

Response to the interactive comment by Ryan Ickert

Pieter Vermeesch
UCL Earth Sciences
p.vermeesch@ucl.ac.uk

I am grateful to Dr. Ickert for his review, which is one of the most careful and detailed ones that I have ever received. The text raises a number of pertinent points, which I will address in the revised manuscript. Following the format of the review, I will first give a general response to the most important points, and this will be followed by a detailed response to the specific comments.

1. Clarity and organisation

Ludwig (1998)'s "Treatment of concordant U/Pb ages" is one of my favourite papers of all time, because it is concise yet provides sufficient mathematical detail to verify the derivations and translate the algorithm into computer code. It was my aim to give my manuscript those same two qualities. However it appears that I have taken the concision too far in some places, whilst providing too much mathematical detail elsewhere. I will expand some of the descriptive text and move some of the mathematical detail to an appendix.

2. Example data

The reviewer points out that the reanalysis of Gibson et al. (2016)'s monazite U-Pb data is "at odds with the published results" due to a combination of true age heterogeneity and initial $^{230}\text{Th}/^{238}\text{U}$ -disequilibrium. The example data used in the manuscript was taken from one specific low-Y monazite crystal (grain #10) in one specific sample (BHE-01). The reported $^{208}\text{Pb}/^{232}\text{Th}$ -ages within this particular grain are fairly uniform, with a weighted mean of 19.9 ± 0.2 Ma. This is significantly older than the U-Th-Pb isochron age (17.8 ± 0.3 Ma). It is unlikely that the difference is due to initial $^{230}\text{Th}/^{238}\text{U}$ -disequilibrium, because correcting for this would move the age into the wrong direction. Repeating the $^{208}\text{Pb}/^{232}\text{Th}$ -age calculations of Gibson et al. (2016) shows that these authors did not apply a common Pb correction to their data. So I have good reasons to believe that the U-Th-Pb isochron age is in fact more accurate than the published values.

The reviewer is correct that the common Pb intercepts are too high. These estimates are imprecise, and the high MSWD reflects the difficulty of the U-Th-Pb isochron algorithm to fit both the U and Th data. So I will follow Dr. Ickert's suggestion and replace this example with two new ones: a carbonate dataset of Parrish et al. (2018) and an allanite dataset of Janots and Rubatto (2014). The carbonate dataset is an example of a low Th/U setting in which the ^{208}Pb -based common Pb correction is more precise than a conventional $^{207}\text{Pb}/^{206}\text{Pb}$ -based common Pb correction (Figure 1). The allanite dataset is an example of a high Th/U setting in which the $^{208}\text{Pb}/^{232}\text{Th}$ method offers greater precision than the U-Pb method. The Janots and Rubatto (2014) study used SIMS and so it is also possible to compare a ^{204}Pb -based common Pb correction with the new ^{208}Pb method. The comparison is favourable to the new U-Th-Pb isochron algorithm (Figure 2).

3. Novelty

Dr. Ickert writes that the isochron method presented in my manuscript "is only a slight modification of [Ludwig's] 'SemiTotal-Pb/U isochron' approach." and that the "advantage in forcing both Th-Pb and U-Pb concordance in constraining the Pb_c/U [...] isn't obvious to [him] from this manuscript."

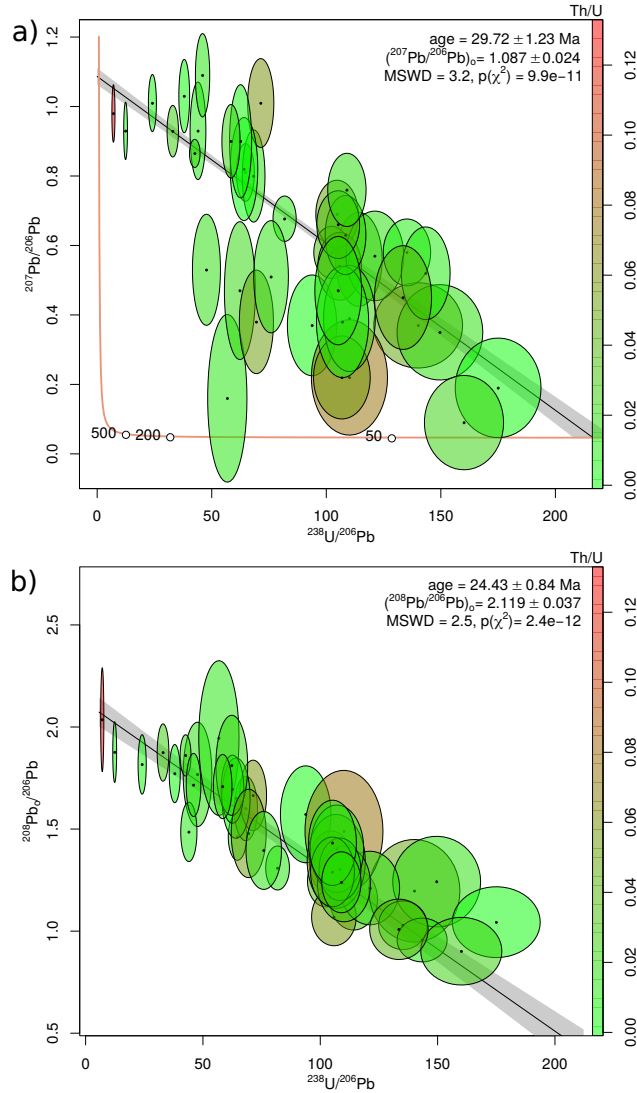


Figure 1: a) SemiTotal-Pb/U isochron (^{207}Pb -based common Pb correction) for Parrish et al. (2018)' chalk data; b) Total-Pb/U-Th isochron (^{208}Pb -based common Pb correction) shown in $^{208}\text{Pb}_o/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space. Colours indicate the Th/U-ratio. All uncertainties are shown at 1σ .

First, the new algorithm is not based on Ludwig (1998)'s **Semi**Total-Pb/U isochron method, but on his **Total**-Pb/U method. Second, the two new datasets will better illustrate the power of including Th-Pb in the isochron analysis. In the case of low-Th/U carbonate data, I will cite the relevant section of Parrish et al. (2018):

“This approach allows common ^{206}Pb to be quantified more robustly than methods using either ^{204}Pb or ^{207}Pb because the $^{208}\text{Pb}_c$ can be determined more precisely than using ^{204}Pb , ^{207}Pb or a combination of the two. In samples with low Th/U ratio this approach has two major advantages: (1) uncertainties of individual analyses are smaller, resulting in less scatter and improved uncertainty of isochron arrays; (2) it allows more reliable calculation of single spot ages and their weighted means. For most analyses, the uncertainties in measurement and consequent estimation of common Pb are smaller for $^{208}\text{Pb}/^{206}\text{Pb}$ than for

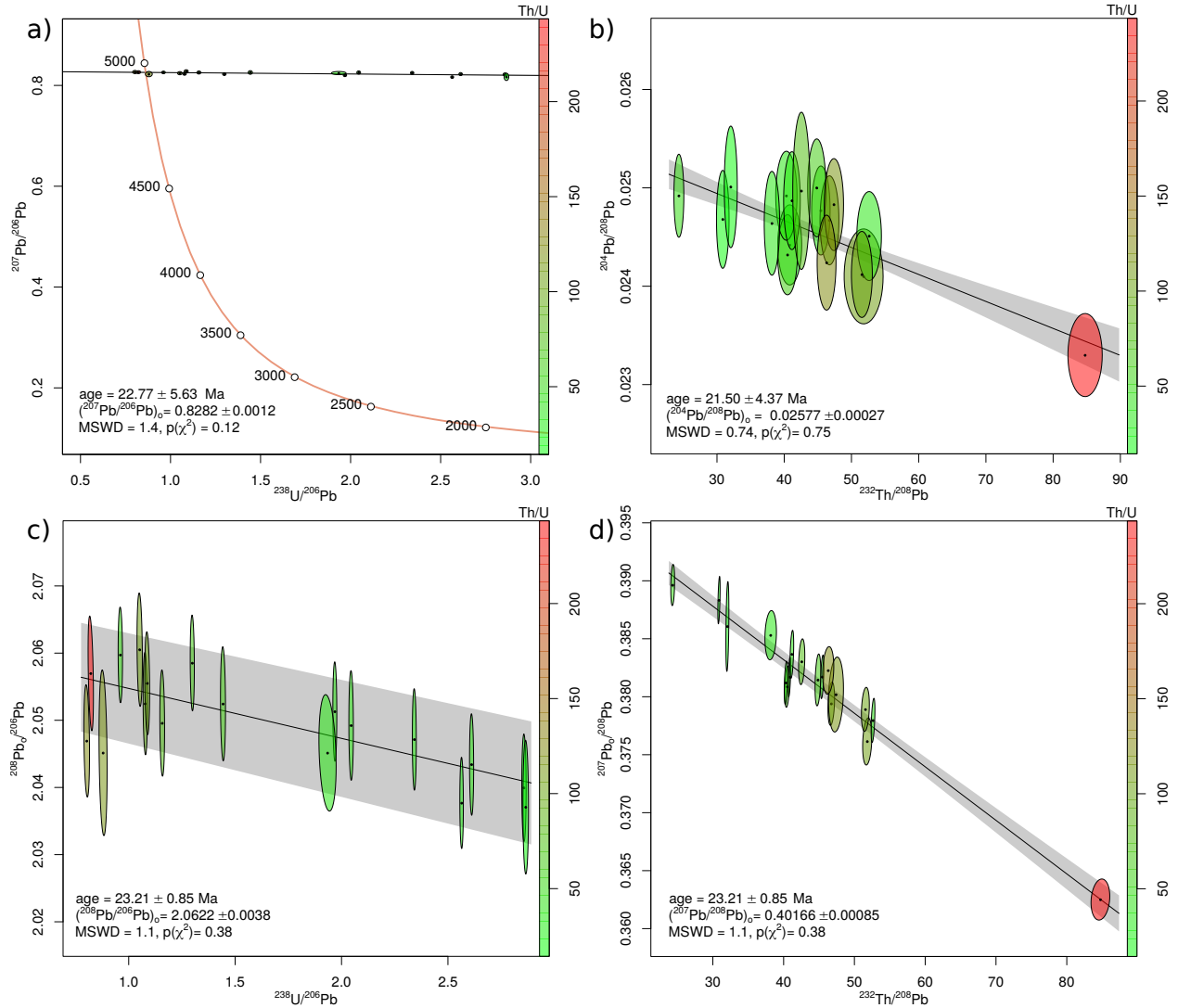


Figure 2: a) SemiTotal-Pb/U isochron (^{207}Pb -based common Pb correction) for Janots and Rubatto (2014)'s allanite data; b) Conventional Pb/Th-isochron (^{204}Pb -based common Pb correction); c) and d) Total-Pb/U-Th isochron (^{208}Pb -based common Pb correction) shown in $^{208}\text{Pb}_0/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space (c) and $^{206}\text{Pb}_0/^{208}\text{Pb} - ^{232}\text{Th}/^{208}\text{Pb}$ space (d). Colours indicate the Th/U-ratio. All uncertainties are shown at 1σ .

$^{207}\text{Pb}/^{206}\text{Pb}$. In all cases in this study, for spots with $>60\%$ radiogenic Pb, both regression ages agree within uncertainty. In all samples the ages and uncertainties of [U-Th-Pb isochron] regressions and weighted means of ^{208}Pb -corrected single spot ages agree within uncertainty, and both generally have smaller uncertainties and less regression scatter than analogous ^{207}Pb -corrected methods.”

For high-Th/U phosphate data, most of the geochronological power lies in the $^{208}\text{Pb}/^{232}\text{Th}$ clock. This chronometer lacks the equivalent of the U-Pb clock's $^{207}\text{Pb}/^{206}\text{Pb}$ -based common-Pb correction. In the absence of ^{204}Pb , the newly developed U-Th-Pb isochron is the only way to account for common Pb. I will add these details to the paper.

4. References

The original manuscript did not cite existing common Pb correction schemes proposed by Andersen (2002), Horstwood et al. (2003), Chew et al. (2014) among others. I will add these references to the revised manuscript, whilst highlighting their underlying assumptions and limitations. More specifically, the method of Andersen (2002) assumes that U-Th-Pb discordance “can be accounted for by a combination of lead loss at a defined time, and the presence of common lead of known composition”. This is clearly not the case for the carbonate and allanite data discussed in the revised manuscript; the ^{204}Pb -based approach of Horstwood et al. (2003) is complicated in the presence of ^{204}Hg and is imprecise due to the low abundance of ^{204}Pb (see Figure 2.b); and the limitations of ^{207}Pb -based methods as discussed by Chew et al. (2014) have already been explained in the quote by Parrish et al. (2018) given above.

Response to the detailed comments

The reviewer was puzzled why

“the Pb_c compositions (0.3685; 2.56; 11.71) and ages (17.71 Ma) appear in [Section 2] with no context.”

The optimal common Pb composition and age could be obtained by trial and error, until the samples plot along a line in $\text{Pb}/\text{Pb}-\text{U}/\text{Pb}$ space. To clarify this point, I will add some truly random guesses for the concordia age to the plot. See Figure 3 of this response letter. Please note that this new figure uses the Janots and Rubatto (2014) data instead of the Gibson et al. (2016) data from the original manuscript.

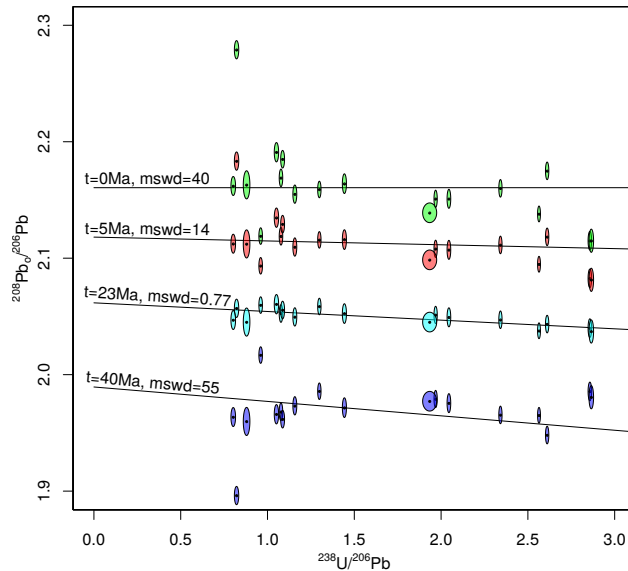


Figure 3: U-Th-Pb data for allanite sample MF482 of Janots and Rubatto (2014) shown on a $^{208}\text{Pb}_0/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ isochron plot. Green ellipses represent the raw data. Red, light and dark blue ellipses show the same measurements with 5, 23 and 40 Ma worth of radiogenic ^{208}Pb removed, respectively. The misfit of the radiogenic ^{208}Pb -corrected data around the best fit line is expressed as weighted square of mean deviates (mswd, McIntyre et al., 1966) values. The y-intercept yields the common Pb isotopic composition. Error ellipses are shown at 1σ .

“the covariance matrix is introduced in equation 11, but not identified until just above equation 18 in the next column.”

Equation 11 contains five different parameters, which are defined in terms of other parameters. Explaining the meaning of all these parameters takes space. I will address this issue by moving lines 110-120 to an Appendix.

“The omegas in equation 11 are never identified.”

Here I simply followed Ludwig (1998): the omegas are defined implicitly in terms of the inverted covariance matrix.

“If the author just wants to write out derivations of equations, they should be in an appendix. 12, 13 and 14 should also be written out with the original variable names ($^{206}\text{Pb}/^{238}\text{U}$, $^{207}\text{Pb}/^{235}\text{U}$ etc.) and the significance of these equations explained to a reader.”

It is not easy to fit the original variables in *GChron*’s two-column format. But what I can do is follow Ludwig (1998) and define the variables before instead of after using them. Equations 18-20, 32-41 and 46-55 will be moved to an appendix.

“there is nothing special about using ^{208}Pb as the index isotope”

^{208}Pb was chosen as an index isotope so as to replace ^{204}Pb in Ludwig (1998)’s Total-Pb/U algorithm. This is different from the alternative formulations proposed by the reviewer, which refer to the SemiTotal-Pb/U algorithm. It is true that the Total-Pb/U regression problem can be redefined in terms of Tera-Wasserburg variables instead of the current Wetherill variables. But the solution is easier and cleaner in Wetherill space.

“Line 5: 232/208 is not as often considered because there are few isotope dilution measurements of ^{232}Th (because they are harder to make by TIMS, and few labs want to do mixed TIMS-MC-ICPMS analyses), because zircon is by far the most well used U-Th-Pb chronometer (where Th-Pb provides little additional information), and because Th/U fractionation occurs in actinide rich minerals (like allanite), complicating the systematics. The lack of statistical tools is very much a second order reason to not jointly consider all the decay schemes.”

^{208}Pb and ^{232}Th are easy to measure by LA-ICP-MS, which has become by far the most widely used analytical technique for U-Th-Pb geochronology. Zircon is indeed the most widely used mineral phase for U-Pb geochronology, but in recent years there has been a rapid rise in the number of studies that use other mineral phases such as apatite, allanite, rutile, and carbonates. Two examples of such studies will be included in the revised manuscript, showcasing the gains in accuracy and precision that can be made with the U-Th-Pb isochron method. The effects of Th/U fraction can quite easily be quantified by comparing the Th/U ratio of the dated mineral with that of the whole rock (Schärer, 1984). This correction has already been implemented in *IsoplotR*.

“It is possible to accurately measure ^{204}Pb in ICPMS measurements but becomes increasingly difficult with decreasing amounts of Pb_c . So Pb_c -rich minerals don’t necessarily suffer from this problem (and these are the minerals for which this correction is most important).”

Speaking from experience, I am unable to accurately measure ^{204}Pb using my quadrupole LA-ICP-MS instrument at UCL, even with gold filters. The blank is more than 90% of the signal. For young and U,Th-poor samples, it is difficult enough to measure the radiogenic Pb, let alone the common ^{204}Pb .

“The point about dwell time is not particularly important. Removing one isotope from a run table doesn’t provide a huge improvement in on-peak time from a practical perspective (it’s a square root problem)”

In the case of Janots and Rubatto (2014)'s allanite study, there is 38 times more $^{208}\text{Pb}_c$ than ^{204}Pb (Figure 2.b). So for the same dwell time, the $^{208}\text{Pb}_c$ measurement would be more than six times more precise than the ^{204}Pb measurement. Conversely, the same precision can be achieved for $^{208}\text{Pb}_c$ in one sixth of the time as ^{204}Pb . Conclusion: the square root problem is important.

“Section 6: This is a very important contribution and it’s unfortunate that it is buried in a small section of a paper on a different topic. It’s far too short to do it any justice and I hope that this receives a much more robust treatment elsewhere in the literature.”

By moving much of the mathematical detail to an appendix, Section 6 will gain prominence. My solution to the problem of asymmetric confidence intervals will be further explored in a forthcoming paper on disequilibrium corrections that I will co-author with Dr. Noah McLean and others later this year.

“Section 7: This is just a constrained Pb_c regression, and it would be useful to refer to the literature where this has been done before.”

I will add another reference to Chew et al. (2014) here.

“What would be useful, and I urge the author to do this, is to demonstrate a specific advantage of this technique (or any of those described herein) over a conventional interpretation. Show both interpretations back-to-back so we can see the advantage.”

See Figures 1 and 2 of this response letter, which will be added to the revised manuscript.

“8. Does the title clearly reflect the contents of the paper? No, it is very general”

The title of Ludwig (1998) is also very general (“On the treatment of concordant uranium-lead ages”). But I will follow the reviewer’s suggestion and change the title to: “Unifying the U–Pb and Th–Pb methods: joint isochron regression and common lead correction”.

References

- Andersen, T. Correction of common lead in U–Pb analyses that do not report ^{204}Pb . *Chemical geology*, 192 (1-2):59–79, 2002.
- Chew, D., Petrus, J., and Kamber, B. U–Pb LA–ICPMS dating using accessory mineral standards with variable common Pb. *Chemical Geology*, 363:185–199, 2014.
- Gibson, R., Godin, L., Kellett, D. A., Cottle, J. M., and Archibald, D. Diachronous deformation along the base of the Himalayan metamorphic core, west-central Nepal. *Geological Society of America Bulletin*, 128 (5-6):860–878, 2016.
- Horstwood, M. S., Foster, G. L., Parrish, R. R., Noble, S. R., and Nowell, G. M. Common-Pb corrected in situ U–Pb accessory mineral geochronology by LA–MC–ICP–MS. *Journal of Analytical Atomic Spectrometry*, 18(8):837–846, 2003.
- Janots, E. and Rubatto, D. U–Th–Pb dating of collision in the external Alpine domains (Urseren zone, Switzerland) using low temperature allanite and monazite. *Lithos*, 184:155–166, 2014.
- Ludwig, K. R. On the treatment of concordant uranium-lead ages. *Geochimica et Cosmochimica Acta*, 62: 665–676, 1998. doi: 10.1016/S0016-7037(98)00059-3.
- McIntyre, G. A., Brooks, C., Compston, W., and Turek, A. The Statistical Assessment of Rb–Sr Isochrons. *Journal of Geophysical Research*, 71:5459–5468, 1966.
- Parrish, R. R., Parrish, C. M., and Lasalle, S. Vein calcite dating reveals Pyrenean orogen as cause of Paleogene deformation in southern England. *Journal of the Geological Society*, 175(3):425–442, 2018.
- Schärer, U. The effect of initial ^{230}Th disequilibrium on young UPb ages: the Makalu case, Himalaya. *Earth and Planetary Science Letters*, 67(2):191–204, 1984.

I would like to thank Dr. Samperton for his positive review. It contains mostly minor suggestions apart from one comment, which prompted me to confront a bigger issue that I had avoided in the original manuscript.

Page 1, Lines 3-4: “The $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{235}\text{U}$ decay systems are routinely combined to improve accuracy”. May be more appropriate to have something along the lines of “...are routinely combined to improve the assessment of accuracy”?

It is true that, in many geological applications, the ^{238}Pb – ^{206}Pb and ^{235}Pb – ^{207}Pb clocks are simply plotted together to assess concordance, after which a simple weighted mean $^{206}\text{Pb}/^{238}\text{U}$ age is calculated. However, Ludwig (1998) showed that the clocks can also be combined to estimate a hybrid (concordia or isochron) age, which in theory is more accurate than either the $^{206}\text{Pb}/^{238}\text{U}$ or $^{207}\text{Pb}/^{235}\text{U}$ age. The aim of the U-Th-Pb isochron paper is to explore this application further but including the $^{208}\text{Pb}/^{232}\text{Th}$ clock as well. So in this case I maintain that “improving accuracy” is a more appropriate term than “improving the assessment of accuracy”.

Page 1, Lines 28-31: “Nevertheless, it manages to fit the data very well. The method should work even better for low-Th phases such as carbonates.” These sentences are far too subjective and informal, please rewrite.

I will add two new datasets to the paper, including a carbonate example (Parrish et al., 2018) and an allanite example (Janots and Rubatto, 2014). By comparing conventional common-Pb corrections for these data with the new Total-Pb/U-Th algorithm, the improvement in precision and accuracy will be much clearer to the reader. See Figures 1 and 2 of this response letter for further details.

Technically, the proportions are a function of the Th/U-ratio, age, AND the $^{238}\text{U}/^{235}\text{U}$ ratio. Here and later in the manuscript the author assumes the mean terrestrial zircon $^{238}\text{U}/^{235}\text{U}$ value (137.818, without uncertainty) of Hiess et al. (2012). While for many (most?) applications of the algorithm this assumption is possibly acceptable, insofar as broadening the general applicability of this approach I think it is worth stating this point explicitly.

The reviewer is correct that the $^{238}\text{U}/^{235}\text{U}$ ratio affects the $^{207}\text{Pb}/^{206}\text{Pb}$ ratio. However as long as all the analyses are cogenetic (which is a requirement for isochron regression), departure of the $^{238}\text{U}/^{235}\text{U}$ ratio from the Hiess et al. (2012) values actually does not hurt the accuracy of the isochron age. This is because, in Equation 12 of the original manuscript, $^{238}\text{U}/^{235}\text{U}$ is multiplied with the common-Pb ratio β . So as long as $^{238}\text{U}/^{235}\text{U}$ and β do not vary between aliquots, an overestimation of one translates into an underestimation of the other without affecting t .

So the uncertainty of the $^{238}\text{U}/^{235}\text{U}$ ratio (U in Equation 12) only matters for the error propagation of β . It is not easy to address this issue with the maximum likelihood formulation of the original manuscript, in which U occurs in a product with γ . If the uncertainty of U is to be propagated, it is no longer possible to reformulate the sum of squares S in terms of the Th/Pb misfit parameter M (Equation 23). Similarly, the analytical uncertainty of the measured $^{232}\text{Th}/^{238}\text{U}$ ratio (W in Equations 12-14) is also difficult to propagate.

The solution to both of these problems is straightforward in theory, but complicated in practice. Recalling the general equation for the sum-of-squares (Equation 11 of the original manuscript):

$$S = \Delta^T (J^T \Sigma J)^{-1} \Delta$$

we can replace Equations 12 (for J) and 13 (for Σ) with

$$\Sigma = \begin{bmatrix} s[X]^2 & s[X, Y] & s[X, Z] & s[X, W] & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ s[Y, X] & s[Y]^2 & s[Y, Z] & s[Y, W] & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ s[Z, X] & s[Z, Y] & s[Z]^2 & s[Z, W] & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ s[W, X] & s[W, Y] & s[W, Z] & s[W]^2 & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & s[\lambda_{35}]^2 & 0 & 0 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0 & s[\lambda_{38}]^2 & 0 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0 & 0 & s[\lambda_{32}]^2 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0 & 0 & 0 & s[U]^2 \end{bmatrix}$$

and

$$J = \begin{bmatrix} 1_{n,n} & 0_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & 1_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & 0_{n \times n} & 1_{n \times n} \\ -U\beta\gamma & -\alpha\gamma & 0_{n \times n} \\ -t_{1 \times n}e^{\lambda_{35}t} & 0_{1 \times n} & 0_{1 \times n} \\ 0_{1 \times n} & -t_{1 \times n}e^{\lambda_{38}t} & 0_{1 \times n} \\ 0_{1 \times n} & 0_{1 \times n} & -t_{1 \times n}e^{\lambda_{32}t} \\ -\beta W\gamma & 0_{1 \times n} & 0_{1 \times n} \end{bmatrix}$$

respectively. Unfortunately, taking matrix derivatives of S is extremely difficult to do by hand for this generalised formulation. In well behaved cases, R's optimisation function manages to calculate them numerically. But the numerical stability of these solutions is much poorer than that of the original algorithm.

An additional advantage of the new formulation is its ability to accommodate a second type of overdispersion model. Section 5 of the original manuscript parameterised the overdispersion in terms of the concordia intercept age. With the generalised formulation of the maximum likelihood problem, it is also possible to attribute the excess dispersion to the common Pb composition. In this case we replace Equation 44 of the original manuscript with the following alternative:

$$J_{\omega} = \begin{bmatrix} -UW\gamma \\ -W\gamma \\ 0_{n \times n} \end{bmatrix}$$

Again, the numerical stability of this formulation is not as good as that of the original algorithm. If I find a way to increase this stability, then I will use the new algorithm. Otherwise I will stick with the original version and be more clear about its limitations.

throughout the manuscript a quantitative blank correction is not addressed, which is fine, but if so a statement should be made here that the equations as currently formulated assume a trivial Pb blank component.

I will add a line to clarify that the data are assumed to have been blank corrected.

Couldn't you pull a representative carbonates dataset to demonstrate this point explicitly? I'd be interested to see this.

A carbonate example will be added to the revised manuscript. See Figure 1 of this response letter.

You mention in passing that data are "overdispersed if.. MSWD $\gg 1$ ". However, I think it worth stating a more general point about the acceptable MSWD range as a function of the number of degrees of freedom (i.e., data points), a la Wendt and Carl (1991). I think it worth citing Wendt and Carl (1991) here, as well as presenting a general formula for the range/uncertainty on the MSWD itself, beyond stating the oversimplification that data are overdispersed when MSWD1.

I will add a reference to Wendt and Carl (1991). However it is also important not to overly rely on MSWDs and p-values. It is possible for a precise dataset with an MSWD value of 100 to be more valuable than an imprecise dataset with an MSWD of 1. What matters is not so much whether a dataset is overdispersed or not, but rather *how* dispersed it is. This key point is addressed in Section 5 of the paper.

You should cite R for those not in the know

I will add the requested citation.

References

- Hiess, J., Condon, D. J., McLean, N., and Noble, S. R. $^{238}\text{U}/^{235}\text{U}$ systematics in terrestrial uranium-bearing minerals. *Science*, 335(6076):1610–1614, 2012.
- Janots, E. and Rubatto, D. U–Th–Pb dating of collision in the external Alpine domains (Urseren zone, Switzerland) using low temperature allanite and monazite. *Lithos*, 184:155–166, 2014.
- Ludwig, K. R. On the treatment of concordant uranium-lead ages. *Geochimica et Cosmochimica Acta*, 62: 665–676, 1998. doi: 10.1016/S0016-7037(98)00059-3.
- Parrish, R. R., Parrish, C. M., and Lasalle, S. Vein calcite dating reveals Pyrenean orogen as cause of Paleogene deformation in southern England. *Journal of the Geological Society*, 175(3):425–442, 2018.

~~U-Th-Pb discordia~~ Unifying the U-Pb and Th-Pb methods: joint isochron regression and common lead correction

Pieter Vermeesch

London Geochronology Centre, Department of Earth Sciences, University College London, United Kingdom

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

Abstract. The actinide elements U and Th undergo radioactive decay to three isotopes of Pb, forming the basis of three coupled geochronometers. The $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{235}\text{U}$ decay systems are routinely combined to improve accuracy. Joint consideration with the $^{208}\text{Pb}/^{232}\text{Th}$ decay system is less common. This paper aims to change this. Co-measured $^{208}\text{Pb}/^{232}\text{Th}$ is particularly useful for discordant samples containing variable amounts of non-radiogenic ('common') Pb.

5 The paper presents a maximum likelihood algorithm for joint isochron regression of the $^{206}\text{Pb}/^{238}\text{Pb}$, $^{207}\text{Pb}/^{235}\text{Pb}$, and $^{208}\text{Pb}/^{232}\text{Th}$ chronometers. Given a set of cogenetic samples, ~~the algorithm~~ this 'Total-Pb/U-Th algorithm' estimates the common Pb composition and concordia intercept age. ~~U-Th-Pb~~ U-Th-Pb data can be visualised on a conventional Wetherill or Tera-Wasserburg concordia diagram, or on a $^{208}\text{Pb}/^{232}\text{Th}$ vs. $^{206}\text{Pb}/^{238}\text{U}$ plot. Alternatively, the results of the new discordia regression algorithm can also be visualised as a $^{208}\text{Pb}_c/^{206}\text{Pb}$ vs. $^{238}\text{U}/^{206}\text{Pb}$ or $^{208}\text{Pb}_c/^{207}\text{Pb}$ vs. $^{238}\text{U}/^{207}\text{Pb}$ isochron, where
10 $^{208}\text{Pb}_c$ represents the common ^{208}Pb component. In its most general form, the Total-Pb/U-Th algorithm accounts for the uncertainties of all isotopic ratios involved, including the $^{232}\text{Th}/^{238}\text{U}$ -ratio, as well as the systematic uncertainties associated with the decay constants and the $^{238}\text{U}/^{235}\text{U}$ -ratio. However, numerical stability is greatly improved when the dependency on the $^{232}\text{Th}/^{238}\text{U}$ -ratio uncertainty is dropped.

For detrital minerals, it is generally not safe to assume a shared common Pb composition and concordia intercept age. In
15 this case the ~~U-Th-Pb discordia~~ Total-Pb/U-Th regression method must be modified by tying it to a terrestrial lead evolution model. Thus also detrital common Pb correction can be formulated in a maximum likelihood sense.

The new method was applied to ~~a published monazite dataset with a three published datasets, including low Th/U-ratio of ~ 10 , resulting in a significant radiogenic ^{208}Pb component. Therefore the case study represents a 'worst case scenario' for the new algorithm. Nevertheless, it manages to fit the data very well. The method should work even better in low-Th phases such
20 as carbonates. The degree to which the dispersion of the data around the isochron line matches the analytical uncertainties can be assessed using the mean square of the weighted deviates (MSWD) statistic. A modified four parameter version of the regression algorithm quantifies this overdispersion, providing potentially valuable geological insight into the processes that control isotopic closure~~ U carbonates, high Th/U allanites and overdispersed monazites. The carbonate example illustrates how the Total-Pb/U-Th method achieves a more precise common-Pb correction than a conventional ^{207}Pb -based approach.
25 The allanite sample shows the significant gain in both precision and accuracy that is made when the Th-Pb decay system is jointly considered with the U-Pb system. Finally the monazite example is used to illustrate how the Total-Pb/U-Th regression algorithm can be modified to include an overdispersion parameter.

All the parameters in the discordia regression method (including the age and the overdispersion parameter) are strictly positive quantities that exhibit skewed error distributions near zero. This skewness can be accounted for using the profile log-likelihood method, or by recasting the regression algorithm in terms of logarithmic quantities. Both approaches yield realistic asymmetric confidence intervals for the model parameters. The new algorithm is flexible enough that it can accommodate disequilibrium corrections and inter-sample error correlations when these are provided by the user. All the methods presented in this paper have been added to the `IsoplotR` software package. This will hopefully encourage geochronologists to take full advantage of the entire ~~U-Th-Pb~~ U-Th-Pb decay system.

35 1 Introduction

The lead content of uranium-bearing minerals comprises two components:

1. Non-radiogenic (a.k.a. initial or ‘common’) Pb is inherited from the environment during crystallisation. It contains all of lead’s four stable isotopes (^{204}Pb , ^{206}Pb , ^{207}Pb and ^{208}Pb) in fixed proportions for a given sample.
2. Radiogenic Pb is added to the common Pb after crystallisation due to the decay of U and Th. It contains only three isotopes (^{206}Pb , ^{207}Pb and ^{208}Pb), which occur in variable proportions as a function of the Th/U-ratio and age.

Denoting the measured and non-radiogenic components with subscripts ‘ m ’ and ‘ c ’ respectively, and assuming initial secular equilibrium, we can write:

$$^{204}\text{Pb}_m = ^{204}\text{Pb}_c \quad (1)$$

$$^{206}\text{Pb}_m = ^{206}\text{Pb}_c + ^{238}\text{U}_m (e^{\lambda_{38}t} - 1) \quad (2)$$

$$50 \quad ^{207}\text{Pb}_m = ^{207}\text{Pb}_c + ^{235}\text{U}_m (e^{\lambda_{35}t} - 1) \quad (3)$$

$$^{208}\text{Pb}_m = ^{208}\text{Pb}_c + ^{232}\text{Th}_m (e^{\lambda_{32}t} - 1) \quad (4)$$

where λ_{38} , λ_{35} and λ_{32} are the decay constants of ^{238}U , ^{235}U and ^{232}Th , respectively, and t is the time elapsed since isotopic closure. In order to accurately estimate t , the common Pb composition is needed. One way to account for common Pb is to normalise all the measurements to ^{204}Pb . For example, using the $^{238}\text{U} - ^{206}\text{Pb}$ decay scheme:

$$50 \quad \left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}} \right]_m = \left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}} \right]_c + \left[\frac{^{238}\text{U}}{^{204}\text{Pb}} \right]_m (e^{\lambda_{38}t} - 1) \quad (5)$$

Applying Equation 5 to multiple cogenetic aliquots of the same sample defines an isochron with slope $(e^{\lambda_{38}t} - 1)$ and intercept $[\frac{^{206}\text{Pb}}{^{204}\text{Pb}}]_c$. Alternatively, and equivalently, an ‘inverse’ isochron line can be defined as:

$$\left[\frac{^{204}\text{Pb}}{^{206}\text{Pb}} \right]_m = \left[\frac{^{204}\text{Pb}}{^{206}\text{Pb}} \right]_c \left\{ 1 - \left[\frac{^{238}\text{U}}{^{206}\text{Pb}} \right]_m (e^{\lambda_{38}t} - 1) \right\} \quad (6)$$

In this case, the isochron is a line whose y-intercept defines the common $^{204}\text{Pb}/^{206}\text{Pb}$ -ratio, and the x-intercept determines the radiogenic $^{238}\text{U}/^{206}\text{Pb}$ -ratio.

The isochron concept can easily be applied to the $^{235}\text{U} - ^{207}\text{Pb}$ system, by replacing ^{206}Pb with ^{207}Pb , ^{238}Pb with ^{235}Pb and λ_{38} with λ_{35} in Equations 5 and 6. The accuracy and precision of the calculation can be further improved by solving the $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{235}\text{U}$ isochron equations simultaneously and requiring t to be the same in both systems. The resulting three-dimensional constrained isochron is known as a ‘Total-Pb/U isochron’ ~~and represents the pinnacle of statistical rigour in~~
60 ~~U-Pb geochronology (Ludwig, 1998).~~ (Ludwig, 1998).

In igneous samples, the conventional ~~total-Pb~~ Total-Pb/U isochron requires isotopic data for two or more cogenetic aliquots. In the simplest case, a two-point isochron can be formed by analysing the U-Pb composition of the U-bearing phase of interest along with a cogenetic mineral devoid of U (e.g. feldspar). In detrital samples, the common Pb intercept of the isochron can be anchored to some nominal value, or to a terrestrial lead evolution model (e.g., Stacey and Kramers, 1975). Thus, the
65 ^{204}Pb -based total U-Pb isochron method is beneficial to nearly all applications of the U-Pb method.

Unfortunately, ^{204}Pb -based common Pb correction is not always practical. First, not all mass spectrometers are able to measure ^{204}Pb with sufficient precision and accuracy. In some ICP-MS instruments, the presence of an isobaric interference with ^{204}Hg precludes accurate ^{204}Pb measurements. And second, because ^{204}Pb is by far the least abundant of lead’s four naturally occurring isotopes, it requires the longest dwell times. For single collector instruments, this ~~degrades~~ reduces the
70 precision of the other isotopes ~~to the point where the analytical cost of measuring~~.

To overcome these problems, alternative common-Pb correction schemes have been proposed that use ^{207}Pb or ^{208}Pb instead of ^{204}Pb . The ‘SemiTotal-Pb/U isochron’ method is based on linear regression of ^{206}Pb – ^{207}Pb – ^{238}U -data in Tera-Wasserburg space (Ludwig, 1998; Williams, 1998; Chew et al., 2011). It assumes that all the samples are cogenetic and form a simple two component mixture between common Pb and radiogenic Pb. The common Pb then marks the intercept with the $^{207}\text{Pb}/^{206}\text{Pb}$ -axis, and the radiogenic Pb can be obtained from the intersection of the isochron with the concordia line. The ~~^{204}Pb may outweigh its benefits.~~ ^{207}Pb -based common Pb correction only works if the assumption of initial concordance is valid, if ^{207}Pb can be measured with sufficient precision, and if there is enough spread in the initial Pb/U-ratios to produce a statistically robust isochron.

~~One way to overcome both problems is to use~~ Andersen (2002) introduced a ^{208}Pb -based common-Pb correction scheme that does not require initial concordance. His method assumes that U–Th–Pb discordance can be accounted for by a combination of
80 lead loss at a defined time, and the presence of common lead of known composition. However in most cases neither the timing of lead loss, nor the composition of the common lead are known. Furthermore, the assumptions that underlie the Andersen (2002) method were tailored to the mineral zircon, but do not apply so much to other minerals such as carbonates, which crystallise at low temperatures and do not experience diffusive lead loss.

This paper introduces a ‘Total-Pb/U–Th isochron’ algorithm that uses the $^{232}\text{Th} - ^{208}\text{Pb}$ decay scheme to determine the common Pb component. ~~Thus, if we can estimate~~ Unlike the Andersen (2002) method, it does not require the common Pb composition to be pre-specified, but assumes that no Pb-loss has occurred. The new algorithm is based on Ludwig (1998)’s Total-Pb/U isochron method, but uses $^{208}\text{Pb}_c$ instead of ^{204}Pb in Equation 4, then Equation 5 can be replaced with 5:

$$\frac{^{206}\text{Pb}_m}{^{208}\text{Pb}_c} = \left[\frac{^{206}\text{Pb}}{^{208}\text{Pb}} \right]_c + \frac{^{238}\text{U}_m}{^{208}\text{Pb}_c} (e^{\lambda_{38}t} - 1) \quad (7)$$

and similarly for Equation 6 and the $^{235}\text{U} - ^{207}\text{Pb}$ equivalents of Equations 5 and 6.

90 This paper introduces a ‘U–Th–Pb isochron’ algorithm that achieves this reformulation. The algorithm is similar to Ludwig (1998) ~~’s total Pb/U isochron, but uses a unified approach that accommodates both random and systematic uncertainties.~~ The algorithms introduced in this paper will be illustrated using ~~a published U–Th–Pb dataset for monazite grain #10 in sample BHE-01~~ three published U–Th–Pb datasets, which showcase how the combined U–Th–Pb approach improves both the precision and accuracy of U–Pb geochronology (Section 4). The cases studies include a carbonate dataset of Parrish et al. (2018), an allanite
 95 ~~dataset of Janots and Rubatto (2014), and an overdispersed monazite dataset of Gibson et al. (2016).~~ dataset of Janots and Rubatto (2014), and an overdispersed monazite dataset of Gibson et al. (2016). ~~With a~~ The carbonate dataset is an example of a low Th/U-ratio of ~ 10 , this sample represents a ‘worst case scenario’ in the sense that the addition of lots of radiogenic ^{208}Pb complicates the removal of the common U setting in which the ^{208}Pb -based common Pb correction is more precise than a conventional ^{207}Pb component. ~~The fact that the new algorithm works very well for monazite implies that it is generally applicable low Th phases such as carbonates~~ ^{206}Pb -based common Pb correction. The allanite dataset is
 100 an example of a high Th/U setting in which the $^{208}\text{Pb}/^{232}\text{Th}$ method offers greater precision than the U–Pb method. The Janots and Rubatto (2014) study used SIMS and therefore also offers an opportunity to compare the new ^{208}Pb method with a conventional ^{204}Pb -based common Pb correction.

Section 5 shows how the isochron regression algorithm can be modified to accommodate strongly skewed uncertainty distributions, using a simple logarithmic change of variables. The Total-Pb/U–Th isochron algorithm assumes that all aliquots
 105 are cogenetic. However Section 6 shows how the algorithm can be adapted to detrital samples, by tying it to the two-stage lead evolution model of Stacey and Kramers (1975). This procedure is similar in spirit to the iterative algorithm of Chew et al. (2011), but uses a maximum likelihood approach that weights the uncertainties of all isotopes in the coupled U–Th–Pb decay system. Finally, Section 7 introduces an implementation of the algorithms described herein, using the IsoplotR software package.

2 U–Th–Pb–U–Th–Pb concordia and the Total-Pb/U–Th isochron

110 In conventional ~~U–Pb–U–Pb~~ U–Pb–U–Pb geochronology, the set of concordant $^{206}\text{Pb}/^{238}\text{U}$ - and $^{207}\text{Pb}/^{235}\text{U}$ -ratios defines a ‘Wetherill’ concordia line. Similarly, ~~U–Th–Pb–U–Th–Pb~~ U–Th–Pb–U–Th–Pb data can be visualised in $^{208}\text{Pb}/^{232}\text{Th}$ - vs. $^{206}\text{Pb}/^{238}\text{U}$ -ratio space (~~Figure ??a~~). In the absence of common Pb, samples whose $^{208}\text{Pb}/^{232}\text{Th}$ -ages equal their $^{206}\text{Pb}/^{238}\text{U}$ -ages plot on a U–Th–Pb–U–Th–Pb concordia line. The addition of common Pb pulls samples away from this line. ~~Common Pb correction amounts to moving samples back to concordia:-~~

$$115 \quad \left[\frac{^{208}\text{Pb}}{^{232}\text{Th}} \right]_* = \left[\frac{^{208}\text{Pb}}{^{232}\text{Th}} \right]_m - \frac{^{208}\text{Pb}_c}{^{232}\text{Th}_m}$$

$$\left[\frac{^{206}\text{Pb}}{^{238}\text{U}} \right]_* = \left[\frac{^{206}\text{Pb}}{^{238}\text{U}} \right]_m - \left[\frac{^{206}\text{Pb}}{^{208}\text{Pb}} \right]_c \frac{^{208}\text{Pb}_c}{^{232}\text{Th}_m} \left[\frac{^{232}\text{Th}}{^{238}\text{U}} \right]_m$$

where ‘*’ marks the radiogenic component. ~~Figure ??a~~ shows the effect of this correction on the Gibson et al. (2016) data, using $\left[\frac{^{206}\text{Pb}}{^{208}\text{Pb}} \right]_c = 0.3685$ and a variable $^{208}\text{Pb}_c/^{232}\text{Th}_m$ -ratio calculated by rearranging Equation 4 for $t = 17.66$ Ma. Binary mixing between common Pb and radiogenic Pb forms linear trends in conventional concordia space, but not in U–Th–Pb

120 concordia space. For example the Janots and Rubatto (2014) data plot above or below the concordia line depending on the Th/U-ratio (Figure 1a).

An alternative visualisation is to plot the $^{208}\text{Pb}/^{206}\text{Pb}$ against $^{238}\text{U}/^{206}\text{Pb}$ (green ellipses in Figure ??b,c). Removing the Figure 1b). The radiogenic ^{208}Pb -component and plotting $^{208}\text{Pb}_c/^{206}\text{Pb}$ against can be removed by rearranging Equation 4 for $^{208}\text{Pb}_c/^{232}\text{Th}_m$. Doing this for different values of t moves the various aliquots vertically on the diagram. Each value of t also corresponds to a radiogenic $^{238}\text{U}/^{206}\text{Pb}$ creates a linear isochron array (blue ellipses in Figure ??ratio, thus marking a point on the horizontal axis of the diagram. We can fit a line through this point and minimise the residual scatter of the data around it, using a least squares criterion such as the mean square of the weighted deviates (MSWD, McIntyre et al., 1966). For the Janots and Rubatto (2014) data, the residual scatter is minimised when $t \approx 23$ Ma (Figure 1b). The x-intercept of this line equals the radiogenic U-Pb composition whilst the y-intercept equals its common Pb composition. The same exercise can be repeated for At this value, the aliquots plot along a simple binary mixture between common Pb and radiogenic Pb. This marks the best estimate for the concordia age. The corresponding common-Pb corrected $^{208}\text{Pb}/^{207}\text{Pb} - ^{232}\text{Th} - ^{206}\text{Pb}/^{207}\text{Pb}$ (Figure ??c). The linear fit corresponds to a concordant U-Pb age of 17.66 Ma and a common Pb composition with $[\frac{^{208}\text{Pb}}{^{206}\text{Pb}}]_c = 2.56$ and $[\frac{^{208}\text{Pb}}{^{207}\text{Pb}}]_c = 11.71$. The next section of this paper introduces an algorithm to automatically find this optimal solution and propagate the corresponding uncertainties. U composition is shown as a tight cluster of blue error ellipses on Figure 1a.

135 3 The U-Th-Pb isochron

In order to formalise this procedure in a mathematical sense, let us first define a number of variables. In analogy to the variable names used by Ludwig (1998), we will refer to the blank corrected isotopic ratios as X , Y , Z , W and U :

$$X = \left[\frac{^{207}\text{Pb}}{^{235}\text{U}} \right]_m, Y = \left[\frac{^{206}\text{Pb}}{^{238}\text{U}} \right]_m, Z = \left[\frac{^{208}\text{Pb}}{^{232}\text{Th}} \right]_m, W = \left[\frac{^{232}\text{Th}}{^{238}\text{U}} \right]_m, U = \left[\frac{^{238}\text{U}}{^{235}\text{U}} \right] \quad (8)$$

The U-Th-Pb isochron line is constrained by three free parameters, the age (t) and the common Pb composition (α where X , β) where:-

$$\alpha = \left[\frac{^{206}\text{Pb}}{^{208}\text{Pb}} \right]_c \quad \text{and} \quad \beta = \left[\frac{^{207}\text{Pb}}{^{208}\text{Pb}} \right]_c$$

Y and Z are vectors, W is a diagonal matrix, and U is a scalar; we will use Greek characters for the unknown common Pb ratios:

$$\alpha = \left[\frac{^{206}\text{Pb}}{^{208}\text{Pb}} \right]_c, \beta = \left[\frac{^{207}\text{Pb}}{^{208}\text{Pb}} \right]_c, \gamma = \frac{^{208}\text{Pb}_c}{^{232}\text{Th}_m} \quad (9)$$

145 where α and β are scalars and γ is a vector; and finally, we will use t as the concordia age so that the radiogenic ratios are given by:

$$\left[\frac{^{208}\text{Pb}}{^{232}\text{Th}} \right]_* = e^{\lambda_{32}t} - 1, \left[\frac{^{207}\text{Pb}}{^{235}\text{U}} \right]_* = e^{\lambda_{35}t} - 1, \left[\frac{^{206}\text{Pb}}{^{238}\text{U}} \right]_* = e^{\lambda_{38}t} - 1 \quad (10)$$

Next, we define three misfit vectors K , L and M containing the difference between the measured and the predicted (i.e. common + radiogenic) isotope ratios:

$$150 \quad K = X - U\beta W\gamma - e^{\lambda_{35}t} + 1 \quad (11)$$

$$L = Y - \alpha W\gamma - e^{\lambda_{38}t} + 1 \quad (12)$$

$$M = Z - \gamma - e^{\lambda_{32}t} + 1 \quad (13)$$

This formulation is a straightforward adaptation of Ludwig (1998)'s ^{204}Pb -based Total-Pb/U isochron equations. And like Ludwig (1998), we can then estimate t , α and β can be estimated by minimising the sum of squares:

$$155 \quad S = \Delta \Sigma_{\Delta}^{-1} \Delta^T \quad (14)$$

with-

$$K = X - U\beta W\gamma - e^{\lambda_{35}t} + 1$$

$$L = Y - \alpha W\gamma - e^{\lambda_{38}t} + 1$$

$$M = Z - \gamma - e^{\lambda_{32}t} + 1$$

160 where U is the $^{238}\text{U}/^{235}\text{U}$ -ratio (= 137.818; Hiess et al., 2012),

$$X = \begin{bmatrix} ^{207}\text{Pb} \\ ^{235}\text{U} \end{bmatrix}_m, Y = \begin{bmatrix} ^{206}\text{Pb} \\ ^{238}\text{U} \end{bmatrix}_m \text{ and } Z = \begin{bmatrix} ^{208}\text{Pb} \\ ^{232}\text{Th} \end{bmatrix}_m$$

are n -element column vectors containing the $^{207}\text{Pb}/^{235}\text{U}$ -, $^{206}\text{Pb}/^{238}\text{U}$ -, and $^{208}\text{Pb}/^{232}\text{Th}$ -measurements,

$$W = \begin{bmatrix} ^{232}\text{Th} \\ ^{238}\text{U} \end{bmatrix}_m$$

is an $n \times n$ diagonal matrix with the $^{232}\text{Th}/^{238}\text{U}$ -measurements, and

$$165 \quad \gamma = \frac{^{208}\text{Pb}_c}{^{232}\text{Th}_m}$$

is an n -element column vector with the inherited $^{208}\text{Pb}/^{232}\text{Th}$ -ratios. γ is unknown but can be estimated from the data along with the scalars t . Δ is the amalgamated misfit vector and Δ^T is its transpose (i.e., α - and β -

$\Delta^T = [K^T \ L^T \ M^T]$). Σ_{Δ} is the covariance matrix of the misfit parameters K , L and M . This matrix is obtained Δ , which can be estimated by error propagation of the isotopic measurement and decay constant uncertainties:;

$$170 \quad \Sigma_{\Delta} = \underline{J^T \Sigma J} \begin{bmatrix} J_x & J_{\lambda} \end{bmatrix} \begin{bmatrix} \Sigma_x & 0_{4n \times 4} \\ 0_{4 \times 4n} & \Sigma_{\lambda} \end{bmatrix} \begin{bmatrix} J_x^T \\ J_{\lambda}^T \end{bmatrix} \quad (15)$$

where-

$$\Sigma = \begin{bmatrix} s[X]^2 & s[X,Y] & s[X,Z] & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ s[Y,X] & s[Y]^2 & s[Y,Z] & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ s[Z,X] & s[Z,Y] & s[Z]^2 & 0_{n \times 1} & 0_{n \times 1} & 0_{n \times 1} \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & s[\lambda_{35}]^2 & 0 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0 & s[\lambda_{38}]^2 & 0 \\ 0_{1 \times n} & 0_{1 \times n} & 0_{1 \times n} & 0 & 0 & s[\lambda_{32}]^2 \end{bmatrix}$$

in which-

$$s[\underline{X}]^2 = I_{n \times n} \begin{bmatrix} s[X_1]^2 \\ \vdots \\ s[X_i] \\ \vdots \\ s[X_n]^2 \end{bmatrix}, \quad s[\underline{X}, \underline{Y}] = I_{n \times n} \begin{bmatrix} s[X_1, Y_1] \\ \vdots \\ s[X_i, Y_i] \\ \vdots \\ s[X_n, Y_n] \end{bmatrix}$$

175 and so forth, where $I_{n \times n}$ marks the $n \times n$ identity matrix-

in which Σ_x is the $4n \times 4n$ covariance matrix of the collated data measurements X, Y, Z and W ; Σ_λ is the 4×4 covariance matrix of the decay constants and U , and $s[X_i]^2$, J_x and $s[X_i, Y_i]$ represent the variance and covariance of X_i and Y_i , respectively. The Jacobian matrix J is given by:-

$$J = \begin{bmatrix} 1_{n,n} & 0_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & 1_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & 0_{n \times n} & 1_{n \times n} \\ -t_{1 \times n} e^{\lambda_{35} t} & 0_{1 \times n} & 0_{1 \times n} \\ 0_{1 \times n} & -t_{1 \times n} e^{\lambda_{38} t} & 0_{1 \times n} \\ 0_{1 \times n} & 0_{1 \times n} & -t_{1 \times n} e^{\lambda_{32} t} \end{bmatrix}$$

180 where $1_{a \times b}$, $0_{a \times b}$ and $t_{a \times b}$ are $a \times b$ matrices filled with 1s, 0s and t s. J_λ are Jacobian matrices with partial derivatives of Δ with respect to the isotopic ratio measurements and the decay constants (plus U), respectively. Further details for Σ_x , Σ_λ , J_x and J_λ are provided in Appendix A.

To minimise S with respect to Equation 14 can be solved for t , α and β , we first need to estimate γ for any given value of these free parameters. To this end, we replace γ by $Z - M - e^{\lambda_{32} t} + 1$ in Equation 13, so that:-

$$185 \quad K = \hat{K} + U\beta W M \text{ with } \hat{K} = X - U\beta W(Z - e^{\lambda_{32} t} + 1) - e^{\lambda_{35} t} + 1$$

and

$$L = \hat{L} + \alpha W M \text{ with } \hat{L} = Y - \alpha W(Z - e^{\lambda_{32} t} + 1) - e^{\lambda_{38} t} + 1$$

Plugging Equations 35 and 36 into Equation 14 and rearranging yields:-

$$\underline{S = M^T AM + BM + M^T C + D}$$

190 where-

$$\underline{A = U^2 \beta^2 W_d \Omega_{1,1} W_d + \alpha^2 W_d \Omega_{2,2} W_d + \Omega_{3,3} + U \alpha \beta W_d (\Omega_{1,2} + \Omega_{2,1}) W_d + U \beta (W_d \Omega_{1,3} + \Omega_{3,1} W_d) + \alpha (W_d \Omega_{2,3} + \Omega_{3,2} W_d)}$$

$$\underline{B = U \beta \hat{K}^T \Omega_{1,1} W_d + \alpha \hat{L}^T \Omega_{2,2} W_d + \alpha \hat{K}^T \Omega_{1,2} W_d + U \beta \hat{L}^T \Omega_{2,1} W_d + \hat{K}^T \Omega_{1,3} + \hat{L}^T \Omega_{2,3}}$$

195 $\underline{C = U \beta W_d \Omega_{1,1} \hat{K} + \alpha W_d \Omega_{2,2} \hat{L} + U \beta W_d \Omega_{1,2} \hat{L} + \alpha W_d \Omega_{2,1} \hat{K} + \Omega_{3,1} \hat{K} + \Omega_{3,2} \hat{L}}$

$$\underline{D = \hat{K}^T \Omega_{1,1} \hat{K} + \hat{K}^T \Omega_{1,2} \hat{L} + \hat{L}^T \Omega_{2,1} \hat{K} + \hat{L}^T \Omega_{2,2} \hat{L}}$$

Taking the matrix derivative of β by iterative methods, but the numerical stability of these methods is not guaranteed. Numerical stability and speed of convergence can be greatly improved if we remove the uncertainties of W from the data covariance matrix

200 Σ_x . If the sum of squares S with respect to M :-

$$\underline{\partial S / \partial M = M^T (A + A^T) + B + C^T}$$

Setting $\partial S / \partial M = 0$ and solving for M :-

$$\underline{M = -(A + A^T)^{-1} (B^T + C)}$$

205 Plugging M back into Equation 13 yields our estimate of γ , which allows us to calculate γ does not depend on the uncertainty of W , then the partial derivatives of S with respect to t, α and β that minimise S are then found by numerical methods, γ and t can be calculated manually, which greatly simplifies the optimisation. Further details about this simplified algorithm are provided in Appendix B.

3 Error propagation and overdispersion

The log-likelihood of the isochron fit is given by

210
$$\mathcal{L} = -\frac{1}{2} \left[33n \ln(2\pi) + \ln |\Sigma_\Delta| + S \right] \quad (16)$$

where $|\Sigma_\Delta|$ is marks the determinant of Σ_Δ . The covariance matrix of the three-fit parameters is then obtained by inverting the matrix of second derivatives of the negative log-likelihood with respect to the vector γ and the scalars t, α, β . This is also

known as the Fisher Information matrix):-

$$\begin{bmatrix} \Sigma_\gamma & s[\gamma, t] & s[\gamma, \alpha] & s[\gamma, \beta] \\ s[t, \gamma] & s[t]^2 & s[t, \alpha] & s[t, \beta] \\ s[\alpha, \gamma] & s[\alpha, t] & s[\alpha]^2 & s[\alpha, \beta] \\ s[\beta, \gamma] & s[\beta, t] & s[\beta, \alpha] & s[\beta]^2 \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \gamma^2} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial t \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial t^2} & \frac{\partial^2 \mathcal{L}}{\partial t \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial t \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \beta^2} \end{bmatrix}^{-1} \quad (17)$$

215 where Σ_γ is an $n \times n$ matrix; $s[\gamma, t]$, $s[\gamma, \alpha]$ and $s[\gamma, \beta]$ are n -element row vectors, $s[t, \gamma]$, $s[\alpha, \gamma]$ and $s[\beta, \gamma]$ are n -element column vectors, and all other elements are scalars. The second derivatives are as follows:-

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma^2} = - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial t} = \left(\frac{\partial^2 \mathcal{L}}{\partial t \partial \gamma} \right)^T = - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}$$

220

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \alpha} = \left(\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \gamma} \right)^T = \begin{bmatrix} 0_{n \times n} \\ W \\ 0_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \Delta - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \beta} = \left(\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma} \right)^T = \begin{bmatrix} UW \\ 0_{n \times n} \\ 0_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \Delta - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}$$

225

$$\frac{\partial^2 \mathcal{L}}{\partial t^2} = \Delta^T \Sigma_\Delta^{-1} \begin{bmatrix} (e^{\lambda_{35} t} \lambda_{35}^2)_{n \times 1} \\ (e^{\lambda_{38} t} \lambda_{38}^2)_{n \times 1} \\ (e^{\lambda_{32} t} \lambda_{32}^2)_{n \times 1} \end{bmatrix} - \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial t \partial \alpha} = \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial t} = - \begin{bmatrix} 0_{1 \times n} \\ W\gamma \\ 0_{1 \times n} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial t \partial \beta} = \frac{\partial^2 \mathcal{L}}{\partial \beta \partial t} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_\Delta^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = - \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}$$

$$235 \quad \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} = \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}$$

[given in Appendix C.](#) The Fisher Information matrix is best solved by block matrix inversion. This is achieved by partitioning Equation 17 into four parts, with $\partial^2 \mathcal{L} / \partial \gamma^2$ defining the first block.

4 Overdispersion

If analytical uncertainty is the only source of data scatter around the discordia line, then the sum of squares S follows a central
240 Chi-square distribution with $2n - 3$ degrees of freedom (i.e., χ_{2n-3}^2). Normalising S by the degrees of freedom gives rise to the so-called reduced Chi-square statistic, which is also known as the Mean Square of the Weighted Deviates (MSWD):

$$MSWD = \frac{S}{2n - 3} \quad (18)$$

Datasets are said to be *overdispersed* if S is greater than the 95% percentile of χ_{2n-3}^2 or, equivalently, if $MSWD \gg 1$. ~~This is the case for the Gibson et al. (2016) dataset, whose MSWD = 8.6 (p-value ≈ 0)~~ [Wendt and Carl \(1991\)](#). The overdispersion
245 can [either](#) be attributed to geological scatter in the concordia intercept age t , [or to excess variability in the common Pb ratios \$\alpha\$ and \$\beta\$](#) . Suppose that ~~this the~~ scatter follows a normal distribution with zero mean and let ω be the standard deviation of this distribution. Then we can redefine Σ_{Δ} ~~as:~~ [Equation 15 as:](#)

$$\Sigma_{\Delta} = \underline{J\Sigma J^T} \begin{bmatrix} J_x & J_{\lambda} \end{bmatrix} \begin{bmatrix} \Sigma_x & 0_{4n \times 4} \\ 0_{4 \times 4n} & \Sigma_{\lambda} \end{bmatrix} \begin{bmatrix} J_x^T \\ J_{\lambda}^T \end{bmatrix} + J_{\omega} \omega^2 J_{\omega}^T \quad (19)$$

where J_{ω} , [a the Jacobian matrix with the partial derivatives of \$\Delta\$ w.r.t. to the dispersion parameter \$\omega\$. If the overdispersion is](#)
250 [attributed to diachronous isotopic closure, then:](#)

$$J_{\omega} = \begin{bmatrix} -\lambda_{35} e^{\lambda_{35} t} I_{n \times n} \\ -\lambda_{38} e^{\lambda_{38} t} I_{n \times n} \\ -\lambda_{32} e^{\lambda_{32} t} I_{n \times n} \end{bmatrix} \quad (20)$$

Alternatively, if the overdispersion is attributed to excess scatter of the common Pb ratios, then:

$$J_\omega = \begin{bmatrix} -UW\gamma \\ -W\gamma \\ 0_{n \times 1} \end{bmatrix}$$

255 ω can then be found by plugging Equation 19 into Equation 16 and maximising \mathcal{L} . Like before, the uncertainty of ω is obtained by inverting the Fisher Information, replacing Equation 17 with

$$\begin{bmatrix} \Sigma\gamma & s[\gamma, t] & s[\gamma, \alpha] & s[\gamma, \beta] & s[\gamma, \omega] \\ s[t, \gamma] & s[t]^2 & s[t, \alpha] & s[t, \beta] & s[t, \omega] \\ s[\alpha, \gamma] & s[\alpha, t] & s[\alpha]^2 & s[\alpha, \beta] & s[\alpha, \omega] \\ s[\beta, \gamma] & s[\beta, t] & s[\beta, \alpha] & s[\beta]^2 & s[\beta, \omega] \\ s[\omega, \gamma] & s[\omega, t] & s[\omega, \alpha] & s[\omega, \beta]^2 & s[\omega]^2 \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \gamma^2} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \omega} \\ \frac{\partial^2 \mathcal{L}}{\partial t \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial t^2} & \frac{\partial^2 \mathcal{L}}{\partial t \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial t \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial t \partial \omega} \\ \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \omega} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \beta^2} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \omega} \\ \frac{\partial^2 \mathcal{L}}{\partial \omega \partial \gamma} & \frac{\partial^2 \mathcal{L}}{\partial \omega \partial t} & \frac{\partial^2 \mathcal{L}}{\partial \omega \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \omega \partial \beta^2} & \frac{\partial^2 \mathcal{L}}{\partial \omega^2} \end{bmatrix}^{-1} \quad (21)$$

where-

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \gamma} \right)^T = -\Delta^T \frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}$$

$$260 \quad \frac{\partial^2 \mathcal{L}}{\partial t \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial t} \right)^T = -\Delta^T \frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \alpha} \right)^T = -\Delta^T \frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \beta} \right)^T = -\Delta^T \frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}$$

$$265 \quad \frac{\partial^2 \mathcal{L}}{\partial \omega^2} = -\frac{1}{2} \left(\frac{\partial^2 \ln |\Sigma_\Delta|}{\partial \omega^2} + \Delta^T \frac{\partial^2 \Sigma_\Delta^{-1}}{\partial \omega^2} \Delta \right)$$

with-

$$\frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} = -\Sigma_\Delta^{-1} \frac{\partial \Sigma_\Delta}{\partial \omega} \Sigma_\Delta^{-1}$$

$$270 \quad \frac{\partial^2 \Sigma_{\Delta}^{-1}}{\partial \omega^2} = - \left(\frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \frac{\partial \Sigma_{\Delta}}{\partial \omega} \Sigma_{\Delta}^{-1} + \Sigma_{\Delta}^{-1} \frac{\partial^2 \Sigma_{\Delta}}{\partial \omega^2} \Sigma_{\Delta}^{-1} + \Sigma_{\Delta}^{-1} \frac{\partial \Sigma_{\Delta}}{\partial \omega} \frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \right)$$

$$\frac{\partial^2 \ln |\Sigma_{\Delta}|}{\partial \omega^2} = \text{Tr} \left(\frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \frac{\partial \Sigma_{\Delta}}{\partial \omega} + \Sigma_{\Delta}^{-1} \frac{\partial^2 \Sigma_{\Delta}}{\partial \omega^2} \right)$$

in which $\text{Tr}(\ast)$ stands for the trace of \ast and

$$\frac{\partial \Sigma_{\Delta}}{\partial \omega} = 2 J_{\omega}^T \omega J_{\omega}$$

$$275 \quad \frac{\partial^2 \Sigma_{\Delta}}{\partial \omega^2} = 2 J_{\omega}^T J_{\omega}$$

For the Gibson et al. (2016) dataset, the maximum likelihood estimate of the overdispersion parameter is 0.672 Ma with a standard error

280 In this case, manual calculation of the second derivatives is only possible if the overdispersion is attributed to t , with formulae shown in Appendix D. The second derivatives are not tractable if the excess dispersion is assigned to α and β . In this case the Fisher Information must always be calculated numerically, which can be difficult.

4 Application to literature data

This section applies the U–Th–Pb isochron algorithm to two published datasets, a carbonate dataset of Parrish et al. (2018) and an allanite dataset of Janots and Rubatto (2014). Parrish et al. (2018) investigated the Palaeogene deformation history of southern England by dating calcite veins in chalk and greensand. The measurements were made by quadrupole LA-ICP-MS, for which it was not possible to measure ^{204}Pb with sufficient precision or accuracy. Figure 2a shows the U–Pb data of one particular sample (CB-2, Isle of Wight) on a conventional Tera-Wasserburg diagram. In the absence of ^{204}Pb , conventional data processing would apply a common-Pb correction using the ^{207}Pb -method. That is, it would infer the concordia intercept age by regression of a Semitotal-Pb/U isochron. Doing so suggests a U–Pb age of 29.72 ± 1.23 Ma. However, this isochron exhibits significant overdispersion with respect to the analytical uncertainties ($\text{MSWD} = 3.2$), casting doubt on the accuracy of the date. The fit also suffers from low precision, caused by the large uncertainties of the ^{207}Pb -measurements. These cause the error ellipses of some spots to cross over into negative $^{207}\text{Pb}/^{206}\text{Pb}$ space.

295 The Th/U-ratios of 0.16 Ma CB-2 are extremely low (< 0.12 , as shown on the colour scale of Figure 2). These low ratios are caused by the low solubility of Th in the vein-forming fluids. As a consequence, less than 1% of the measured ^{208}Pb is of radiogenic origin. At the same time, the sample contains between 2 and 20 times more ^{208}Pb than it does ^{207}Pb . This makes the ^{208}Pb -based Total-Pb/U–Th correction far more precise than the conventional ^{207}Pb -based Semitotal-Pb/U correction. Figure 2b shows the Total-Pb/U–Th isochron of CB-2 in $^{208}\text{Pb}_o/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space. The scatter around this line is much tighter than that of the Semitotal-Pb/U fit, and the MSWD is only 2.5, despite the high precision of the added ^{208}Pb

data. The isochron intercept age has dropped to 24.43 ± 0.84 Ma, which is significantly younger than the ^{207}Pb -corrected age estimate. Importantly, the two age estimates do not overlap within the stated uncertainties.

300 It is not possible to formally prove that the ^{208}Pb -corrected age is more accurate than the ^{207}Pb -corrected age for the carbonate dataset. However, an independent assessment of accuracy *is* possible for our second case study. Janots and Rubatto (2014)'s allanite dataset used SIMS instead of LA-ICP-MS, making it possible to compare a ^{204}Pb -based common lead correction with the new ^{208}Pb method. Figure 3a shows the U–Pb data of one particular allanite sample (MF482) on a conventional Tera-Wasserburg concordia diagram, yielding a Semitotal-Pb/U isochron age of 22.77 ± 5.63 Ma. As before, the Th/U-ratios
305 are shown as shades of green to red. These values range from 23 to 235, which is three orders of magnitude higher than Parrish et al. (2018)'s carbonate data. Consequently, most of the chronometric power of the allanite data is contained in the Th–Pb system and not in the U–Pb method. 90 – 97% of the ^{208}Pb is radiogenic, as opposed to 0.3 – 1.0% of the ^{206}Pb , and only 0.06 – 0.016% of the ^{207}Pb .

Figure 3b shows the Th–Pb data in $^{204}\text{Pb}/^{208}\text{Pb} - ^{232}\text{Th}/^{208}\text{Pb}$ space, where they form an isochron with a Th–Pb age of 21.50
310 ± 4.37 Ma. This agrees within error with the ^{207}Pb -corrected U–Pb age, but has a slightly smaller uncertainty and a much lower MSWD (0.74 instead of 1.4). Combining the U–Pb and Th–Pb systems together, Figure 3c shows allanite sample MF482 in $^{208}\text{Pb}_o/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space, where it defines an 23.21 ± 0.85 Ma isochron. This falls within the uncertainties of the U–Pb and Th–Pb age estimates, but is more than five times more precise than the previous age estimates. An alternative visualisation of the Total-Pb/U–Th isochron is shown in Figure 3d. Here, the common-Pb corrected $^{207}\text{Pb}_o/^{208}\text{Pb}$ -ratio is plotted against
315 $^{232}\text{Th}/^{208}\text{Pb}$. Thus, we use the ^{207}Pb as a common-Pb indicator instead of the ^{204}Pb used in Figure 3b. The >15 times greater abundance of ^{207}Pb compared to ^{204}Pb nearly quadruples the precision of the data, producing a tight fit around the isochron.

5 Dealing with skewed error distributions

All the free parameters in the regression algorithm (t , α , β and ω) are strictly positive quantities. This positivity constraint manifests itself in skewed error distributions. For example, when the four parameter algorithm of Section ??-3 is applied to
320 datasets that exhibit little or no overdispersion ($\omega \approx 0$), then the usual '2-sigma' error bounds can cross over into physically impossible negative data space. This section of the paper introduces two ways to deal with this problem.

A first solution is to obtain asymmetric uncertainty bounds for ω using a profile likelihood approach (Galbraith, 2005; Vermeesch, 2018). First, maximise Equation 16 for the four parameters t , α , β and ω . Denote the corresponding log-likelihood value by \mathcal{L}_m . Second, consider a range of values for ω around the maximum likelihood estimate. For each of these values,
325 maximise \mathcal{L} for t , α and β whilst keeping ω fixed. Denote the corresponding log-likelihood by \mathcal{L}_ω . Finally, a 95% confidence region for ω is obtained by collecting all the values of ω for which $\mathcal{L}_\omega > \mathcal{L}_m - 3.85/2$, where 3.85 corresponds to the 95th percentile of a chi-square distribution with one degree of freedom \mp .

(Figure 4 illustrates the profile likelihood method using the Gibson et al. (2016) dataset). The same procedure can also be applied to t , α and β , in order to obtain asymmetric confidence intervals for those parameters if needed. This would be particularly
330 useful for very young samples.

A second and more pragmatic approach to dealing with the positivity constraint is to simply redefine the regression parameters in terms of logarithmic quantities. This is done by replacing Equations 11, 12 and 19 with:

$$K = X - U \exp[\beta_*] W \gamma - \exp[\lambda_{35} e^{t_*}] + 1 \quad (22)$$

$$L = Y - \exp[\alpha_*] W \gamma - \exp[\lambda_{38} e^{t_*}] + 1 \quad (23)$$

$$335 \quad \Sigma_{\Delta} = \underline{J \Sigma J^T} \begin{bmatrix} J_x & J_{\lambda} \end{bmatrix} \begin{bmatrix} \Sigma_x & 0_{4n \times 4} \\ 0_{4 \times 4n} & \Sigma_{\lambda} \end{bmatrix} \begin{bmatrix} J_x^T \\ J_{\lambda}^T \end{bmatrix} + J_{\omega} \exp[\omega_*]^2 J_{\omega}^T \quad (24)$$

respectively, and maximising Equation 16 with respect to t_* , α_* , β_* and ω_* . The standard errors for these log parameters (again obtained from the Fisher Information matrix) can then be converted to asymmetric confidence intervals for t , α , β and ω .

~~Applying this approach to the Gibson et al. (2016) dataset~~ This approach yields results that are similar to those obtained using the profile log-likelihood method (as illustrated in Figure 4) ~~for monazite grain #10 in sample BHE-01 of Gibson et al. (2016)~~

340 . This sample experienced a diachronous crystallisation history, resulting in an overdispersed Total-Pb/U–Th isochron fit (MSWD = 8). Quantifying the excess dispersion with a model-3 fit yields an overdispersion parameter $\omega = 0.67$ Ma with asymmetric confidence bounds of +0.48/–0.23 Ma. Besides generating realistic confidence regions, the logarithmic reparameterisation of the likelihood function has the added benefit increasing the numerical stability of the maximum likelihood method.

6 Detrital samples

345 So far we have assumed that all the ~~U–Th–Pb~~ U–Th–Pb measurements are cogenetic and share the same common Pb composition. This assumption is generally not valid for detrital minerals, which tend to contain a mixture of provenance components. In this case the different crystals in a sample are not expected to plot along a single isochron line. However it is still possible to remove the common Pb component by making certain assumptions about the common Pb composition. One way to do this is to assume that the mineral of interest was extracted from a reservoir of known ~~U–Th–Pb composition~~ U–Th–Pb composition.

350 For example, using the two-stage lead evolution model of Stacey and Kramers (1975), it is possible to predict the $^{206}\text{Pb}/^{208}\text{Pb}$ and $^{207}\text{Pb}/^{208}\text{Pb}$ ratios of the reservoir for any given time t . ~~In other words, any given value of the concordia intersection age determines parameters α and β of Equation ??.~~ More specifically, if $t < 3.7$ Ga, then

$$\alpha(t) = \frac{\left[\frac{^{206}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} + \left[\frac{^{238}\text{U}}{^{204}\text{Pb}} \right]_{sk} (e^{\lambda_{38} 3.7} - e^{\lambda_{38} t})}{\left[\frac{^{208}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} + \left[\frac{^{232}\text{Th}}{^{204}\text{Pb}} \right]_{sk} (e^{\lambda_{32} 3.7} - e^{\lambda_{32} t})} \quad (25)$$

$$\beta(t) = \frac{\left[\frac{^{207}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} + \frac{1}{U} \left[\frac{^{238}\text{U}}{^{204}\text{Pb}} \right]_{sk} (e^{\lambda_{35} 3.7} - e^{\lambda_{35} t})}{\left[\frac{^{208}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} + \left[\frac{^{232}\text{Th}}{^{204}\text{Pb}} \right]_{sk} (e^{\lambda_{32} 3.7} - e^{\lambda_{32} t})} \quad (26)$$

355 where ~~$\left[\frac{^{206}\text{Pb}/^{204}\text{Pb}}{3.7} \right] = 11.152$, $\left[\frac{^{208}\text{Pb}/^{204}\text{Pb}}{3.7} \right] = 31.23$, $\left[\frac{^{207}\text{Pb}/^{204}\text{Pb}}{3.7} \right] = 12.998$, $\left[\frac{^{238}\text{U}/^{204}\text{Pb}}{sk} \right] = 9.74$, and $\left[\frac{^{232}\text{Th}/^{204}\text{Pb}}{sk} \right] = 36.84$.~~ $\left[\frac{^{208}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} = 31.23$, $\left[\frac{^{207}\text{Pb}}{^{204}\text{Pb}} \right]_{3.7} = 12.998$, $\left[\frac{^{238}\text{U}}{^{204}\text{Pb}} \right]_{sk} = 9.74$, and $\left[\frac{^{232}\text{Th}}{^{204}\text{Pb}} \right]_{sk} = 36.84$. Substituting $\alpha(t)$ and $\beta(t)$ for α and β in Equations 11–13 reduces the number of free parameters from three (α , β and t) to one (t). This provides a quick and numeri-

cally robust mechanism for common-Pb correction of detrital minerals. [It is the maximum likelihood equivalent of the heuristic approach used by Chew et al. \(2011\).](#)

360 7 Implementation in IsoplotR

The algorithms presented in this paper have been implemented in the IsoplotR software toolbox for geochronology (Vermeesch, 2018). The easiest way to use the [U-Th-Pb-U-Th-Pb](#) isochron functions is via an online graphical user interface at <http://isoplotr.london-geochron.com> [http:// isoplotr. london-geochron. com](http://isoplotr.london-geochron.com). Alternatively, the same functions can also be accessed from the command line, using the R programming language ([R Core Team, 2020](#)). This section of the paper presents some code snippets to illustrate the key functions involved. This brief tutorial assumes that the reader has R and IsoplotR installed on her/his computer. Further details about this are provided by Vermeesch (2018), and on the aforementioned website. First, we need to load IsoplotR into R:

```
library(IsoplotR)
```

Two new data formats have been added to IsoplotR's existing six [U-Pb-U-Pb](#) formats, to accommodate datasets comprising ^{232}Th and ^{208}Pb . ~~The Gibson et al. (2016) dataset~~ [Sample Ga2 of Janots and Rubatto \(2014\)](#) has been included in the IsoplotR package as two data files (`UPb7.csv` and `UPb8.csv`), ~~which are also available in the supplementary information of this paper.~~

`UPb7.csv` specifies the [U-Th-Pb-U-Th-Pb](#) composition using the 'Wetherill' ratios $^{207}\text{Pb}/^{235}\text{U}$, $^{206}\text{Pb}/^{238}\text{U}$, $^{208}\text{Pb}/^{232}\text{Th}$ and $^{232}\text{Th}/^{238}\text{U}$, whereas `UPb8.csv` uses the 'Tera-Wasserburg' ratios $^{238}\text{U}/^{206}\text{Pb}$, $^{207}\text{Pb}/^{206}\text{Pb}$, $^{208}\text{Pb}/^{206}\text{Pb}$ and $^{232}\text{Th}/^{238}\text{U}$. ~~Both data formats require that the analytical uncertainties and error correlations of all the ratios are specified. The key difference between the two formats is the strength of the internal error correlations, which is greater for format 7 than it is for format 8.~~ The following commands load the contents of [UPb7](#) `UPb8.csv` into a variable called `UPb`, and plot the data on a $^{208}\text{Pb}/^{232}\text{Th}$ vs. $^{206}\text{Pb}/^{238}\text{U}$ -concordia diagram:

```
380 UPb <- read.data('UPb8.csv', method='U-Pb', format=8)
concordia(UPb, type=3)
```

Performing a discordia regression and visualising the results as a $^{208}\text{Pb}_c/^{206}\text{Pb}$ vs. $^{238}\text{U}/^{206}\text{Pb}$ isochron:

```
isochron(UPb, type=1)
```

which performs a three parameter regression without overdispersion. Accounting for overdispersion is done using the optional model argument:

```
fit <- isochron(UPb, type=1, model=3)
```

where `fit` is a variable that stores the numerical results of the isochron regression. This is a list of items that can be inspected by typing `fit` at the R command prompt. For example, the maximum likelihood estimates for t , α , β and ω are stored in

390 `fit$par` and the covariance matrix in `fit$cov`. Changing `type` to 2 plots the regression results as a $^{208}\text{Pb}_c/^{207}\text{Pb}$ vs. $^{235}\text{U}/^{207}\text{Pb}$ isochron. The isochron results can also be visualised on the concordia diagram:

```
concordia (UPb, type=2, show.age=2)
```

where `type=2` produces a Tera-Wasserburg diagram and the `show.age` argument adds a three-parameter regression line to it. Change this to `show.age=4` for a four-parameter fit.

8 Discussion and future developments

395 This paper introduced a ~~new algorithm~~ 'Total-Pb/U-Th algorithm' for common Pb correction by joint regression of all Pb isotopes of U and Th. ~~The algorithm was successfully applied to a monazite dataset by Gibson et al. (2016). With a Th/U-ratio of ~ 10 , the Gibson et al. (2016) sample represents a 'worst case scenario', because the presence of significant amounts of radiogenic ^{208}Pb complicates the~~ For samples that are low in Th (such as carbonates), ^{208}Pb offers the most precise way to correct for common lead, because ^{208}Pb tends to be more abundant than both $^{208}\text{204}\text{Pb}$ -based common Pb correction. The fact
400 ~~that the Gibson et al. (2016) test case is successful holds great promise for the application of the new algorithm to Th-poor materials such as carbonates. Pb and ^{207}Pb . For samples that are high in Th, the $^{208}\text{Pb}/^{232}\text{Th}$ clock adds chronometrically valuable information to the joint U-Pb decay systems.~~

The ingrowth of radiogenic Pb described by Equations 2–4 assumes initial secular equilibrium between all the intermediate daughters in the ~~U-Th-Pb~~ U-Th-Pb decay chains. The new discordia regression algorithm can be modified to accommodate
405 departures from this assumption. In fact, such disequilibrium corrections have already been implemented in `IsoplotR`, using the matrix derivative approach of McLean et al. (2016). A manuscript detailing these calculation is in preparation by the latter author. The disequilibrium correction is particularly useful for applications to young carbonates, whose initial $^{234}\text{U}/^{238}\text{U}$ and $^{230}\text{Th}/^{238}\text{U}$ activity ratios may be far out of equilibrium.

The new discordia regression algorithm is based on the method of maximum likelihood, and accounts for correlated uncer-
410 tainties between variables. ~~For example, the analytical uncertainties of the $^{208}\text{Pb}/^{232}\text{Th}$ and $^{232}\text{Th}/^{238}\text{U}$ ratios in the Gibson et al. (2016) dataset are characterised by correlation coefficients of ca. -0.6 . However geochronological datasets are often associated with equally significant error correlations *between samples* (Vermeesch, 2015). The algorithm presented in this paper easily handles such correlations, which carry *systematic uncertainty*. It suffices to replace the zero values in the upper left $[3n \times 3n]$ sub-matrix of Equation ?? with non-zero values. Unfortunately, there currently exist no algorithms that keep track of inter-sample error
415 ~~correlations in the context of U-Pb geochronology. Doing so~~ These are represented by the off-diagonal terms of the covariance matrix Σ_x in Equation 15. However, to use this option in practical applications will require a new generation of low level data processing software.~~

This new generation software will also need to deal with a second issue that negatively affects the accuracy of the U(-Th)-Pb method, which is apparent from ~~Figures ??b and ??e~~ Figure 1. After removing the radiogenic ^{208}Pb -component from
420 the ~~Gibson et al. (2016) Janots and Rubatto (2014)~~ dataset, the 95% confidence ellipse of one of the aliquots crosses over into negative $^{208}\text{Pb}_c/^{206}\text{232}\text{Pb}$ and $^{208}\text{Pb}_c/^{207}\text{Pb-Th}$ ratios. This nonsensical result is related to the issues discussed in Section 5.

Isotopic data are strictly positive quantities that exhibit skewed error distributions. ‘Normal’ statistical operations such as averaging and the calculation of ‘2-sigma’ confidence intervals can produce counter-intuitive results when applied to such data.

425 In Section 5, the skewness of the fit parameters was removed by reformulating the regression algorithm in terms of logarithmic quantities. Similarly, Vermeesch (2015) showed that the skewness of isotopic compositions can be removed using log-ratios, in the context of $^{40}\text{Ar}/^{39}\text{Ar}$ geochronology. McLean et al. (2016) introduced the same approach to in-situ ~~U-Pb~~ U-Pb geochronology by LA-ICP-MS. Future software development will allow analysts to export their ~~U-Th-Pb~~ U-Th-Pb isotopic data directly as logratios and covariance matrices. Such a data structure can still be analysed with the new discordia
430 regression algorithm, after a logarithmic change a variables for X , Y , Z and W in Equations 11, 12 and 13.

The ~~U-Pb~~ U-Pb method is one of the most powerful and versatile methods in the geochronological toolbox. With two isotopes of the same parent (^{235}U and ^{238}U) decaying to two different isotopes of the same daughter (^{207}Pb and ^{206}Pb), the ~~U-Pb~~ U-Pb method offers an internal quality control that is absent from most other geochronological techniques. U-bearing minerals often contain significant amounts of Th, which decays to ^{208}Pb . However until this day geochronologists have not
435 frequently used this additional parent-daughter pair to its full potential. It is hoped that the algorithm and software presented in this paper will change this situation.

Appendix A

The covariance matrix of the isotopic ratio measurements X , Y , Z and W is given by:

$$\Sigma_x = \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} & \Sigma_{X,Z} & \Sigma_{X,W} \\ \Sigma_{Y,X} & \Sigma_Y & \Sigma_{Y,Z} & \Sigma_{Y,W} \\ \Sigma_{Z,X} & \Sigma_{Z,Y} & \Sigma_Z & \Sigma_{Z,W} \\ \Sigma_{W,X} & \Sigma_{W,Y} & \Sigma_{W,Z} & \Sigma_W \end{bmatrix} \quad (27)$$

440 where

$$\Sigma_X = \begin{bmatrix} s[X_1]^2 & s[X_1, X_2] & \dots & s[X_1, X_n] \\ s[X_2, X_1] & s[X_2]^2 & \dots & s[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ s[X_n, X_1] & s[X_n, X_2] & \dots & s[X_n]^2 \end{bmatrix}, \quad (28)$$

$$\Sigma_{X,Y} = \begin{bmatrix} s[X_1, Y_1]^2 & s[X_1, Y_2] & \dots & s[X_1, Y_n] \\ s[X_2, Y_1] & s[X_2, Y_2] & \dots & s[X_2, Y_n] \\ \vdots & \vdots & \ddots & \vdots \\ s[X_n, Y_1] & s[X_n, Y_2] & \dots & s[X_n, Y_n] \end{bmatrix}, \quad (29)$$

and so forth, in which $s[a]^2$ is the variance of a and $s[a, b]$ is the covariance of a and b , for any a and b . Σ_λ is the covariance matrix of the decay constants and the $^{238}\text{U}/^{235}\text{U}$ -ratio:

$$445 \quad \Sigma_\lambda = \begin{bmatrix} s[\lambda_{35}]^2 & 0 & 0 & 0 \\ 0 & s[\lambda_{38}]^2 & 0 & 0 \\ 0 & 0 & s[\lambda_{32}]^2 & 0 \\ 0 & 0 & 0 & s[U]^2 \end{bmatrix}, \quad (30)$$

Here the covariance terms have been set to zero, but nonzero values could also be accommodated. Finally, the Jacobian matrices J_x and J_λ are given by:

$$J_x = \begin{bmatrix} I_{n \times n} & 0_{n \times n} & 0_{n \times n} & -U\beta I_{n \times n}\gamma \\ 0_{n \times n} & I_{n \times n} & 0_{n \times n} & -\alpha I_{n \times n}\gamma \\ 0_{n \times n} & 0_{n \times n} & I_{n \times n} & 0_{n \times n} \end{bmatrix} \quad (31)$$

and

$$450 \quad J_\lambda = \begin{bmatrix} -t_{n \times 1}e^{\lambda_{35}t} & 0_{n \times 1} & 0_{n \times 1} & -\beta W\gamma \\ 0_{n \times 1} & -t_{n \times 1}e^{\lambda_{38}t} & 0_{n \times 1} & 0_{n \times 1} \\ 0_{n \times 1} & 0_{n \times 1} & -t_{n \times 1}e^{\lambda_{32}t} & 0_{n \times 1} \end{bmatrix} \quad (32)$$

where $0_{a \times b}$ and $1_{a \times b}$ mark a $a \times b$ matrices of zeros and ones, respectively, whereas $I_{n \times n}$ is the $n \times n$ identity matrix.

Appendix B

The numerical stability of the optimisation is greatly enhanced by dropping the dependency of the sum of squares S on the uncertainty of the Th/U ratios W . Thus, we replace Equation 27 by

$$455 \quad \Sigma_x = \begin{bmatrix} \Sigma_X & \Sigma_{X,Y} & \Sigma_{X,Z} \\ \Sigma_{Y,X} & \Sigma_Y & \Sigma_{Y,Z} \\ \Sigma_{Z,X} & \Sigma_{Z,Y} & \Sigma_Z \end{bmatrix} \quad (33)$$

and Equation 31 by the $3n \times 3n$ identity matrix (i.e., $J_x = I_{3n \times 3n}$). Let us define Ω to be the inverse covariance matrix of Δ , so that

$$\Sigma_{\Delta}^{-1} \equiv \Omega = \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} & \Omega_{1,3} \\ \Omega_{2,1} & \Omega_{2,2} & \Omega_{2,3} \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{bmatrix} \quad (34)$$

460 Then, we can directly estimate γ for any given value of t , α and β , by replacing γ with $Z - M - e^{\lambda_{32}t} + 1$ in Equation 13, so that:

$$K = \hat{K} + U\beta W M \text{ with } \hat{K} = X - U\beta W(Z - e^{\lambda_{32}t} + 1) - e^{\lambda_{35}t} + 1 \quad (35)$$

and

$$L = \hat{L} + \alpha W M \text{ with } \hat{L} = Y - \alpha W(Z - e^{\lambda_{32}t} + 1) - e^{\lambda_{38}t} + 1 \quad (36)$$

Plugging Equations 35 and 36 into Equation 14 and rearranging yields:

$$465 \quad S = M^T A M + B M + M^T C + D \quad (37)$$

where

$$A = U^2 \beta^2 W_d \Omega_{1,1} W_d + \alpha^2 W_d \Omega_{2,2} W_d + \Omega_{3,3} + U \alpha \beta W_d (\Omega_{1,2} + \Omega_{2,1}) W_d + U \beta (W_d \Omega_{1,3} + \Omega_{3,1} W_d) + \alpha (W_d \Omega_{2,3} + \Omega_{3,2} W_d) \quad (38)$$

$$B = U \beta \hat{K}^T \Omega_{1,1} W_d + \alpha \hat{L}^T \Omega_{2,2} W_d + \alpha \hat{K}^T \Omega_{1,2} W_d + U \beta \hat{L}^T \Omega_{2,1} W_d + \hat{K}^T \Omega_{1,3} + \hat{L}^T \Omega_{2,3} \quad (39)$$

470

$$C = U \beta W_d \Omega_{1,1} \hat{K} + \alpha W_d \Omega_{2,2} \hat{L} + U \beta W_d \Omega_{1,2} \hat{L} + \alpha W_d \Omega_{2,1} \hat{K} + \Omega_{3,1} \hat{K} + \Omega_{3,2} \hat{L} \quad (40)$$

$$D = \hat{K}^T \Omega_{1,1} \hat{K} + \hat{K}^T \Omega_{1,2} \hat{L} + \hat{L}^T \Omega_{2,1} \hat{K} + \hat{L}^T \Omega_{2,2} \hat{L} \quad (41)$$

Taking the matrix derivative of S with respect to M :

$$475 \quad \partial S / \partial M = M^T (A + A^T) + B + C^T \quad (42)$$

Setting $\partial S / \partial M = 0$ and solving for M :

$$M = -(A + A^T)^{-1} (B^T + C) \quad (43)$$

Plugging M back into Equation 13 yields our estimate of γ , which allows us to calculate S . The values of t , α and β that minimise S are then found by numerical methods.

Explicit formulae for the Fisher Information matrix (Equation 17) are possible for the simplified algorithm, in which the uncertainty of W is ignored:

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma^2} = - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix} \quad (44)$$

$$485 \quad \frac{\partial^2 \mathcal{L}}{\partial \gamma \partial t} = \left(\frac{\partial^2 \mathcal{L}}{\partial t \partial \gamma} \right)^T = - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix} \quad (45)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \alpha} = \left(\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \gamma} \right)^T = \begin{bmatrix} 0_{n \times n} \\ W \\ 0_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \Delta - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix} \quad (46)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \beta} = \left(\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \gamma} \right)^T = \begin{bmatrix} UW \\ 0_{n \times n} \\ 0_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \Delta - \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix} \quad (47)$$

490

$$\frac{\partial^2 \mathcal{L}}{\partial t^2} = \Delta^T \Sigma_{\Delta}^{-1} \begin{bmatrix} (e^{\lambda_{35} t} \lambda_{35}^2)_{n \times 1} \\ (e^{\lambda_{38} t} \lambda_{38}^2)_{n \times 1} \\ (e^{\lambda_{32} t} \lambda_{32}^2)_{n \times 1} \end{bmatrix} - \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix} \quad (48)$$

$$\frac{\partial^2 \mathcal{L}}{\partial t \partial \alpha} = \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial t} = - \begin{bmatrix} 0_{1 \times n} \\ W\gamma \\ 0_{1 \times n} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix} \quad (49)$$

$$495 \quad \frac{\partial^2 \mathcal{L}}{\partial t \partial \beta} = \frac{\partial^2 \mathcal{L}}{\partial \beta \partial t} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix} \quad (50)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = - \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix} \quad (51)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix} \quad (52)$$

500

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} = \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha} = - \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix}^T \Sigma_{\Delta}^{-1} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix} \quad (53)$$

Appendix D

Additional derivatives are required to propagate the uncertainty of the overdispersion parameters ω . This can only be done manually if the overdispersion is attributed to the concordia intercept age t , using the simplified model (ignoring the uncertainty of W). In that case

505

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \gamma} \right)^T = -\Delta^T \frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \begin{bmatrix} U\beta W \\ \alpha W \\ I_{n \times n} \end{bmatrix} \quad (54)$$

$$\frac{\partial^2 \mathcal{L}}{\partial t \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial t} \right)^T = -\Delta^T \frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \begin{bmatrix} (\lambda_{35} e^{\lambda_{35} t})_{n \times 1} \\ (\lambda_{38} e^{\lambda_{38} t})_{n \times 1} \\ (\lambda_{32} e^{\lambda_{32} t})_{n \times 1} \end{bmatrix} \quad (55)$$

510

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \alpha} \right)^T = -\Delta^T \frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \begin{bmatrix} 0_{n \times 1} \\ W\gamma \\ 0_{n \times 1} \end{bmatrix} \quad (56)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \omega} = \left(\frac{\partial^2 \mathcal{L}}{\partial \omega \partial \beta} \right)^T = -\Delta^T \frac{\partial \Sigma_{\Delta}^{-1}}{\partial \omega} \begin{bmatrix} UW\gamma \\ 0_{n \times 1} \\ 0_{n \times 1} \end{bmatrix} \quad (57)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \omega^2} = -\frac{1}{2} \left(\frac{\partial^2 \ln |\Sigma_\Delta|}{\partial \omega^2} + \Delta^T \frac{\partial^2 \Sigma_\Delta^{-1}}{\partial \omega^2} \Delta \right) \quad (58)$$

515 with

$$\frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} = -\Sigma_\Delta^{-1} \frac{\partial \Sigma_\Delta}{\partial \omega} \Sigma_\Delta^{-1} \quad (59)$$

$$\frac{\partial^2 \Sigma_\Delta^{-1}}{\partial \omega^2} = - \left(\frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \frac{\partial \Sigma_\Delta}{\partial \omega} \Sigma_\Delta^{-1} + \Sigma_\Delta^{-1} \frac{\partial^2 \Sigma_\Delta}{\partial \omega^2} \Sigma_\Delta^{-1} + \Sigma_\Delta^{-1} \frac{\partial \Sigma_\Delta}{\partial \omega} \frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \right) \quad (60)$$

$$520 \frac{\partial^2 \ln |\Sigma_\Delta|}{\partial \omega^2} = \text{Tr} \left(\frac{\partial \Sigma_\Delta^{-1}}{\partial \omega} \frac{\partial \Sigma_\Delta}{\partial \omega} + \Sigma_\Delta^{-1} \frac{\partial^2 \Sigma_\Delta}{\partial \omega^2} \right) \quad (61)$$

in which $\text{Tr}(\cdot)$ stands for the trace of \cdot and

$$\frac{\partial \Sigma_\Delta}{\partial \omega} = 2J_\omega^T \omega J_\omega \quad (62)$$

$$\frac{\partial^2 \Sigma_\Delta}{\partial \omega^2} = 2J_\omega^T J_\omega \quad (63)$$

525 Explicit formulae for the second derivatives are not available for the common-Pb based overdispersion model. In that case, the Fisher Information matrix must be computed numerically.

Data availability. Datasets `UPb7.csv` and `UPb8.csv` are included in the supplementary information.

Author contributions. As the sole author of this paper, PV did all the work and wrote the entire paper.

Competing interests. PV does not have any competing interests.

530 *Acknowledgements.* PV would like to thank Randy Parrish, Phil Hopley and John Cottle for stimulating discussions that prompted the writing of this paper; and Ryan Ickert and Kyle Samperton for careful reviews.

References

- Andersen, T.: Correction of common lead in U–Pb analyses that do not report ^{204}Pb , *Chemical Geology*, 192, 59–79, 2002.
- Chew, D. M., Sylvester, P. J., and Tubrett, M. N.: U–Pb and Th–Pb dating of apatite by LA-ICPMS, *Chemical Geology*, 280, 200–216, 2011.
- Galbraith, R. F.: *Statistics for fission track analysis*, CRC Press, 2005.
- 535 Gibson, R., Godin, L., Kellett, D. A., Cottle, J. M., and Archibald, D.: Diachronous deformation along the base of the Himalayan metamorphic core, west-central Nepal, *Geological Society of America Bulletin*, 128, 860–878, 2016.
- Hiess, J., Condon, D. J., McLean, N., and Noble, S. R.: $^{238}\text{U}/^{235}\text{U}$ systematics in terrestrial uranium-bearing minerals, *Science*, 335, 1610–1614, 2012.
- Janots, E. and Rubatto, D.: U–Th–Pb dating of collision in the external Alpine domains (Urseren zone, Switzerland) using low temperature
540 allanite and monazite, *Lithos*, 184, 155–166, 2014.
- Ludwig, K. R.: On the treatment of concordant uranium-lead ages, *Geochimica et Cosmochimica Acta*, 62, 665–676, [https://doi.org/10.1016/S0016-7037\(98\)00059-3](https://doi.org/10.1016/S0016-7037(98)00059-3), 1998.
- McIntyre, G. A., Brooks, C., Compston, W., and Turek, A.: The Statistical Assessment of Rb–Sr Isochrons, *Journal of Geophysical Research*, 71, 5459–5468, 1966.
- 545 McLean, N. M., Bowring, J. F., and Gehrels, G.: Algorithms and software for U–Pb geochronology by LA-ICPMS, *Geochemistry, Geophysics, Geosystems*, 17, 2480–2496, 2016.
- McLean, N. M., Smith, C. J. M., Roberts, N. M. W., and Richards, D. A.: Connecting the U–Th and U–Pb Chronometers: New Algorithms and Applications, AGU Fall Meeting Abstracts, 2016.
- Parrish, R. R., Parrish, C. M., and Lasalle, S.: Vein calcite dating reveals Pyrenean orogen as cause of Paleogene deformation in southern
550 England, *Journal of the Geological Society*, 175, 425–442, 2018.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2020.
- Stacey, J. and Kramers, J.: Approximation of terrestrial lead isotope evolution by a two-stage model, *Earth and Planetary Science Letters*, 26, 207–221, 1975.
- 555 Vermeesch, P.: Revised error propagation of $^{40}\text{Ar}/^{39}\text{Ar}$ data, including covariances, *Geochimica et Cosmochimica Acta*, 171, 325–337, 2015.
- Vermeesch, P.: *IsoplotR*: a free and open toolbox for geochronology, *Geoscience Frontiers*, 9, 1479–1493, 2018.
- Wendt, I. and Carl, C.: The statistical distribution of the mean squared weighted deviation, *Chemical Geology: Isotope Geoscience Section*, 86, 275–285, 1991.
- Williams, I. S.: U–Th–Pb geochronology by ion microprobe, *Reviews in Economic Geology*, 7, 1–35, 1998.

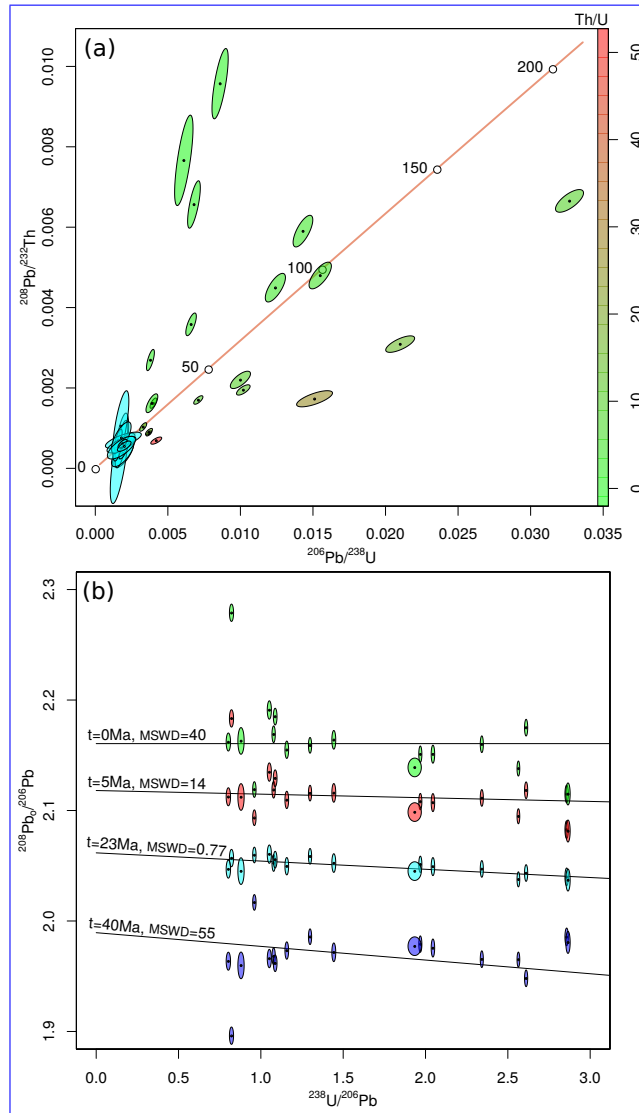


Figure 1. U-Th-Pb data for allanite sample MF482 of Janots and Rubatto (2014) shown on (a) U-Th-Pb concordia diagram showing $^{206}\text{Pb}/^{238}\text{U}$ and (b) a $^{208}\text{Pb}/^{206}\text{Pb}$ - $^{238}\text{U}/^{206}\text{Pb}$ and $^{208}\text{Pb}/^{232}\text{Th}$ ratios for monazite grain #10 in sample BHE-01 of Gibson et al. (2016) ^{206}Pb isochron plot. Uncorrected ratios are shown in shades of green and common Pb-corrected ratios to red on the concordia diagram, in blue. b. $^{238}\text{U}/^{232}\text{Th}$, $^{206}\text{Pb}/^{208}\text{Pb}$ and c. $^{235}\text{U}/^{232}\text{Th}$, $^{207}\text{Pb}/^{208}\text{Pb}$ isochron plots for proportion to the Gibson et al. (2016) dataset Th/U ratio. Green ellipses represent the raw data are shown as green ellipses on the isochron diagram. Blue, Red, light and dark blue ellipses show the same measurements with the 5, 23 and 40 Ma worth of radiogenic ^{208}Pb removed, respectively. The x-intercept misfit of the two resulting isochrons yield the radiogenic U-Pb composition, and correspond to an age ^{208}Pb -corrected data around the best fit line is expressed as weighted square of 17.66 Mamean deviates (MSWD, McIntyre et al., 1966) values. The y-intercepts yield the common Pb isotopic composition, with $[\text{Pb}/^{206}\text{Pb}]_c = 2.71$ and $[\text{Pb}/^{207}\text{Pb}]_c = 12.41$. Error ellipses are shown at 1σ .

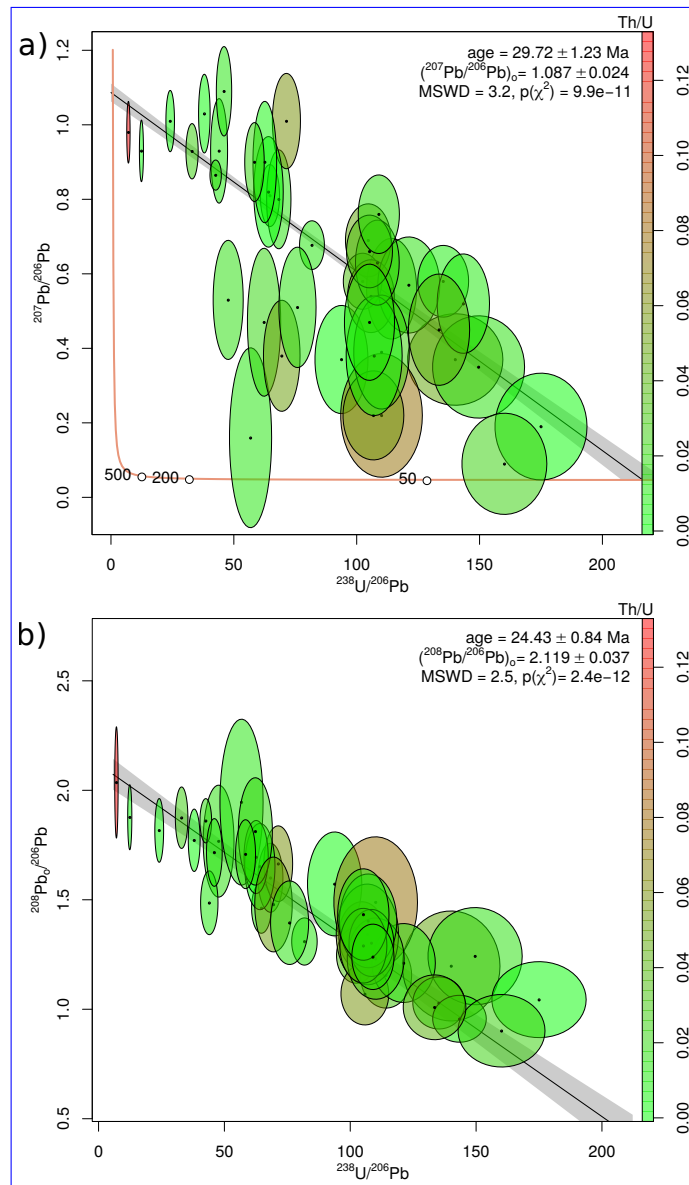


Figure 2. a) SemiTotal-Pb/U isochron (^{207}Pb -based common Pb correction) for Parrish et al. (2018)' chalk data; b) Total-Pb/U–Th isochron (^{208}Pb -based common Pb correction) shown in $^{208}\text{Pb}_c/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space. Colours indicate the Th/U-ratio. All uncertainties are shown at 1σ .

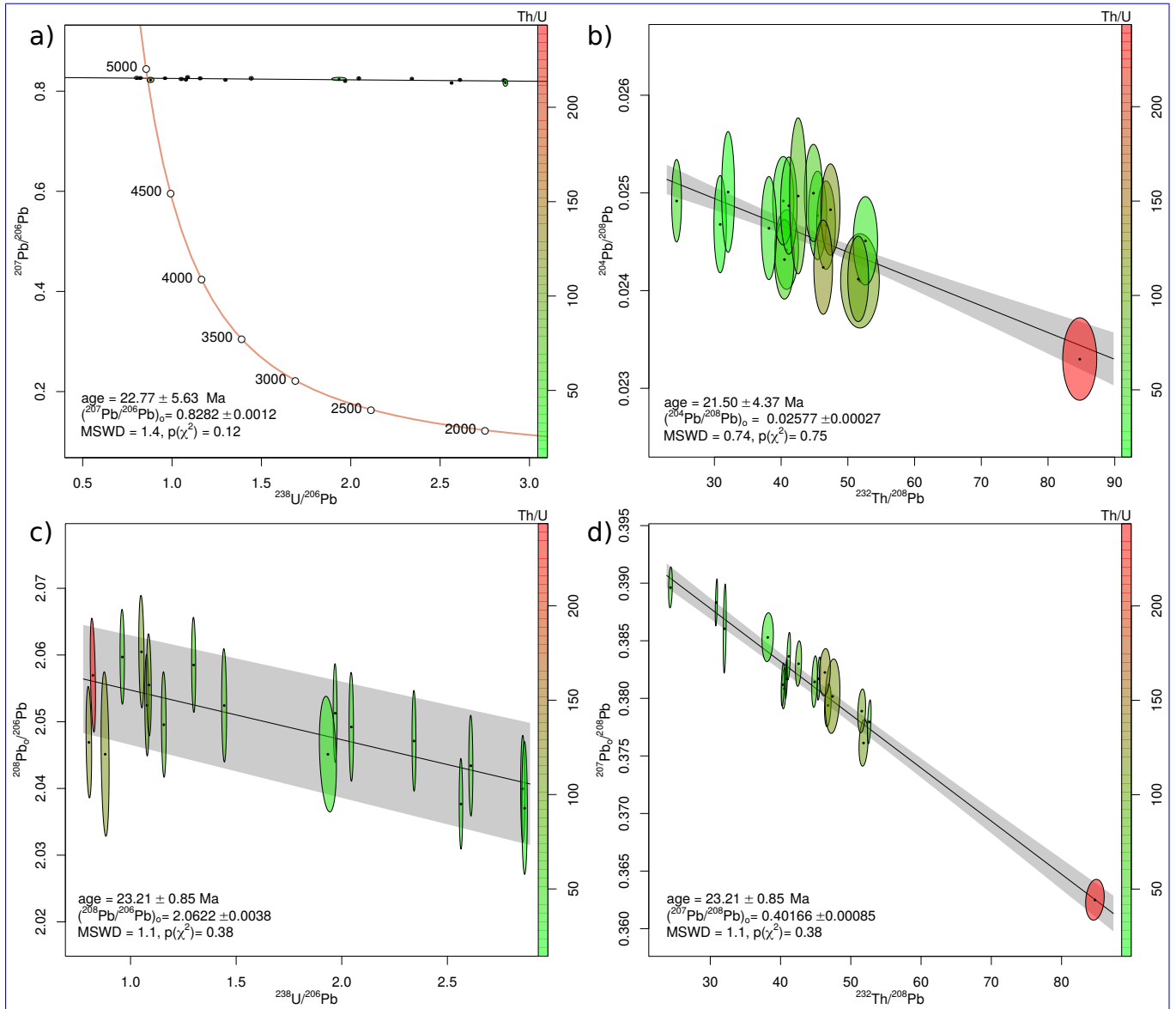


Figure 3. a) SemiTotal-Pb/U isochron (^{207}Pb -based common Pb correction) for Janots and Rubatto (2014)'s allanite data; b) Conventional Pb/Th-isochron (^{204}Pb -based common Pb correction); c) and d) Total-Pb/U–Th isochron (^{208}Pb -based common Pb correction) shown in $^{208}\text{Pb}_o/^{206}\text{Pb} - ^{238}\text{U}/^{206}\text{Pb}$ space (c) and $^{206}\text{Pb}_o/^{208}\text{Pb} - ^{232}\text{Th}/^{208}\text{Pb}$ space (d). Colours indicate the Th/U-ratio. All uncertainties are shown at 1σ .

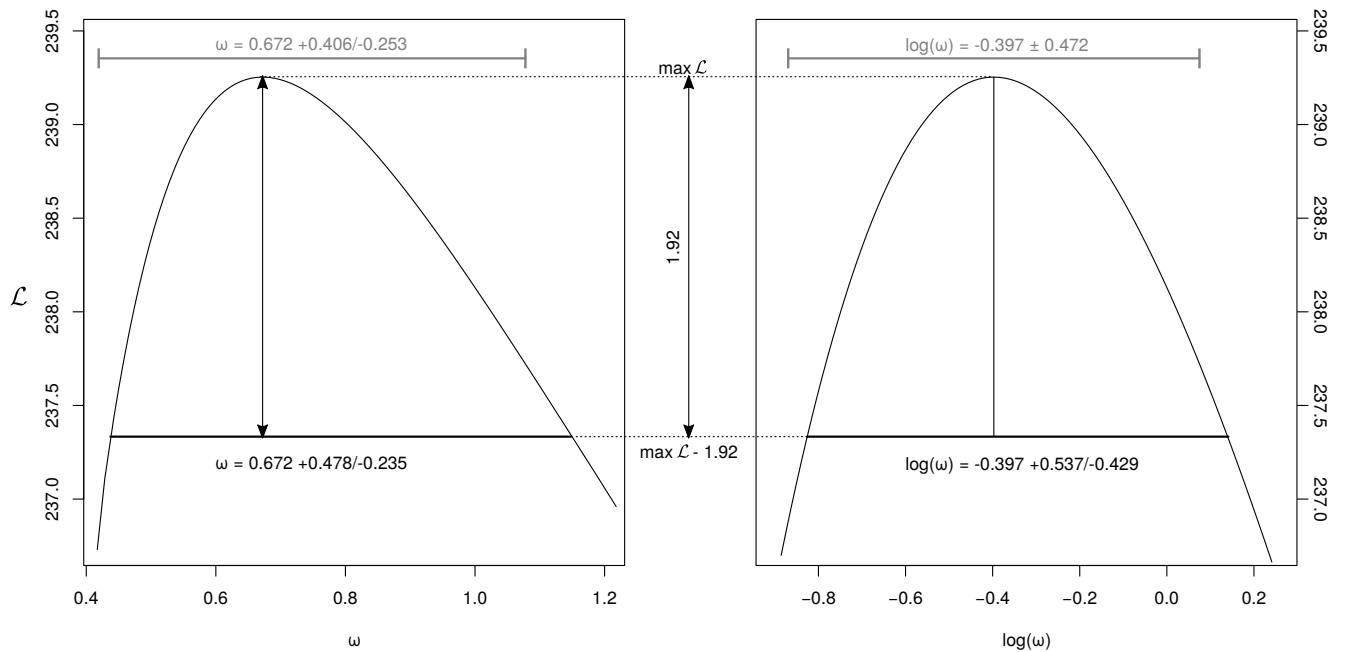


Figure 4. Profile log-likelihood intervals of the overdispersion parameter ω (black, left) and $\log(\omega)$ (black, right) for the Gibson et al. (2016) dataset. The set of ω -values whose log-likelihood fall within a range of 1.92 from the maximum value define an asymmetric 95% confidence interval. Alternatively, a standard symmetric confidence interval for $\log(\omega)$ (grey, right) can be mapped to an asymmetric confidence interval for ω (grey, left). The two approaches yield similar results.