Science Bulletin 64 (2019) 189–197

Contents lists available at ScienceDirect

Science Bulletin

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

Science Bulletin
www.scibull.com

Article

Monte Carlo sampling for error propagation in linear regression and applications in isochron geochronology

Yang Li [a,b,c,*], Shuang Zhang [c], Richard Hobbs [b], Camila Caiado [d], Adam D. Sproson [b,e], David Selby [b], Alan D. Rooney [c]

[a] State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] Department of Earth Sciences, Durham University, Durham DH1 3LE, UK
[c] Department of Geology and Geophysics, Yale University, New Haven, CT 06511, USA
[d] Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK
[e] Atmosphere and Ocean Research Institute, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa 277-8564, Japan

ARTICLE INFO

Article history:
Received 24 September 2018
Received in revised form 27 November 2018
Accepted 28 November 2018
Available online 22 December 2018

Keywords:
Linear regression
Isochron
Geochronology
Uncertainty propagation
Monte Carlo
Isoplot

ABSTRACT

Geochronology is essential for understanding Earth's history. The availability of precise and accurate isotopic data is increasing; hence it is crucial to develop transparent and accessible data reduction techniques and tools to transform raw mass spectrometry data into robust chronological data. Here we present a Monte Carlo sampling approach to fully propagate uncertainties from linear regressions for isochron dating. Our new approach makes no prior assumption about the causes of variability in the derived chronological results and propagates uncertainties from both experimental measurements (analytical uncertainties) and underlying assumptions (model uncertainties) into the final age determination. Using synthetic examples, we find that although the estimates of the slope and *y*-intercept (hence age and initial isotopic ratios) are comparable between the Monte Carlo method and the benchmark "Isoplot" algorithm, uncertainties from the later could be underestimated by up to 60%, which are likely due to an incomplete propagation of model uncertainties. An additional advantage of the new method is its ability to integrate with geological information to yield refined chronological constraints. The new method presented here is specifically designed to fully propagate errors in geochronological applications involves linear regressions such as Rb-Sr, Sm-Nd, Re-Os, Pt-Os, Lu-Hf, U-Pb (with discordant points), Pb-Pb and Ar-Ar.

1. Introduction

Geochronology is an essential aspect of Earth sciences, and advances in this field have resulted in many breakthroughs in understanding the history of our solar system and the evolution of life on Earth [1]. In general, extracting geologically meaningful ages from rocks and minerals starts with sample collection, followed by sample processing, and isotopic ratio measurements via mass spectrometry. The raw isotopic ratios generated by mass spectrometers then need to be transformed into atomic ratios, and eventually into chronological dates with propagation of associated uncertainties (e.g., [2,3]). Over the past three decades, a great number of analytical innovations and instrumentation advances have emerged, which gave rise to unprecedented levels of accuracy and precision for isotopic ratio measurements as well as pioneering new radiometric systems for questions ranging from early solar system evolution to Anthropocene climate change. Advances in the precision and accuracy as well as the expansion of available geochronometers have been facilitated by a combination (often iteratively) of better analytical approaches and robust, transparent and accessible data reduction tools (e.g., [4–13]). To more fully harness these technical improvements, it is critical to concomitantly develop data reduction techniques and appropriate visualization methods. Although there have been significant progresses made in data reduction techniques for U-Th-Pb and Ar-Ar systems [3,6,7,14–20], fewer advances have been seen in isochron dating, a method utilized for systems including Rb-Sr, Sm-Nd, Re-Os, Pt-Os, Lu-Hf, U-Pb (with discordant points), Pb-Pb and Ar-Ar.

Isochron dating is based on linear regression in which one determines the slope, *y*-intercept and associated uncertainties of the best fitting line to the parent and daughter isotopic ratios (including their uncertainties and error correlations). The fundamental assumptions behind isochron dating include: (1) all

* Corresponding author.
E-mail addresses: geoliy@outlook.com, geoliy@mail.iggcas.ac.cn (Y. Li).

https://doi.org/10.1016/j.scib.2018.12.019
2095-9273/

co-genetic samples have near-identical initial daughter isotopic compositions; (2) samples begin accumulating daughter isotopes via radiogenic decay at the same time; (3) these samples remain closed in terms of both parent and daughter isotopes following the accumulation of the daughter isotope. A further requirement is that these samples should have variable parent isotope (or daughter isotope) ratios to define a line. This linear regression is routinely carried out by the "Isoplot" program that is based on a Microsoft Excel macro [2,21] and includes York's algorithm [22–24]. This algorithm performs a least-squares fit to data with normally distributed but correlated uncertainties, and assumes that the data points lie along a straight line (isochron) and offsets from this line are due to imperfect measurements, otherwise known as analytical uncertainties. In reality however, the data points might not fall on a straight line even if they could be measured perfectly because of differences in initial isotopic composition, varying ages and/or open system behavior, which we will refer as model uncertainties. To address this, the "Isoplot" program uses two different techniques (additional options are discussed below) for error propagation and decides which one to use based on the probability of how well the data "fits" to the line. If the probability of fit is satisfactory, "Isoplot" assumes that analytical uncertainty is the only cause of scatter and uses York's algorithm to propagate only analytical uncertainties to produce a so-called Model 1 age. If the fit of the data to a common line is not satisfactory resulting in a violation of York's assumption (i.e., in the case of over-dispersion), "Isoplot" uses an adapted regression that accounts for an unknown but normally distributed variation in the initial isotopic ratios of the samples [2,25], producing a Model 3 age. Though the users can choose the cutoff value between the two Models (between 0.05 and 0.3 with a default of 0.15), in the absence of additional geologic constraints, there is no standard criteria to choose this cutoff value, which can lead to inconsistencies in chronological results if this value is not properly documented.

"Isoplot" also offers a Model 2 solution in which case equal weights and zero error correlations are assigned to the samples, as opposed to those used in Model 1 and Model 3 where each sample has a weighting proportional to the inverse square of its analytical uncertainties (also accounts the error correlation). When the assumption that residuals (observed scatter) of the data-points from a straight line have a normal (Gaussian) distribution is invalid, "Isoplot" has an option called "Robust regression" which makes no assumptions about the cause(s) of the observed scatter of the data from a straight line. We do not discuss these two options further as they are rarely used and beyond the scope of this study.

As pointed out by Ludwig [26], uncertainty determined by Monte Carlo sampling is the most reliable approach, therefore in this paper we propose an method to determine the slope, *y*-intercept and their uncertainties, based on Monte Carlo sampling and simple linear regression. Unlike the Monte Carlo method in York et al. [24], the proposed method here propagates not only analytical uncertainties, but also uncertainties arising from the underlying assumptions (model uncertainties). This approach differs from Model 1 and Model 3 solutions from Isoplot as our new method propagates uncertainties in a consistent manner regardless of the probability of fit and hence avoids subjective choosing of the cut-off value discussed above. Our method can be applied to data with any goodness of fit and distinguishes between analytical and model uncertainties. This paper discusses three key aspects: (1) the Monte Carlo based method; (2) the examination of differences and similarities to Isoplot; and (3) the use of a synthetic dataset to demonstrate the potential to integrate independent geological information for refined chronologic constraints.

2. Monte Carlo simulation

2.1. Experimental data and their uncertainties

The parent and daughter isotopic ratios (x, y) of a sample are measured experimentally, with their uncertainties (δx, δy) inherited from the analytical procedure. Additionally, the uncertainties of the parent and daughter isotopic ratios are typically correlated due to the utilization of a common isotope for converting absolute atomic numbers into isotopic ratios (e.g., ^{86}Sr in $^{87}Rb/^{86}Sr$ and $^{87}Sr/^{86}Sr$; ^{144}Nd in $^{147}Sm/^{144}Nd$ and $^{143}Nd/^{144}Nd$), which is quantified by a correlation coefficient denoted by ρ or *rho* [27]. Experimental data with the same parent and daughter isotopic ratios and uncertainties, but variable error correlations are graphically illustrated in Fig. 1a as error ellipses at the 2-sigma level (all uncertainties are presented at the 2-sigma level in absolute values unless otherwise stated). By definition, a high error correlation indicates that the sources of δx and δy are predominately from one contributor, which for isotope geochemistry is likely to be caused by analytical uncertainty of the stable isotope used to convert absolute atomic numbers into isotopic ratios. As emphasized by Ludwig [26] and illustrated in Fig. 1a, the 2-sigma error ellipses including error correlation extend farther than the 2-sigma range of δx and δy, which is a non-intuitive characteristic of joint distributions. As such, excluding error correlations for linear regressions will yield an incorrect uncertainty for the slope and its uncertainty [28]. Hence it is critical to report and use accurate error correlations for the experimental data in all geochronological studies which can be estimated through differentiation and observation [2]. The analytical uncertainties with error correlation can also be presented as probability density functions (PDFs, Fig. 1b). This probability density function is the basis for the resampling process used in our Monte Carlo method.

2.2. Propagation of analytical uncertainties

We demonstrate the principles of our Monte Carlo based technique using a synthetic example consisting of five samples. The parent and daughter isotopic ratios and associated uncertainties including error correlations of the five samples are graphically illustrated in Fig. 2a as error ellipses. To propagate analytical uncertainties, we perform the following steps:

(1) For each of the five samples, we randomly select a coordinate from its corresponding probability density function as that defined in Fig. 1b. Each sampled coordinate is considered to be a pair of absolute values without uncertainty (Fig. 2a);
(2) Once a coordinate has been selected for each of the five samples, the parameters (slope and *y*-intercept) of the regression line are determined (Fig. 2a) following a least-square estimation [29]. The slope and *y*-intercept of this regression line is plotted in Fig. 2b;
(3) Repeating steps 1 and 2 yields a distribution representing the probability of slope and *y*-intercept of the five samples. By increasing the iteration times (Fig. 2c, e), the shape of the resulting probability distribution becomes apparent (Fig. 2d, f). We acknowledge here that more iterations will yield a more accurate distribution, but will also increase computing time. A discussion on how to balance the iteration time and computing resource is presented in Section 2.4 below.

This approach only propagates analytical uncertainties but not uncertainties from the linear regression itself. This is illustrated

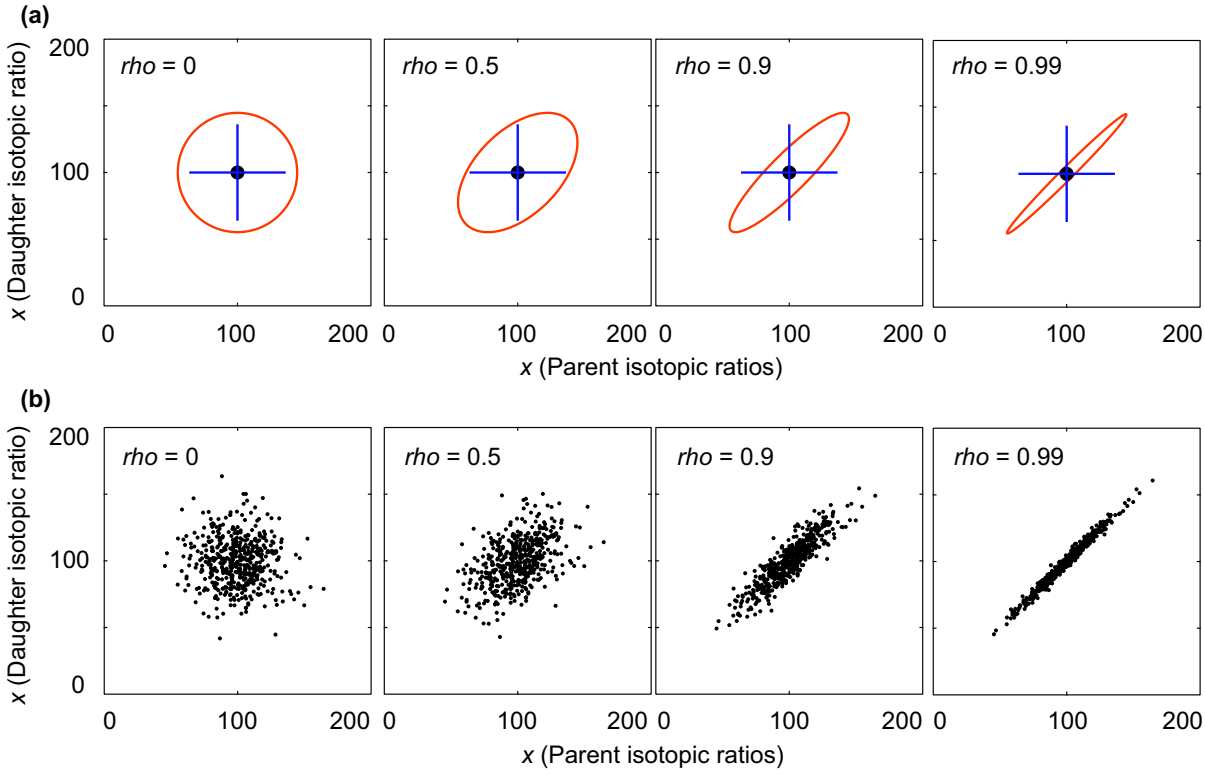


Fig. 1. Data and uncertainties with associated error correlations (*rho*) are presented as error ellipse and error bar (a) as well as probability density function (b). The plots show the same data and their uncertainties (cross hairs) and only vary in their correlation (values are indicated on each plot). All uncertainties are presented at the 2-sigma level (95.45% confidence).

by the example in Fig. 3. For a dataset consisting of five samples that have no analytical uncertainty and do not plot on a common line (Fig. 3a), using the above algorithm will result in no uncertainty for the slope and *y*-intercept (Fig. 3b), which is not a plausible result because the fitted line does not pass through all the five samples. We term these non-analytical uncertainties as the model uncertainty. The primary contributors of this model uncertainty include differences in the initial isotope composition, ages, or those which arise from open isotopic system behavior violating the fundamental assumptions behind isochron dating. In realistic scenarios it is likely that both analytical and model uncertainties will be present at some level though careful selection of samples and refined measurements maybe used to minimize their effect. Using the simple Monte Carlo algorithm described above which only propagates analytical uncertainties and fails to capture this extra source of uncertainty. We therefore propose an extension of our method to account for this as described below.

2.3. Propagation of model uncertainties

Uncertainties for the slope and *y*-intercept of the regression line in each sampling step in Section 2.2 (Fig. 2) are calculated as standard errors following that of James et al. [29]. Further, these uncertainties are correlated as defined by the correlation coefficient (*C*):

$$C = -\sum_{i=1}^{n}(x_i)/\left(n \times \left(\sum_{i=1}^{n}\left(x_i^2\right)\right)^{0.5}\right), \quad (1)$$

where n is the number of samples (e.g., 5 for the example in Fig. 2), and x_i denotes the sampled point's *x*-axis. Knowing these uncertainties and error correlation for each sampling step, it is possible to include them by replacing the outcome of each sampling step by a new probability density distribution. This process is illustrated in Fig. 3, where one of the outcomes from the sampling step (Fig. 3b) is replaced by a new probability density distribution (Fig. 3d). For input data with analytical uncertainties (Fig. 3e), when model uncertainties are included for all simulations, a final distribution (blue in Fig. 3f) is obtained. This final distribution includes both analytical and model uncertainties, and we term them as total uncertainties. In the presence of both analytical uncertainties and model uncertainties, we cannot determine exactly whether the scatter in the final distribution is inherited from analytical uncertainties or caused by model uncertainties, or a combination of both.

Statistical analysis is applied to the final distribution to quantify the uncertainties for data interpretation. We use the means and two standard deviations of the slope and *y*-intercept, plus the correlation between them, to assess the significance of this final distribution. Additionally, the contribution of analytical uncertainties to the total uncertainties (analytical + model uncertainties) could be assessed. Here we emphasis that as discussed above, analytical uncertainties could be an additional source of model uncertainties, hence the contribution only can be assessed semi-quantitatively.

The advantage of this method is that regardless of the degree of fit, both analytical and model uncertainties are propagated into the final distribution. In other words, the degree of fit is not a prerequisite to alter the strategy of error propagation. As such, the proposed method ensures that quoted uncertainties can be fairly compared as they are calculated in a consistent manner.

2.4. The iteration times

To achieve a representative final distribution for the given sample set, a high number of iterations are required at the expense of consuming more computing resources and time. In this regard,

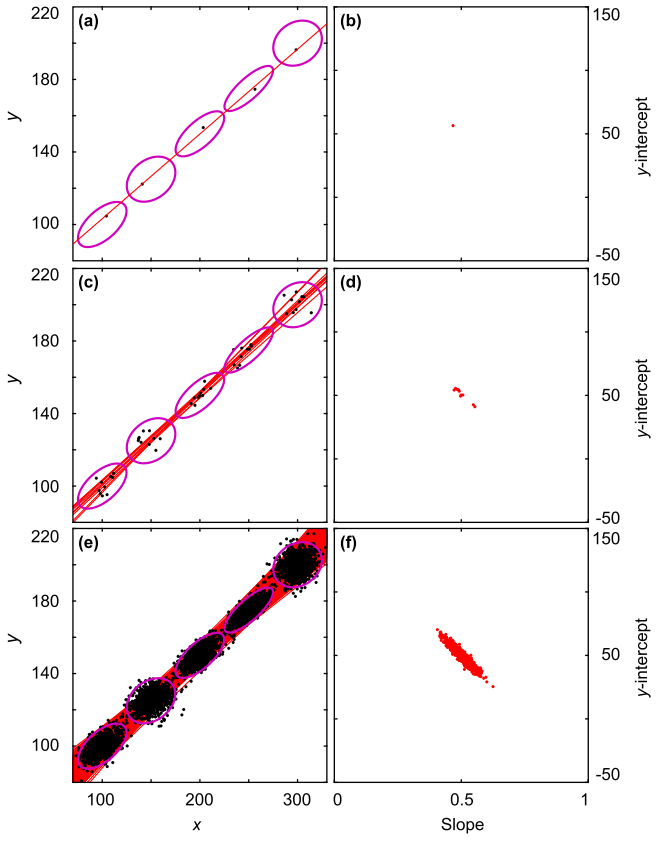


Fig. 2. The principle of the Monte Carlo based simulation is illustrated by an example comprising five samples. (a) Randomly sampling a data point from the PDFs of each of the five samples and estimating its slope and *y*-intercept using a simple least-squares method. The slope and *y*-intercept from (a) are plotted in (b). (c, d) 10 and (e, f) 1000 iterations of the procedure described in (a) and (b). The accuracy of the final distribution (f) improves with increasing iterations/sampling.

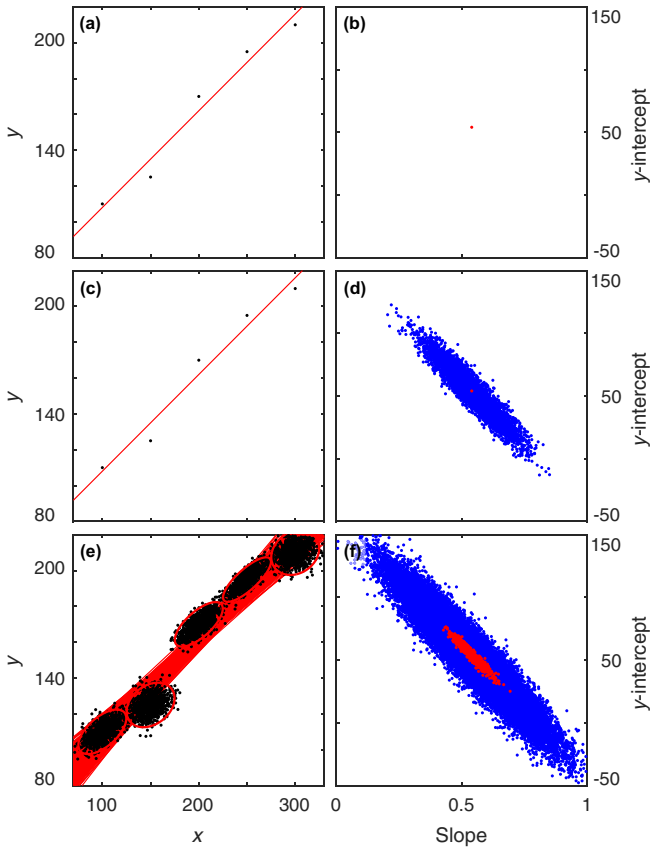


Fig. 3. The presence of model uncertainties. As illustrated by a synthetic example comprising five samples not plotting on a line, assuming no analytical uncertainties (a), sampling according to their PDFs will yield a distribution without uncertainties (b) although in fact it has uncertainty. This indicates the presence of non-analytical uncertainties, which are defined as model uncertainties and need to be accounted for. Using the same samples without analytical uncertainties (b), the model uncertainty has been illustrated by a new distribution in blue (c, d). A more realistic data set, in which data have analytical uncertainties (e), model uncertainties have been added to all resampled regressions, a final distribution (blue points) is obtained (f) which includes both analytical and model uncertainties.

the iteration times should be balanced between the accuracy of the final distribution and the simulation time. Here we monitor the mean and standard deviation of the final distribution and stop iteration once this mean and standard deviation are stabilized. Our preliminary experiment suggests that an iteration count of about 10^6 is sufficient in most cases, and could be increased when necessary.

3. Comparison with Isoplot

It is important to compare the results from the Monte Carlo based approach with those from the Isoplot program to understand differences in the assumptions and how they propagate into the resultant age estimations. In the following section, we construct a synthetic experimental data set to highlight the magnitude of these differences and explore implications in isochron dating.

3.1. Synthetic experimental dataset

Using the Re-Os isotopic system as an example, where ^{187}Re decays to ^{187}Os with a decay constant of 1.666×10^{-11} a^{-1} [30,31], we generate synthetic examples for the experiment (Table 1). To be representative of geological scenarios, the examples are designed to cover plausible scenarios in isochron dating, as represented by the probability of fit which varies between 0 and 1 (Fig. 4). For uncertainty propagation using the Isoplot program, we follow the default approach in the Isoplot program to set the cut-off value as 0.15. As can be seen from the following discussion, using different cut-off values should not bias our conclusion. Below we outline the approaches generating these examples.

(1) An age and an initial daughter isotopic ratio (i.e., $^{187}Os/^{188}Os_{initial}$) are randomly assigned between 100 and 4,500 Ma and 0.2–1.2, respectively, following uniform distributions.
(2) The number of samples, *n*, used to construct an isochron is randomly chosen between 5 and 30 following a uniform distribution.
(3) For the *n* samples, their parent isotopic ratios (i.e., $^{187}Re/^{188}Os$) at present day are randomly selected following uniform distributions between 100 and 1,000. Specifically, for each example, we first randomly pick a lowest ratio and a highest ratio which lie between 100 and 1,000. Afterwards, we randomly pick *n*-2 ratios following a uniform distribution between that lowest ratio and highest ratio. The purpose of this specific approach is to guarantee that for these examples, the variety of the parent isotopic ratios (spread of the isochron) in each example follows a uniform distribution.
(4) The daughter isotopic ratios (e.g., $^{187}Os/^{188}Os$) at present day of the *n* samples are calculated individually following Eq. (2) using the *t*, initial daughter isotopic ratio and parent isotope ratios generated in step 1, 2 and 3, respectively.

$$^{187}Os/^{188}Os = {}^{187}Os/^{188}Os_{initial} + {}^{187}Re/^{188}Os * \left(e^{\lambda t} - 1\right). \quad (2)$$

Table 1
Parameters for the synthetic dataset.

Age (Ma)	*n*	Initial	*x*	d*x*	*y*	d*y*	Scatter	*rho*
100–4,500 uniform	5–30 uniform	0.2–1.2 uniform	100–1,000 uniform	0.2%–1% uniform	Equation (2)	0.2%–1% uniform	0.2%–1.2% uniform	0.4–0.999 uniform

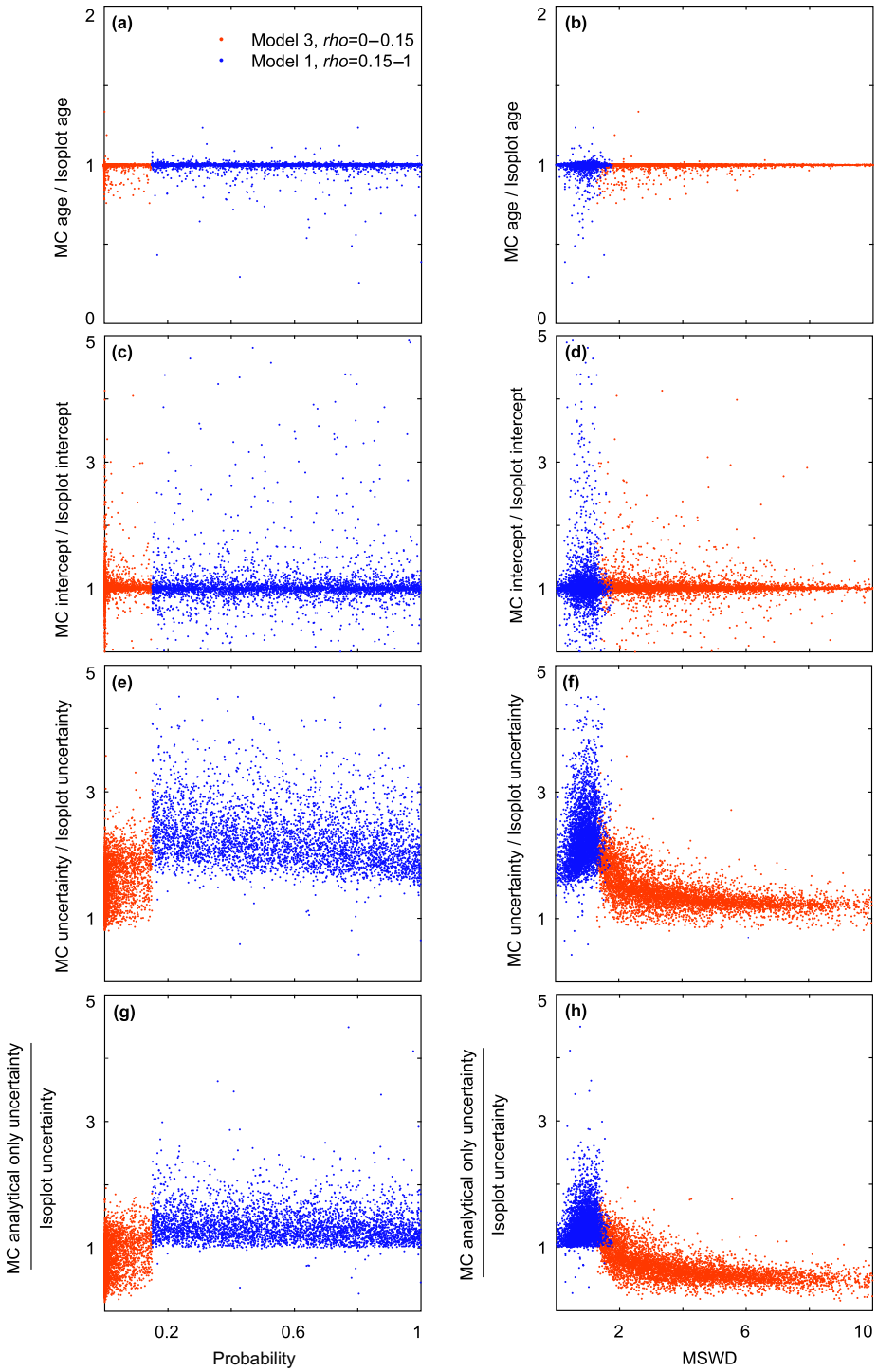


Fig. 4. Comparing results from Isoplot and Monte Carlo methods using synthetic examples. Note the relationship of uncertainties between Monte Carlo method and Isoplot program has an abrupt change at $p = 0.15$, likely due to the contrasting strategies of error propagation in Model 1 and Model 3 solutions used in Isoplot. Comparison of the slope estimate as a function of the probability of fit (a) and MSWD (b) and *y*-intercept estimate as a function of the probability of fit (c) and MSWD (d). The slope and *y*-intercept estimates, hence age and initial isotopic ratio estimates, from the two methods are comparable. In cases when the analytical and model uncertainties are taking into account (e, f), the uncertainties of the slopes and *y*-intercepts from the Monte Carlo based simulation are larger than those from the Isoplot program. When only the analytical uncertainties are considered (g, h), the Isoplot Model 1 age uncertainty is comparable but slightly larger than the Monte Carlo based approach. See text for a detailed discussion.

(5) We then introduce scatter to the daughter isotopic ratios by adding or subtracting a value ranging from 0.2% to 1.2% of the corresponding daughter isotopic ratios following uniform distributions, and the decision whether to add or subtract is also random. Note that this scatter serves to imitate model uncertainties. The model uncertainties are introduced through modifying the daughter isotopic ratios, which cover all the potential causes of model uncertainties including variations in initial isotopic composition and age, as well as open system behaviour to the isotopic system and imperfect measurements.
(6) The 2-sigma relative uncertainty (i.e., percentage uncertainty) of the parent and daughter isotope ratios are randomly assigned between 0.2% and 1% following uniform distributions, with their error correlations randomly given between 0.4 and 0.999, which also follow uniform distributions.

The data generated above are processed by our new Monte Carlo method as well as the Isoplot program. Therefore, one age and one initial isotopic composition plus their associated uncertainties (2-sigma) will be obtained from the Isoplot program either following Model 1 ($p > 0.15$) or Model 3 ($0 < p < 0.15$) solutions. For the Monte Carlo simulation, one age, one initial isotopic ratio, and associated total uncertainties (analytical uncertainties + model uncertainties) are obtained for each example. We perform this process 10,000 times, and as expected, the probability of these examples varies between 0 and 1 with the corresponding Mean Square Weighted Deviation (MSWD) ranging from >10 to 0.

3.2. Results from Monte Carlo method and the Isoplot program

Regardless of which linear regression tool is employed, the slopes and *y*-intercepts, hence ages and initial isotopic ratios, are the same (Fig. 4a–d). Minimal scatter exists when the spread in the synthetic data points is limited, which renders an accurate age estimation difficult. Notably, uncertainties obtained from the Monte Carlo simulation are consistently larger than those from the Isoplot program (Fig. 4e, f). Here we use the $R_{MC/Iso}$ to illustrate these results, where $R_{MC/Iso}$ equals to age uncertainties (total) from the Monte Carlo method divided by age uncertainties from the Isoplot program. When p decreases from 1 to 0.15, the running mean of $R_{MC/Iso}$ increases from 2 to 2.5, and indicates a progressively increasing degree of underestimation of uncertainties by the Isoplot program. When p decreases from 0.15 to 0, we observe a significant decrease in the running mean of the $R_{MC/Iso}$ from ~2 to ~1.5, and then gradually decrease to >1. This relationship can further be illustrated by plotting $R_{MC/Iso}$ as a function of MSWD (which is dependent on p, Fig. 4f), and shows that $R_{MC/Iso}$ reduce from 2.5 to 1.5 as the MSWD increases from 1.3 to 2.5, ultimately $R_{MC/Iso}$ approaches one when the scatter is sufficiently large (i.e., MSWD >> 2.5). A notable feature here is the abrupt change in the relationship between $R_{mc/Iso}$ and probability (or MSWD in equivalent) when p approaches 0.15. Such an abrupt transition is mainly due to the contrasting error propagation strategies in Isoplot caused by the utilization of an arbitrary cut-off value.

These results indicate that uncertainties following the Model 1 scenarios in the Isoplot program are underestimated by 50%–60% compared to total uncertainties derived from the Monte Carlo method (as calculated by the difference between the uncertainties relative to the Monte Carlo based total uncertainties). For the Model 3 age in Isoplot, the uncertainties can also be underestimated by as much as 60%, though uncertainties become more comparable for increasing MSWD.

An underestimation of uncertainty could be detrimental in geological studies when high temporal resolution is essential. For example, when verifying the relationship between two geological processes that are indistinguishable in time (e.g., 1,000 ± 0.6 Ma and 999 ± 0.6 Ma), an underestimation of the uncertainties by 50% will yield ages of 1,000 ± 0.3 Ma and 999 ± 0.3 Ma, which could lead to a conclusion that the two geological events were not contemporaneous in time, hence rejecting a direct causal link between them. In contrast, with full propagation of the uncertainties, a potential causal link cannot be ruled out.

We speculate that the underestimation of uncertainties in the Model 1 ages arises from only considering analytical uncertainties without incorporating model uncertainties. This is supported by the observations that the analytical-only uncertainties from the Monte Carlo based method are comparable (though slightly larger, discussed below) to those from the Model 1 scenario in Isoplot program (Fig. 4g, h). The underestimation of uncertainties in the Model 3 ages is less transparent, but most likely due to an incomplete propagation of model uncertainties.

A further feature is that when $p > 0.15$, the analytical only uncertainties from our Monte Carlo method are slightly larger than those from the Model 1 solution (Fig. 4g, h). Such a discrepancy is expected based on York et al. [24] — the uncertainties from Monte Carlo method only becomes comparable with those from the least square method when sampling the least-squares-adjusted data points (i.e., the projection of the observed data point onto the isochron) by Monte Carlo, rather than sampling the observed data points as has been done here.

4. Potential to integrate geological information

An additional advantage of using the Monte Carlo based method is that the resulting distribution of age and initial isotopic ratios can be adjusted to integrate with geological information and produce improved chronological constraints. We demonstrate this by using a synthetic example consisting of 12 samples. Their $^{187}Re/^{188}Os$ and $^{187}Os/^{188}Os$ ratios and associated uncertainties including error correlations (Table 2) are used to determine their

Table 2
Re-Os data for the synthetic samples.

Sample No.	$^{187}Re/^{188}Os$	2-sigma	$^{188}Os/^{188}Os$	2-sigma	*rho*
1	100.000	1.540	1.504	0.023	0.936
2	200.000	2.940	2.407	0.024	0.473
3	300.000	4.830	3.311	0.037	0.764
4	400.000	3.960	4.215	0.078	0.565
5	500.000	7.150	5.118	0.057	0.635
6	600.000	10.200	6.022	0.090	0.484
7	700.000	11.620	6.926	0.082	0.949
8	800.000	12.880	7.830	0.078	0.945
9	900.000	12.780	8.733	0.096	0.910
10	1,000.000	10.900	9.637	0.165	0.994
11	1,100.000	19.580	10.541	0.065	0.477
12	1,200.000	7.440	11.444	0.161	0.452

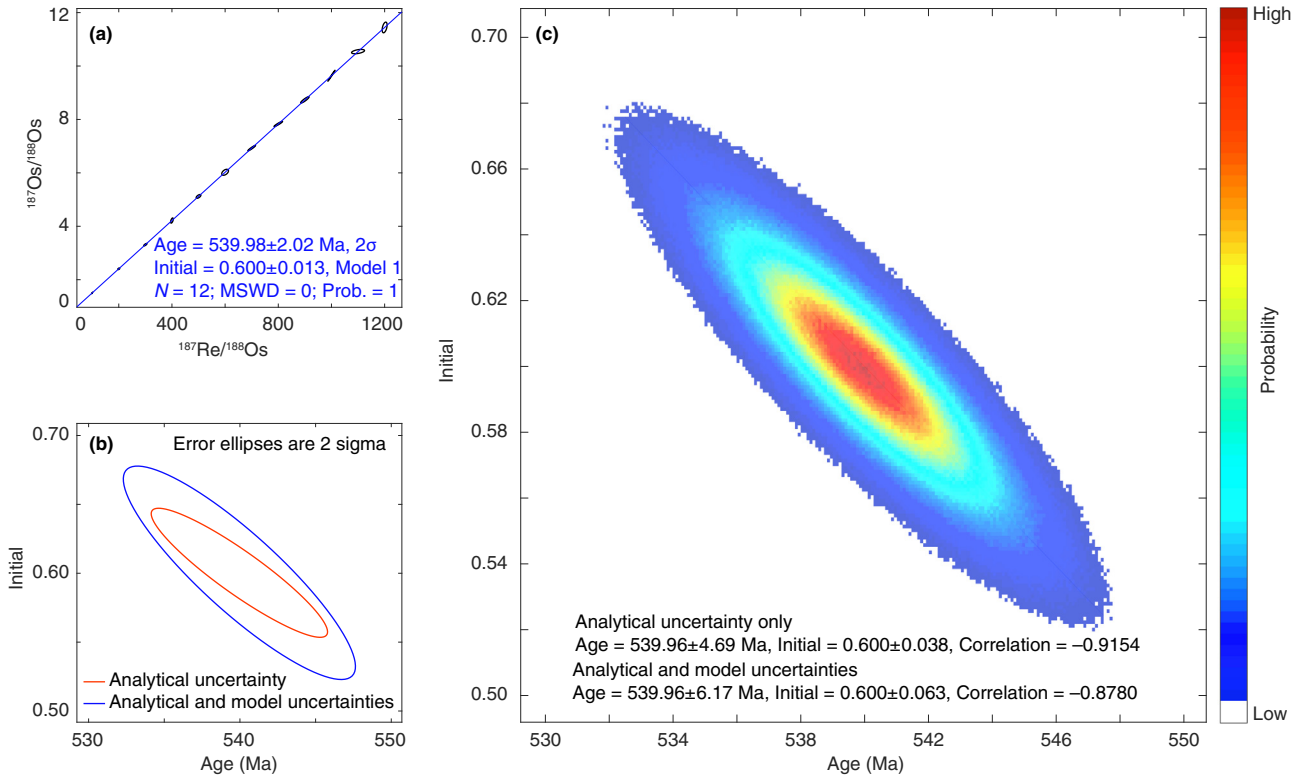


Fig. 5. Re-Os chronological results of the 12 synthetic samples using the Monte Carlo based method and the Isoplot program. (a) Isochron diagram using the algorithm of the Isoplot program; (b) results with analytical only and analytical + model uncertainties obtained from the Monte Carlo method at the 2-sigma level; (c) the final distribution of age and initial isotopic composition from the Monte Carlo based method is visualized.

age and initial isotopic ratio. Results obtained from the Monte Carlo method and the algorithm of the Isoplot program are presented in Fig. 5. The ages and initial isotopic ratios from the two methods are essentially the same (Isoplot age = 540 ± 2 Ma, initial $^{187}Os/^{188}Os = 0.600 \pm 0.013$; Monte Carlo age = 540 ± 6, initial $^{187}Os/^{188}Os = 0.600 \pm 0.063$), but uncertainties from the Isoplot program are significantly smaller as discussed above. If there is evidence that these samples are younger than 541 Ma, i.e., based on independent geological constraints, it is reasonable to discard regression results that are older than 541 Ma from the final distribution (Fig. 5). By doing so, the final distribution is altered, and skewed to younger ages and higher initial isotopic ratios (Fig. 6). If we consider quantiles to interpret uncertainties for this distribution, the age estimate changes to 539^{+2}_{-6} Ma and the initial isotopic composition is $0.616^{+0.026}_{-0.035}$ at the 95% percentile level. Similarly, if the initial isotopic ratio can be independently constrained, this information can also be integrated into the Monte Carlo method. This approach is analogous to a common practice in isochron dating, where a sample or a mineral containing low or negligible parent isotope is selected together with samples bearing high parent isotope for isochron dating (e.g., using matrix and garnet with low and high $^{176}Lu/^{177}Hf$ ratios, respectively for Lu-Hf dating; using plagioclase and pyroxene with low and high $^{147}Sm/^{144}Nd$ ratios, respectively for Sm-Nd dating), through which the *y*-intercept of the isochron is "fixed" by the sample (e.g., matrix and plagioclase) plotting near or at the *y*-intercept. It is possible that the independent constrained geological information would also have uncertainties or follow a certain distribution, these also can be considered in our Monte Carlo method.

In addition, with semi-quantitatively constrained contributions of analytical uncertainties to the total uncertainties, the new method provides guidance on how to yield refined chronological constrains. For example, if the uncertainties are dominated by analytical approaches, then improving experimental techniques would be an obvious next step to generate improved chronological information. In contrast, if analytical uncertainty is not the primary contributor to the total uncertainty, then the studied samples may not meet the criteria for isochron dating, and better sampling strategy would be the solution for refined chronological constrains.

5. Conclusions

A Monte Carlo based method is developed to estimate parameters (slope, *y*-intercept) in linear regression with full propagation of their uncertainties, which is then applied to data reduction for isochron geochronology. Crucially, the new method propagates both analytical and model uncertainties in a consistent manner, and also allows for the user to employ a posteriori geological criterion to yield refined chronological constrains and interpret the significance of the analytical/model uncertainty. Using a synthetic data set, results obtained from the Monte Carlo method and those from the Isoplot program are compared. The comparison indicates that although the estimates of the slope (age) and *y*-intercept (initial isotopic ratio) from both methods are similar, uncertainties following the Model 1 approach in the Isoplot program are underestimated by ~60%. For Model 3 solution in the Isoplot program, the uncertainties can be underestimated by as much as 60% depending on the goodness of fit, and the results from the two methods only start to converge when the goodness of fit approaches 0 (i.e., MSWD >> 2.5). We further demonstrate that geological information can be integrated into our Monte Carlo based method to yield improved chronological constraints.

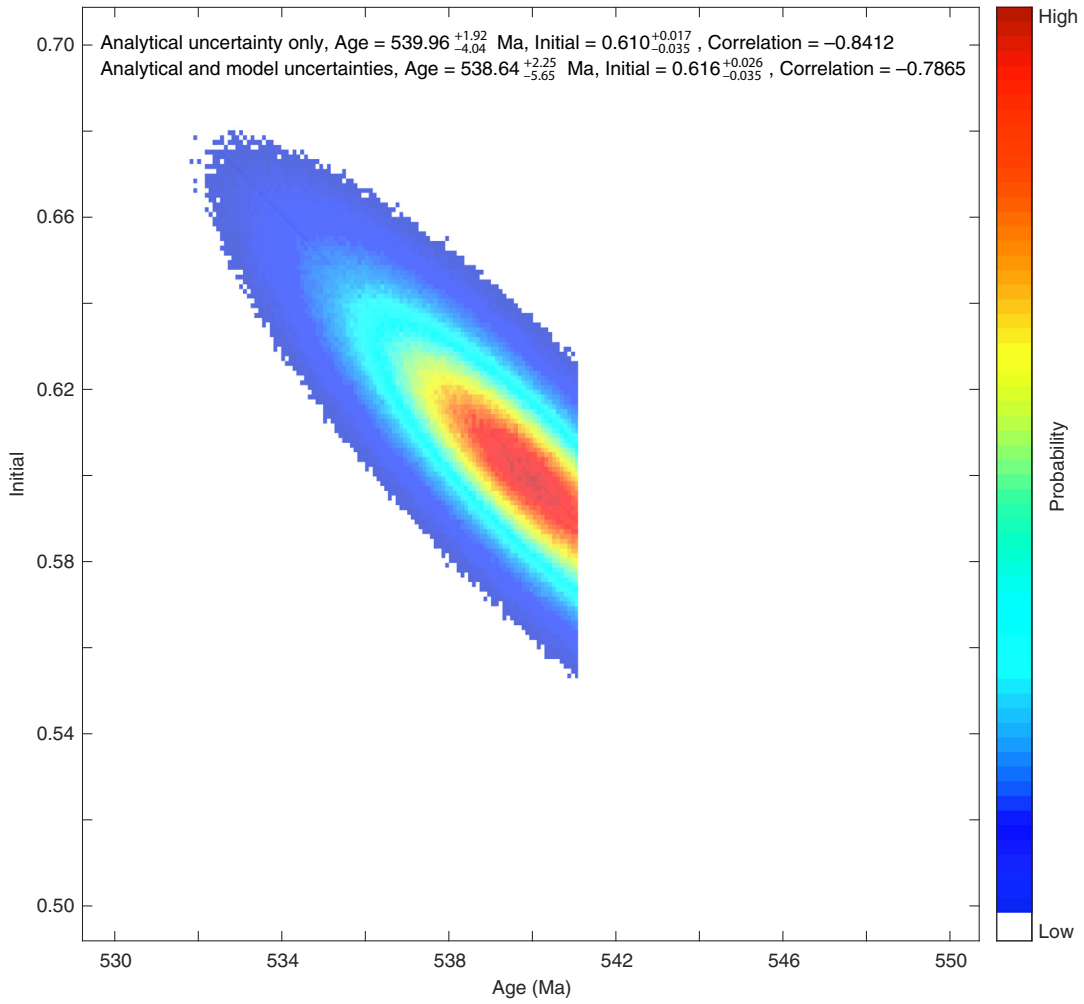


Fig. 6. Improving chronological constraints through integrating geological information for the synthetic example from Fig. 5. In this example, we assume that the samples are younger than 541 Ma from independent constraints, and hence simulation results >541 Ma could be removed to yield a better constrained chronological result. The approach demonstrated here is analogous to a common practice in isochron dating, where a sample or a mineral containing low or negligible parent isotope is selected together with samples bearing higher parent isotope for isochron dating, through which the *y*-intercept of the isochron is "fixed" by the sample plotting near or at the *y*-intercept.

Conflict of interest

The authors declare that they have no conflict of interest.

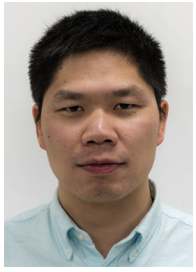
Acknowledgments

Yang Li acknowledges the NERC Numerical Earth Science Modelling courses at Durham University for developing coding skills, especially the help from Jeroen van Hunen and Dimitrios Michelioudakis. This work was funded by the State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences (SKL-K201706). David Selby acknowledges the total endowment fund. Alan D. Rooney and Yang Li acknowledge Yale University for support. The code associating this manuscript is available at https://github.com/wadesnoopy/mc.

References

[1] Harrison TM, Baldwin SL, Caffee M, et al. It's about time: opportunities and challenges for US geochronology. Institute of Geophysics and Planetary Physics Publication 6539. Los Angeles: University of California; 2015.
[2] Ludwig KR. User's manual for Isoplot 3.00: a geochronological toolkit for Microsoft Excel. Berkeley Geochronology Center Special Publication No. 4a 2003.
[3] Schmitz MD, Schoene B. Derivation of isotope ratios, errors, and error correlations for U-Pb geochronology using ^{205}Pb-^{235}U-(^{233}U)-spiked isotope dilution thermal ionization mass spectrometric data. Geochem Geophys Geosys 2007;8:1–20.
[4] Mattinson JM. Zircon U-Pb chemical abrasion ("CA-TIMS") method: combined annealing and multi-step partial dissolution analysis for improved precision and accuracy of zircon ages. Chem Geol 2005;220:47–66.
[5] McLean NM, Condon DJ, Schoene B, et al. Evaluating uncertainties in the calibration of isotopic reference materials and multi-element isotopic tracers (EARTHTIME Tracer Calibration Part II). Geochim Cosmochim Acta 2015;164:481–501.
[6] Renne PR, Mundil R, Balco G, et al. Joint determination of ^{40}K decay constants and ^{40}Ar*/^{40}K for the Fish Canyon sanidine standard, and improved accuracy for ^{40}Ar/^{39}Ar geochronology. Geochim Cosmochim Acta 2010;74:5349–67.
[7] Rivera TA, Storey M, Zeeden C, et al. A refined astronomically calibrated ^{40}Ar/^{39}Ar age for Fish Canyon sanidine. Earth Planet Sci Lett 2011;311:420–6.
[8] Creaser RA, Papanastassiou DA, Wasserburg GJ. Negative thermal ion mass-spectrometry of osmium, rhenium, and iridium. Geochim Cosmochim Acta 1991;55:397–401.
[9] Völkening J, Walczyk T, Heumann KG. Osmium isotope ratio determinations by negative thermal ionization mass spectrometry. Int J Mass Spectr Ion Process 1991;105:147–59.
[10] Jicha BR, Singer BS, Sobol P. Re-evaluation of the ages of ^{40}Ar/^{39}Ar sanidine standards and supereruptions in the western U.S. using a Noblesse multi-collector mass spectrometer. Chem Geol 2016;431:54–66.
[11] Condon DJ, Schoene B, McLean NM, et al. Metrology and traceability of U-Pb isotope dilution geochronology (EARTHTIME Tracer Calibration Part I). Geochim Cosmochim Acta 2015;164:464–80.
[12] Li XH, Li QL. Major advances in microbeam analytical techniques and their applications in Earth science. Sci Bull 2016;61:1785–7.
[13] Chen S, Wang XH, Niu YL, et al. Simple and cost-effective methods for precise analysis of trace element abundances in geological materials with ICP-MS. Sci Bull 2017;62:277–89.

[14] McLean NM, Bowring JF, Bowring SA. An algorithm for U-Pb isotope dilution data reduction and uncertainty propagation. Geochem Geophys Geosys 2011;12:1–26.
[15] McLean NM, Bowring JF, Gehrels G. Algorithms and software for U-Pb geochronology by LA-ICPMS. Geochem Geophys Geosys 2016;17:2480–96.
[16] Bowring JF, McLean NM, Bowring SA. Engineering cyber infrastructure for U-Pb geochronology: Tripoli and U-Pb_Redux. Geochem Geophys Geosys 2011:12.
[17] Horstwood MSA, Kosler J, Gehrels G, et al. Community-derived standards for LA-ICP-MS U-(Th-)Pb Geochronology - uncertainty propagation, age interpretation and data reporting. Geostand Geoanal Res 2016;40:311–32.
[18] Vermeesch P. Revised error propagation of $^{40}Ar/^{39}Ar$ data, including covariances. Geochim Cosmochim Acta 2015;171:325–37.
[19] Dutton A, Rubin K, McLean N, et al. Data reporting standards for publication of U-series data for geochronology and timescale assessment in the earth sciences. Quat Geochronol 2017;39:142–9.
[20] Keller CB, Schoene B, Samperton KM. A stochastic sampling approach to zircon eruption age interpretation. Geochem Perspect Lett 2018;8:31–5.
[21] Ludwig KR. ISOPLOT for MS-DOS, a plotting and regression program for radiogenic-isotope data, for IBM-PC compatible computers, version 1.00. US Geological Survey; 1988.
[22] York D. Least squares fitting of a straight line with correlated errors. Earth Planet Sci Lett 1968;5:320–4.
[23] York D. Least-squares fitting of a straight line. Canad J Phys 1966;44:1079–86.
[24] York D, Evensen NM, Martinez ML, et al. Unified equations for the slope, intercept, and standard errors of the best straight line. Am J Phys 2004;72:367–75.
[25] McIntyre GA, Brooks C, Compston W, et al. The statistical assessment of Rb-Sr isochrons. J Geophys Res 1966;71:5459–68.
[26] Ludwig KR. Mathematical–statistical treatment of data and errors for $^{230}Th/U$ geochronology. Rev Mineral Geochem 2003;52:631–56.
[27] Cumming GL. A recalculation of the age of the solar system. Canad J Earth Sci 1969;6:719–35.
[28] Ludwig KR, Titterington DM. Calculation of $^{230}Th/U$ isochrons, ages, and errors. Geochim Cosmochim Acta 1994;58:5031–42.
[29] James G, Witten D, Hastie T, et al. An Introduction to Statistical Learning, Vol. 112. Berlin: Springer; 2013.
[30] Smoliar MI, Walker RJ, Morgan JW. Re-Os ages of group IIA, IIIA, IVA, and IVB iron meteorites. Science 1996;271:1099–102.
[31] Selby D, Creaser RA, Stein HJ, et al. Assessment of the ^{187}Re decay constant by cross calibration of Re–Os molybdenite and U-Pb zircon chronometers in magmatic ore systems. Geochim Cosmochim Acta 2007;71:1999–2013.



Yang Li is an associate professor at the Institute of Geology and Geophysics, Chinese Academy of Sciences. He received his B.Eng. and Ph.D. from China University of Geosciences, Wuhan and Durham University in 2011 and 2017, respectively, followed by postdoctoral training at Durham and Yale University. His current research broadly seeks to understand how metalliferous volatiles are concentrated in the upper crust and their fates near the surface through integrating field geology, high-precision geochronology and isotopic microanalysis. Additional research interests include further developments and geological applications of the Re-Os chronometer.