its cost-effectiveness, precision and rapidity of data acquisition, laser ablation inductively coupled plasma-mass spectrometry (LA-ICP-MS) has revolutionised Earth sciences in the last three decades, including in the field of zircon geochronology. In that time, the precision of zircon U-Th and U-Pb dating has improved to the level of secondary ionisation mass spectrometry (SIMS) with relative measurement precision approaching $\sim 0.5\%$ $2s$ (Guillong *et al.* 2016, Sliwinski *et al.* 2017). However, this number omits uncertainties related to systematic errors, which constrain the 'reproducibility' of validation reference material (VRM) U/Pb ratios, and ultimately limit the overall uncertainty of the method to $\sim 2\text{--}4\%$ (Klötzli *et al.* 2009, Horstwood *et al.* 2016). This is due to so-called 'matrix effects', which describe the variability between samples in laser-induced elemental fractionation (LIEF) and plasma-induced elemental fractionation (here, U/Pb) caused by differences in zircon radiation damage, opacity, crystal orientation, trace element

mass fractions, crystal defects and the robustness of the plasma between sessions (Allen and Campbell 2012, Marillo-Sialer *et al.* 2014, Solari *et al.* 2015, Marillo-Sialer *et al.* 2016, Sliwinski *et al.* 2017).

Systematic uncertainties are by far the greatest impediment to obtaining precise U/Pb zircon data by LA-ICP-MS but are difficult to satisfactorily quantify. Horstwood *et al.* (2016) laid out a consistent framework for laboratories to achieve the best and most reproducible U-Pb ages, which is in large part emulated by commercial data reduction software like Lolite (Paton *et al.* 2011). In brief, this includes the following: (a) the measurement and subtraction of gas blank from the raw signal, calculation of elemental ratios and correction for LIEF; (b) the calculation of the mean of the elemental ratios (e.g., Pb/U); (c) normalisation of this ratio using reference values for the calibration reference material, as well as the excess scatter necessary to make the MSWD

doi: 10.1111/ggr.12442

© 2022 The Authors. *Geostandards and Geoanalytical Research* published by John Wiley & Sons Ltd on behalf of the International Association of Geoanalysts.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

of the calibration $RM = 1$; (d) correction for common Pb and (e) propagation of systematic uncertainties. Uncertainties in steps 1, 2 and 4 are random uncertainties that are propagated into the weighted mean calculations of population ages, while systematic uncertainties (3 and 5) are propagated *after* the weighted means are calculated and can be used to compare the overlap between two populations. For example, two zircon populations with ages 10 ± 1 and 13 ± 1 Ma measured on separate days may seem distinct from one another on the basis of their non-overlapping uncertainties. However, if there is a large source of systematic uncertainty that changes these values to 10 and 13 ± 2 Ma, the same distinction cannot be made. Although some components of systematic uncertainty, such as decay constant uncertainties, model common-Pb ratio uncertainties and reference material isotope ratio uncertainties, impart a small degree of inaccuracy, the long-term excess variance of VRMs (ϵ') is responsible for the largest offsets in data and requires long-term, consistent monitoring to provide a meaningful value (Horstwood *et al.* 2016). The calculation of this value is performed individually by LA-ICP-MS laboratories without the aid of a common data reduction software package, and there is, therefore, a lack of consensus about the best practices for deriving quality data.

The purpose of the present study is to present a case study for quantifying $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ uncertainty through the intra-session excess variance (ϵ), and more importantly, the inter-session long-term excess variance (ϵ') using a suite of well-characterised and commonly used reference materials. These data provide an extensive long-term compilation of zircon reference material U-Pb and Pb-Pb ages and demonstrate some limitations to current LA-ICP-MS dating techniques, shedding light on the considerations needed when interpreting and publishing LA-ICP-MS (and possibly SIMS) data that concern the discrimination of zircon age populations. The geochronology community (specifically those dealing with zircon) can benefit from the data presented here, as we present an example of how to compile large data sets and assess the 'reproducibility' (strictly the *intermediate measurement precision*) of long-term zircon age data.

Experimental design, materials and methods

Reference materials and samples

The zircon reference material GJ-1 (601.86 ± 0.37 Ma, Jackson *et al.* 2004, Horstwood *et al.* 2016) was used as the calibration RM. The validation reference materials used were: AusZ7-1 (~ 38.9 Ma, Kennedy *et al.* 2014), AusZ7-5

(2.4082 Ma ± 0.0022 , von Quadt *et al.* 2016), OD-3 (33.0 ± 0.1 Ma, Iwano *et al.* 2013), Plešovice (337.15 Ma, Sláma *et al.* 2008, Horstwood *et al.* 2016), Mud Tank (731.65 ± 0.49 , Black and Gulson 1978, Horstwood *et al.* 2016), Temora2 (416.78 ± 0.33 Ma, Black *et al.* 2004) and 91500 (1063.51 ± 0.39 Ma, Wiedenbeck *et al.* 1995, Horstwood *et al.* 2016).

LA-ICP-MS instrumentation and data reduction

Data were collected using an ASI Resolution 193 nm ArF laser from within a Laurin Technic S155 constant geometry two-volume ablation cell, connected to a ThermoScientific Element XR sector field ICP-MS. Ablation was performed under a pure He atmosphere ($0.5\text{--}0.7$ l min^{-1}), after which the ablated aerosol was mixed with Ar in the ablation funnel within the cell, before homogenisation in a signal smoothing device prior to ionisation in the plasma. More analytical parameters can be found in Table 1.

Samples were bracketed by the GJ-1 calibration RM every ten to thirty analyses, and each measurement session saw the measurement of each VRM anywhere from six to twenty-eight times (95% confidence interval, median = 13). Data were reduced using VizualAge (Petrus and Kamber 2012) on the Lolite platform [v2.5, Paton *et al.* (2011)]. Although trace elements were occasionally measured together with Hg, U, Th and Pb for geochronology, the data reduction scheme for U-Pb remained unchanged over the course of 6 years, consisting of: (a) baseline subtraction of raw counts on masses 202, 204, 206, 207, 208, 232, 235 and 238; (b) calculation of U/Pb ratios by the mean-of-ratios method; (c) down-hole fractionation by cubic or exponential spline, as relevant (Paton *et al.* 2010); (d) approximate linear instrumental drift correction; and (e) normalisation to the ID-TIMS U/Pb ratio of the chemically abraded (CA) or non-abraded calibration RM (for CA and non-CA samples, respectively). Data were not corrected for Th disequilibrium or alpha radiation dosage.

Because of the complicated and nested nature of the data processing, and the variety of terms used in different sources, a quick note about terminology is warranted here. Horstwood *et al.* (2016) defines two forms of excess variance (scatter). Session-based excess variance (ϵ) is calculated as the total uncertainty that must be added to VRM analyses to bring the MSWD to 1 (i.e., is the VRM forming a coherent age population within the measurement session?). This added uncertainty may account for inhomogeneities in the VRM ages or material properties relative to the calibration RM. Meanwhile, long-term variance of reference materials (ϵ') assesses the variation among all

Table 1.
LA-ICP-MS analytical conditions at the ETH Zürich

Laboratory and sample preparation	
Laboratory name	Dept. of Earth Science, ETH Zürich
Sample type/mineral	Zircon
Sample preparation	Conventional mineral separation, 1-in resin mount, 1 µm polish
Laser ablation system	
Make, model and type	ASI Resolution
Ablation cell and volume	Laurin Technic S155, constant geometry, aerosol dispersion volume < 1 cm ³
Laser wavelength	193 nm
Pulse width	25 ns
Energy density/Fluence	~ 2.0–2.5 J cm ⁻² (1.5–2.5) ^a
Repetition rate	4–5 Hz (2–5)
Spot size	19–30 µm
Ablation rate	~ 75 nm pulse ⁻¹ at 2.5 J cm ⁻²
Sampling mode/pattern	Single hole drilling, ~ 3 cleaning pulses
Carrier gas and flow	100% He, 0.5–0.7 l min ⁻¹
Ablation duration	30–40 s (20–75)
ICP-MS instrument	
Make, model and type	Thermo Element XR SF-ICP-MS
Sample introduction	Ablation aerosol only, squid-like aerosol homogenisation device
RF power	~ 1550 W (1150–1600)
Make-up gas flow	~ 0.95 l min ⁻¹ Ar (gas mixed to He carrier inside ablation cell funnel) (0.92–1.10)
Detection system	Single detector triple mode secondary electron multiplier, analogue, Faraday
Masses measured	202, 204, 206, 207, 208, 232, 235, 238 amu
Integration time per peak (mass)	~ 10 ms (202, 204, 208, 232, 235), ~ 20 ms (238), ~ 75 ms (206, 207)
Integration time per reading	0.25 s
Dead time	20 ns
Typical oxide rate (ThO/Th)	~ 0.15% (0.04–0.47)
Typical (++) rate (Ba ²⁺ /Ba ⁺)	2.5% (0.9–7.3)
Data processing	
Gas blank	~ 20 s (10–30)
Calibration strategy	GJ-1 used as calibration RM; bracketing 2 per ~ 20 samples
Reference material information	601.86 ± 0.37 Ma, Horstwood <i>et al.</i> (2016), U = 312 and Th = 10.8 µg g ⁻¹ (in-house)
Data processing package used	lolite v2.5 using VizualAge
Mass discrimination	Mass bias correction for all ratios normalised to calibration reference material
Uncertainty level and propagation	Ages are quoted at 2s absolute. Propagation was by quadratic addition

^a Typical run conditions; total range of parameter in parentheses.

measurement sessions' means (i.e., is the analyst consistently reporting the same isotope ratio value for each individual VRM?). It is possible that a VRM will demonstrate excellent repeatability within session ($\epsilon \sim 0$) but have a shifted isotopic ratio. Repeat analysis under such conditions may have apparently low ϵ (e.g., Table S2), while ϵ' demonstrates considerable uncertainty (Table 3). Long-term excess

variance (ϵ') can be calculated in IsoplotR (Vermeesch 2018), with slightly different terminology: given a population, IsoplotR calculates a 'dispersion' term, which represents how much uncertainty must be added to each data point (in quadrature) for the population to have an MSWD = 1, that is,

$$\sigma_i = \sqrt{\sigma_w^2 + (\epsilon')^2} \quad (1)$$

where ϵ' represents the long-term excess variance (i.e., 'dispersion' in IsoplotR) and σ_w represents the weighted mean uncertainty accounting for all non-systematic sources of uncertainty [i.e., the output resulting from step 7 in the framework laid out by Horstwood *et al.* (2016)]. The term σ_i represents the final uncertainty and is calculated by iteratively varying ϵ' such that:

$$\text{MSWD} = \frac{1}{n-1} * \sum_{i=1}^n \frac{(x_i - x_{\text{avg}})^2}{\sigma_i^2} = 1 \quad (2)$$

where n is the number of measurements, x is the i th measurement, σ_i is the uncertainty on the measurement and x_{avg} is the weighted mean of n measurements, such that:

$$\bar{x} = \frac{\sum_{i=1}^n (x_i * W_i)}{\sum_{i=1}^n W_i} \quad (3)$$

where W_i is the weight of the measurement, defined as:

$$W_i = \frac{1}{\sigma_i^2} \quad (4)$$

Note that in these calculations, ϵ can be calculated in a similar manner to ϵ' , but while the latter uses session means and weighted mean uncertainties, the former uses individual zircon ratios and uncertainties.

Long-term data aggregation was performed in the statistical software R, using geochronology functions from the IsoplotR package (Vermeesch 2018). Summary statistics were performed in two steps, outlined in Figure 1: (a) data were grouped by session date and identity, and for every group a weighted mean was calculated [using a random effects model and detecting outliers by a modified Chauvenet criterion (Vermeesch 2018)]; (b) the output from step (a) was grouped by identity. Summary statistics from step 1 (Table S2) therefore represent the weighted means of ²⁰⁶Pb/²³⁸U and ²⁰⁷Pb/²⁰⁶Pb ages from each measurement session for each RM (e.g., the weighted means from Plešovice over 100 sessions), together with an excess variance term that describes how repeatable each RM was on each day (ϵ). Summary statistics from step 2 (Table 2)

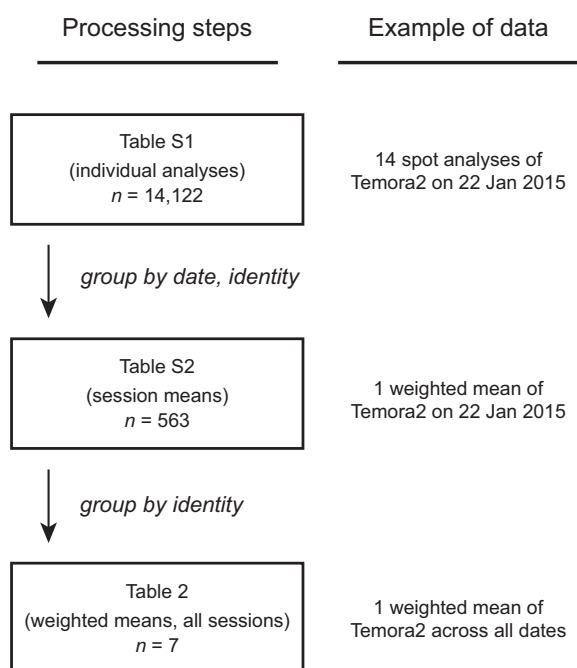


Figure 1. Data processing workflow (left) with an example of data available at each step: (1) Table S1 provides 14122 individual zircon reference material ages. These are grouped by date and identity to generate (2) Table S2 with session-weighted means. Finally, data from Table S2 is grouped by identity to generate long-term weighted means for each reference material and an associated long-term excess variance.

represent the weighted means of the same ratios over all measurement sessions, together with a long-term excess variance term (ϵ') that describes how reproducible each RM was between all sessions. A detailed description of the data tables and their relationships is as follows (and can be found in Figure 1):

- Table 1 (Table_1_instrumentation.docx): LA-ICP-MS analytical conditions.
- Table S1 (Table_S1_all_refs.xlsx): raw data; all reference material zircon U-Pb data ($n = 14122$).
- Table S2 (Table_S2_refs_comp1.xlsx): analysed, filtered data; Compilation of all reference material data (Table S1), with summary statistics reported after grouping by session date and identity ($n = 563$).
- Table 2 (Table_2.docx): analysed, filtered data; compilation of reference material weighted means (Table S2), with summary statistics reported after grouping by identity ($n = 9$).

- Table 3 (Table_3.docx): Summary of age biases in weighted means from each RM.

- Figure S1 (Fig_S1_all_RMs.pdf): Long-term $^{206}\text{Pb}/^{238}\text{U}$ ratios for 91500, Temora2, OD-3, Plešovice, AusZ7-1 and AusZ7-5, with reference ratio marked in red (from Table S2).

- Figure S2 (Fig_S2.pdf): Summary of typical uncertainties on $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ ratios for each reference material within each measurement session (from Table S2). Only 91500, Temora2, Plešovice and Mud Tank are reported with $^{207}\text{Pb}/^{206}\text{Pb}$ ratios.

- Table S3 (Table_S3.xlsx): extended version of Table 2.

Results and discussion

Given the complex nature of the data presented here, an example is provided that demonstrates the relationships of different data sheets. Temora2 was analysed over the course of eighty-six measurement sessions (1229 analyses), and every spot analysis was compiled in Table S1 using the standard iolite output reporting U/Pb and Pb/Pb ratios and ages. The data here allow for an estimate of the typical 1s uncertainty on each ratio, using the median measurement precision across all analyses (Table 2). The ratios from Temora2 are grouped from Table S1 by analysis date into eighty-six individual session weighted means, including MSWDs and ϵ values for $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ ratios (Table S2). Here, a typical ϵ can be estimated as the median ϵ needed to bring the MSWD of each measurement session to unity (Table 2), however individual sessions may have more or less ϵ (Figure S2). Finally, the data from Table S2 are compiled again into a long-term weighted mean and MSWD (Table 2), with outlier rejection. At this stage, ϵ' can be calculated as the additional uncertainty necessary to bring the long-term MSWD to unity.

Within individual measurement sessions, MSWD for RM are typically close to unity yet variable, as are the corresponding excess variances (ϵ) for each session (Table S2). A comparison of these values with typical uncertainties demonstrates the relative importance of ϵ . For example, AusZ7-1 is typically analysed with 1.3% relative uncertainty in $^{206}\text{Pb}/^{238}\text{U}$ (i.e., the median measurement precision in each measurement session (1s) is $\sim 1.3\%$), and typically requires an ϵ of 0% (up to 1.5–2%), demonstrating that the repeatability within session (Figure S2) covers (or masks) any excess variance (i.e., MSWD is close to 1). For the other RMs: AusZ7-5 typically has a measurement precision (1s) and ϵ (1s) of 5.6% and 0% (up to 2–6%), respectively; OD-3 has 0.9% and 1.0%; Plešovice has 0.7% and 0% (up to 2%);

Table 2.
Summary statistics for all reference materials

RM	No. sessions	$^{206}\text{Pb}/^{238}\text{U}^a$	1SE ^b	MSWD	Typ. Prec. (1s %) ^c	Typical ϵ (1s %) ^d	ϵ' (1s %) ^e	$^{207}\text{Pb}/^{206}\text{Pb}^a$	1SE ^b	MSWD	Typ. Prec. (1s %) ^c	Typical ϵ (1s %) ^d	ϵ' (1s %) ^e
AusZ7-1	115	0.00595	5.53E-06	4.7	1.3	0.0	0.84						
AusZ7-5	74	0.00036	7.38E-06	0.8	5.6	0.0	0						
Mud Tank	12	0.1175	5.30E-04	43.6	1.0	0.0	1.47	0.06335	1.84E-04	6.5	1.8	0.0	0.87
OD-3	19	0.005056	6.40E-06	1.6	0.9	1.0	0.28						
Plešovice	100	0.05405	3.44E-05	5.7	0.7	0.0	0.54	0.05334	1.80E-05	2.0	0.9	0.0	0.21
Temora2	86	0.06638	4.71E-05	7.5	0.8	0.0	0.58	0.05524	2.43E-05	1.3	1.4	0.0	0.17
91500	146	0.1784	8.33E-05	7.0	0.8	0.0	0.49	0.07494	2.40E-05	1.7	1.3	0.0	0.20

^a Weighted mean calculated in IsoplotR.

^b 1 standard error of weighted mean.

^c Typical analytical precision on point analyses (1s); calculated as median of all analytical precision measurements; see Figure S2 for full details.

^d Typical excess uncertainty per measurement session (%); calculated as median of all excess uncertainties; see Figure S2 for full details.

^e Long-term excess uncertainty across all measurement sessions (%).

Table 3.
Summary of age biases in various reference materials

	AuZ7-1	AusZ7-5	Mud Tank	OD-3	Plešovice	Temora2	91500
Number of measurement sessions	115	74	12	19	100	86	146
Prior to ϵ' propagation							
% of means with age bias (%) ^a	74	5	92	58	58	59	51
% of age biases that are negative (%)	100	75	100	100	3	90	95
After ϵ' propagation							
% of means with age bias (%)	47	5	33	58	26	16	14
% of age biases that are negative (%)	100	75	100	100	4	93	100

^a Age bias defined as a weighted mean that is > 2s greater than or less than the reference age.

Temora2 has 0.8% and 0% (up to 1.5%); Mud Tank has 1.0% and 0.7% and 91500 has 0.8% and 0% (up to 1%). These results highlight that within-session repeatabilities of RMs are typically excellent at the measurement precisions achieved (i.e., MSWD ~ 1), requiring little or no propagation for excess variance ($\epsilon \sim 0\%$). This is reflected in $^{207}\text{Pb}/^{206}\text{Pb}$ ratios as well: Plešovice has typical uncertainties of 0.9% (1s) and $\epsilon = 0\%$; Temora2 has 1.4% and 0%; Mud Tank has 1.8% and 0% and 91500 has 1.3% and 0%. With higher levels of measurement precision, within-session excess variance (ϵ) is expected to become more significant.

Despite the fact that within-session excess variance (ϵ) is typically 0, long-term excess variance (ϵ') is typically higher, due to the variability in the weighted mean $^{206}\text{Pb}/^{238}\text{U}$ or $^{207}\text{Pb}/^{206}\text{Pb}$ ratio between sessions. For $^{206}\text{Pb}/^{238}\text{U}$, these values (1s) are: AusZ7-1 = 0.84%, AusZ7-5 = 0%, OD-3 = 0.28%, Temora2 = 0.58%, Plešovice = 0.54% and 91500 = 0.49%. For $^{207}\text{Pb}/^{206}\text{Pb}$ ages, smaller excess variances are observed: Temora2 = 0.17%, Plešovice =

0.21%, Mud Tank = 0.87% and 91500 = 0.20%. While a detailed discussion of the origins of such systematic uncertainty is beyond the scope of this study, it is possible that it relates strongly to matrix effects (e.g., Allen and Campbell 2012) and daily variations in analytical parameters.

The value of these findings is threefold. First, they present a large database of reference material analyses and a methodical calculation of a value (ϵ') that may often be quoted as a certain percentage (e.g., Gutiérrez *et al.* 2018) but is not rigorously quantified. Second, it provides a methodology for quantifying $^{206}\text{Pb}/^{238}\text{U}$ or $^{207}\text{Pb}/^{206}\text{Pb}$ ϵ' values from years-long databases in a way that could be emulated by other LA-ICP-MS laboratories in the future. Third, in laboratories where long-term databases are not present but similar analytical instrumentation is used, this study provides a first-order set of estimates for ϵ' . In particular, the variability in $^{206}\text{Pb}/^{238}\text{U}$ or $^{207}\text{Pb}/^{206}\text{Pb}$ ϵ' values for different reference materials provides the basis to argue for a

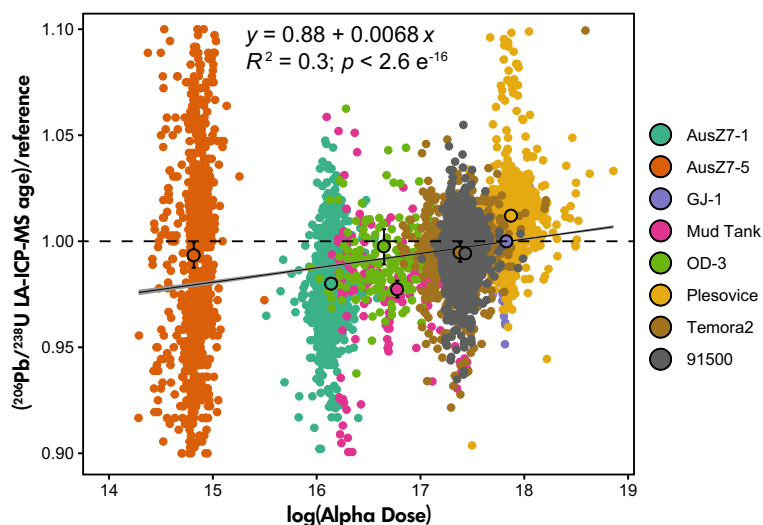


Figure 2. Age bias (measured age/reference age) versus log(alpha dose) for eight non-thermally annealed reference materials, demonstrating a tendency of low-radiation RMs to record anomalously young ages. Range bars denote 2SE. Note that even data were corrected for alpha dose, some age biases remain.

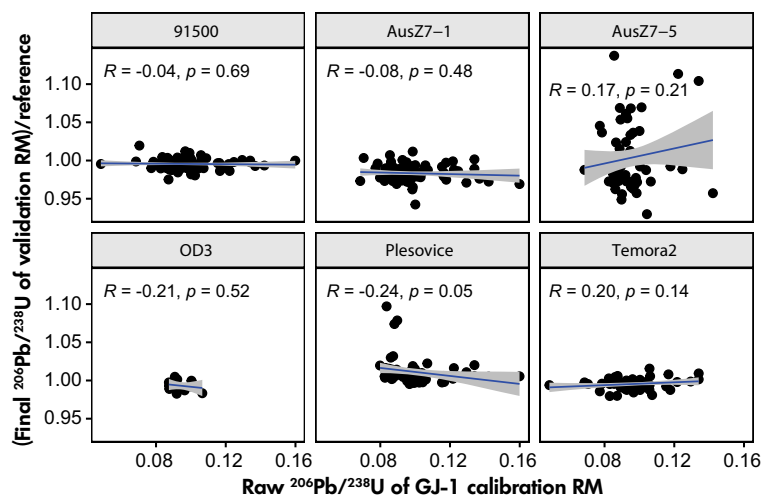


Figure 3. Raw $^{206}\text{Pb}/^{238}\text{U}$ ratios of GJ-1 versus normalised $^{206}\text{Pb}/^{238}\text{U}$ final ratios of six reference materials (Mud Tank omitted due to small sample size). Each point represents the average over one measurement session. Note that the final $^{206}\text{Pb}/^{238}\text{U}$ ratio of VRMs is not correlated with the raw ratio of the calibration RM. Blue lines are least square regression lines and grey bands represent 95% confidence intervals.

particular ϵ' value. $^{207}\text{Pb}/^{206}\text{Pb}$ ϵ' is typically $\sim 0.4\%$ (2s) for zircons older than ~ 300 Ma. $^{206}\text{Pb}/^{238}\text{U}$ ϵ' is typically $\sim 1\%$ (2s) for similarly aged zircons, but likely higher for Cenozoic zircons (e.g., AusZ7-1), as discussed later. Using ϵ' estimates as rough guides, one may, for example, attempt to distinguish suites of zircons of similar age using a modified estimate of total uncertainty. Two hypothetical zircon suites with weighted means of 400 and 410 Ma each have an uncertainty of 0.5% (2s), amounting to ~ 2 Ma. If the two

suites were analysed in the same session, an argument could be made that they are distinct populations because they are distinct at the 95% confidence level. If, however, they were from different sessions, then one would have to assume some degree of excess variance. An estimate of 1.2% (2s, taken from Temora2) propagated into the weighted mean uncertainty results in a total uncertainty > 5 Ma, which places the distinctiveness of the two suites into question.

Data mining exercises such as these may be useful in determining the source of systematic uncertainties as well. In this instance, the non-zero ϵ' values are the result of variations in weighted mean averages of reference materials. Given that all final $^{206}\text{Pb}/^{238}\text{U}$ are attained after normalising to the calibration RM, one question that arises is whether the age offset in VRMs is the result of an imperfect normalisation to the calibration RM. In other words, is the offset of the final $^{206}\text{Pb}/^{238}\text{U}$ ratio correlated with the raw $^{206}\text{Pb}/^{238}\text{U}$ ratio of the calibration RM? Data aggregation suggests this is not the case (Figure 3). While there are age biases present during many measurement sessions [i.e., $(^{206}\text{Pb}/^{238}\text{U})/(\text{reference})$ not equal to 1], these biases are not significantly correlated with the raw $^{206}\text{Pb}/^{238}\text{U}$ ratio of the GJ-1 calibration RM in those sessions.

However, these data do demonstrate that reference materials tend to show an age bias in each measurement session, which in almost all cases is negative (Table 3). In particular, all RMs aside from AusZ7-5 demonstrate age bias (reference age outside of 2s uncertainty bounds of mean age) in at least 50% of measurement sessions, with positive biases dominating in Plešovice and negative biases dominating in the rest. After propagation of ϵ' , the proportion of biased weighted means drops from 74% to 47% for AusZ7-1, from 92% to 33% for Mud Tank, 58% to 26% for Plešovice, 59% to 16% for Temora2 and 51% to 14% for 91500. The proportions of biased weighted means of AusZ7-5 (5%) and OD-3 (58%) remain unchanged after ϵ' propagation.

The predominantly negative age bias most likely results from the properties of the calibration RM. Here, the high U content and age of GJ-1 result in heavy radiation damage relative to other RMs, inducing a higher degree of down-hole LIEF that causes over-correction of the VRM analyses and generates anomalously young ages in the VRMs (Allen and Campbell 2012). This effect is lessened, but not eliminated by thermal annealing/chemical abrasion (Sliwinski *et al.* 2017). Also, the extent of age bias does not seem to be correlated with only the alpha radiation dosage, and likely results from a number of other factors including crystal optical parameters and trace element content (Marillo-Sialer *et al.* 2016), which becomes apparent when examining the persistent positive age bias in Plešovice (with otherwise very similar alpha dosage to the calibration RM, Figure 2). The effects of alpha dosage are quite apparent in young RMs (e.g., AusZ7-1), where the greatest proportion of weighted means demonstrate negative age bias (92%), and where the ϵ' is higher than with older samples (0.84% 1s). The latter is likely due to the variability in the magnitude of age bias between sessions. That is, while the low alpha dosage

almost always leads to negative age bias, the magnitude of this bias changes from day to day and introduces increased scatter between measurement sessions (Figure 2, Sliwinski *et al.* 2017). Very young zircons such as AusZ7-5 demonstrate some negative bias, but their long-term excess uncertainty seems to be masked by the high analytical uncertainty resulting from their low Pb mass fractions.

It should be noted that variations in material properties are not solely responsible for LIEF and resulting age biases. Many analytical parameters change between sessions (e.g., plasma temperature, torch position), although these parameters are usually assumed to affect calibration RMs and VRMs equally. One notable exception is the location of the RMs in the laser ablation cell and the presence of gas impurities (e.g., O_2 , N_2) introduced during sample exchange. Thompson *et al.* (2018) demonstrated pronounced age offsets due to atmospheric air in improperly flushed cells, as well as the alleviation of age biases by placing samples under sustained vacuum prior to analysis. This suggests that gas/water adsorbed onto the surface of a sample may interfere with the ablation behaviour and therefore the age determination of a sample. While further exploration of laser-sample interaction and gas flow are beyond the scope of this study, it is clear that maintaining consistency in as many analytical parameters as possible is essential to generating consistent age determinations in RMs, minimising age biases and limiting ϵ' . This includes limiting gas contamination by properly flushing the cell (Thompson *et al.* 2018), running the instrument for a set amount of time before analysis, as well as maintaining consistency in laser ablation and ICP-MS parameters.

Conclusions

The present study provides a comprehensive, long-term assessment of uncertainties (ϵ and ϵ' , or intra- and inter-session excess uncertainty) using a suite of commonly used reference materials. While intra-session excess uncertainty is generally low (reference material populations show MSWDs close to 1), inter-session excess uncertainty may vary. Typically, this effect requires the addition of < 2% excess uncertainty (2s), but illustrates the importance of characterising long-term excess variance (ϵ') by LA-ICP-MS U-Pb laboratories and its influence in determining total age uncertainty for a sample. This excess variance can be added in quadrature to the weighted mean uncertainties, and the resulting total uncertainty can be used to compare different populations of zircons for age overlap. Used in this fashion, these long-term data sets can resolve systematic biases between RMs, improve our understanding of their origin and enable the definition of practices to reduce or eliminate them.

Acknowledgements

This study was supported by ETH Research Grant ETH-34 15-2. We thank Pieter Vermeesch for extremely helpful and timely input regarding the statistical treatment of large data sets, particularly while using IsoPlotR. Open access funding provided by Eidgenössische Technische Hochschule Zurich.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its online supplementary materials.

References

- Allen C.M. and Campbell I.H. (2012) Identification and elimination of a matrix-induced systematic error in LA-ICP-MS $^{206}\text{Pb}/^{238}\text{U}$ dating of zircon. *Chemical Geology*, 332–333, 157–165.
- Black L. and Gulson B. (1978) The age of the Mud Tank carbonatite, Strangways range, Northern Territory. *BMR Journal of Australian Geology and Geophysics*, 3, 227–232.
- 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. and Foudoulis C. (2004) Improved $^{206}\text{Pb}/^{238}\text{U}$ 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.
- Guillong M., Sliwinski J.T., Schmitt A., Forni F. and Bachmann O. (2016) U-Th zircon dating by laser ablation single collector inductively coupled plasma-mass spectrometry (LA-ICP-MS). *Geostandards and Geoanalytical Research*, 40, 377–387.
- Gutiérrez F., Payacán I., Szymanowski D., Guillong M., Bachmann O. and Parada M. (2018) Lateral magma propagation during the emplacement of La Gloria pluton, Central Chile. *Geology*, 46, 1051–1054.
- Horstwood M.S.A., Košler J., Gehrels G., Jackson S.E., McLean N.M., Paton C., Pearson N.J., Sircombe K., Sylvester P., Vermeesch P., Bowring J.F., Condon D.J. and Schoene B. (2016) 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.
- Iwano H., Orihashi Y., Hirata T., Ogasawara M., Danhara T., Horie K., Hasebe N., Sueoka S., Tamura A., Hayasaka Y., Katsube A., Ito H., Tani K., Kimura J.-I., Chang Q., Kouchi Y., Haruta Y. and Yamamoto K. (2013) An inter-laboratory evaluation of OD-3 zircon for use as a secondary U-Pb dating standard. *Island Arc*, 22, 382–394.
- Jackson S.E., Pearson N.J., Griffin W.L. and Belousova E.A. (2004) The application of laser ablation-inductively coupled plasma-mass spectrometry to *in situ* U-Pb zircon geochronology. *Chemical Geology*, 211, 47–69.
- Kennedy A.K., Wotzlaw J.-F., Schaltegger U., Crowley J.L. and Schmitz M. (2014) Eocene zircon reference material for microanalysis of U-Th-Pb isotopes and trace elements. *The Canadian Mineralogist*, 52, 409–421.
- Klötzli U., Klötzli E., Günes Z. and Košler J. (2009) Accuracy of laser ablation U-Pb zircon dating: Results from a test using five different reference zircons. *Geostandards and Geoanalytical Research*, 33, 5–15.
- Marillo-Sialer E., Woodhead J., Hanchar J., Reddy S., Greig A., Hergt J. and Kohn B.J. (2016) An investigation of the laser-induced zircon ‘matrix effect’. *Chemical Geology*, 438, 11–24.
- Marillo-Sialer E., Woodhead J., Hergt J., Greig A., Guillong M., Gleadow A., Evans N. and Paton C. (2014) The zircon ‘matrix effect’: Evidence for an ablation rate control on the accuracy of U-Pb age determinations by LA-ICP-MS. *Journal of Analytical Atomic Spectrometry*, 29, 981–989.
- Paton C., Hellstrom J., Paul B., Woodhead J. and Hergt J. (2011) lolite: Freeware for the visualisation and processing of mass spectrometric data. *Journal of Analytical Atomic Spectrometry*, 26, 2508–2518.
- Paton C., Woodhead J. D., Hellstrom J. C., Hergt J. M., Greig A. and Maas R. (2010) Improved laser ablation U-Pb zircon geochronology through robust downhole fractionation correction. *Geochemistry Geophysics Geosystems*, 11.
- Petrus J. A. and Kamber B. S. (2012) VizualAge: A novel approach to laser ablation ICP-MS U-Pb geochronology data reduction. *Geostandards and Geoanalytical Research*, 36, 247–270.
- Sláma J., Košler J., Condon D.J., Crowley J.L., Gerdes A., Hanchar J.M., Horstwood M.S.A., Morris G.A., Nasdala L., Norberg N., Schaltegger U., Schoene B., Tubrett M.N. and Whitehouse M.J. (2008) Plešovice zircon – A new natural reference material for U-Pb and Hf isotopic microanalysis. *Chemical Geology*, 249, 1–35.
- Sliwinski J.T., Guillong M., Liebske C., Dunkl I., von Quadt A. and Bachmann O. (2017) Improved accuracy of LA-ICP-MS U-Pb ages of Cenozoic zircons by alpha dose correction. *Chemical Geology*, 472, 8–21.
- Solari L.A., Ortega-Obregón C. and Bernal J.P. (2015) U-Pb zircon geochronology by LA-ICP-MS combined with thermal annealing: Achievements in precision and accuracy on dating standard and unknown samples. *Chemical Geology*, 414, 109–123.
- Thompson J.M., Meffre S. and Danyushevsky L. (2018) Impact of air, laser pulse width and fluence on U-Pb dating of zircons by LA-ICP-MS. *Journal of Analytical Atomic Spectrometry*, 33, 221–230.

Vermeesch P. (2018)

IsoplotR: A free and open toolbox for geochronology. *Geoscience Frontiers*, 9, 1479–1493.

von Quadt A., Wotzlaw J.-F., Buret Y., Large S.J.E., Peytcheva I. and Trinquier A. (2016)

High-precision zircon U/Pb geochronology by ID-TIMS using new 10^{13} ohm resistors. *Journal of Analytical Atomic Spectrometry*, 31, 658–665.

Wiedenbeck M., Allé P., Corfu F., Griffin W., Meier M., Oberli F., von Quadt A., Roddick J.C. and Spiegel W. (1995)

Three natural zircon standards for U-Th-Pb, Lu-Hf, trace element and REE analyses. *Geostandards Newsletter*, 19, 1–23.

Supporting information

The following supporting information may be found in the online version of this article:

Figure S1. All reference material zircon U-Pb data ($n = 14122$).

Figure S2. Analysed and filtered data of Table S1.

Table S1. Extended version of Table 2.

Table S2. Long-term $^{206}\text{Pb}/^{238}\text{U}$ ratios for 91500, Temora2, OD-3, Plešovice, AusZ7-1 and AusZ7-5.

Table S3. Summary of typical uncertainties on $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ ratios for each reference material within each measurement session (from Table S2).

This material is available from: <http://onlinelibrary.wiley.com/doi/10.1111/ggr.12442/abstract> (This link will take you to the article abstract).