

We would like to thank the reviewer for his effort of putting together a detailed and thoughtful review. In it, he indeed exposed some weaknesses in our study, which we would like to address below and correct in a future version of our manuscript.

This contribution describes an analytical workflow for laser-ablation-based zircon (U-Th)/He dating, introducing several modifications in data reduction compared to existing procedures, and tests these on a large dataset acquired on Fish Canyon Tuff (FCT) and LGC zircon. It investigates the factors that contribute to age dispersion by (1) a comparison of ablation-pit-volume measurements by white-light-interferometry and μ CT; (2) a machine-learning model of age dispersion depending on a series of analytical parameters; and (3) an α -redistribution model capturing effects of parent zoning and implantation from grain parts that were polished away.

The authors present the main workflow and results in the main manuscript, while the technical details of sample preparation, data reduction and modeling are presented in several Appendices. The accompanying code allows to carry out the full data reduction and machine-learning model as described in the text. The authors demonstrate how taking into account non-stoichiometric zircon composition and the effect of α -redistribution during data reduction improve the age variation and accuracy of the FCT reference material and how the observed age dispersion depends strongly on measured pit volumes and inclusion-related parameters.

This is a highly welcome publication dealing with the problem of age “overdispersion” in laser-ablation-based (U-Th)/He dating involving theoretical, analytical, and data-reduction aspects. I expect that the large (U-Th)/He dataset, detailed technical Appendix, and the related software will help the community to improve existing procedures and to compare results across laboratories.

My personal highlights are the innovative data-reduction approach using oxide-sum normalization and the correction of LA-ICP-MS-pit contributions to the measured helium, which fits well into the ongoing research on optimal pit-size configurations and dealing with zoned grains (e.g., Maier et al., 2006; Hoemann et al., in review)

We would like to thank the reviewer for his encouraging comments on our manuscript.

While this manuscript is clearly suitable for *Gchron*, there are three issues that I'd like the authors to address:

First, the documentation of the analytical data is insufficient. The manuscript presents a large set of FCT and LGC in-situ (U-Th)/He data, but there is no dedicated supplementary table besides the data table for testing the machine-learning code, which, as far as I can tell, only contains 800 of the FCT analyses and little description. I suggest documenting all FCT and LGC data together with the analytical parameters for each session, e.g., following the community guidelines for (U-Th)/He dating (Flowers et al., 2023). The methods section also needs to be more specific, e.g., regarding the laser-ablation pit sizes and depths for the acquisition of the FCT and LGC data (see comments on L. 88, 127, 128), and the elements analyzed by LA-ICP-MS (comment on L. 133).

We will provide the full dataset, including all analytical parameters, of both FCT and LGC with a future version of this manuscript. We will also include the different ages derived from the data reduction schemes tested in this study for each sample. The uploaded dataset contains 800 analyses instead of the 811 mentioned in the text, as a few grains that were obviously bad data due to a faulty ^4He /ICPMS/interferometer measurement were excluded from the ML run. We will address this in the text. Specific comments are addressed below.

Second, the authors have acquired a large dataset on the zircon reference material FCT to test their improvements for the laser-ablation (U-Th)/He workflow. However, while they provide a thorough explanation with respect to the improved fit of the FCT data to the reference age, other results are also worth exploring and will likely make a stronger case for the proposed modifications. In particular, the LGC dataset could provide an important distinction between improvements that address heterogeneous zircon from more general effects (see comments on L. 62, Fig. 6). Another aspect is the effect of the oxide-sum-based data reduction on the U and Th concentrations of the secondary reference materials, and a comparison of the present FCT dataset to existing laser-ablation (U-Th)/He data from the same material (see comment on L. 283–295).

We agree. We reduced the 100 LGC-3 spots with all four schemes (Table R1, Figure R1). For this unzoned megacryst, the dispersion is low for every scheme (σ_{MAD} of 1.2–1.7%), whereas oxide normalization moves the median toward the reference age of 476.4 ± 5.7 Ma (offset of -5.4% and -2.8% with Zr and Si normalization, -1.7% with oxide normalization). Depth weighting changes individual LGC ages by only $\sim 0.2\%$. As anticipated by the reviewer, oxide normalization mainly improves accuracy, whereas depth weighting mostly matters for the zoned FCT grains. AusZ2 (our secondary reference) and the comparison with published FCT datasets are addressed in our reply to L. 283–295. We will add LGC as a second row to Fig. 6 and discuss it briefly in the text.

Table R1. Median ages $\pm \sigma_{MAD}$ for the four data-reduction schemes. FCT values are those of Fig. 6 ($n = 811$); LGC-3 values were recalculated from the 100 spots of the LGC-3 grid. Offsets are relative to the LGC reference age of 476.4 ± 5.7 Ma (Tian et al., 2017).

Reduction scheme	FCT (Ma)	LGC-3 (Ma)	LGC-3 offset
Zr internal standard	23.9 ± 5.8 (24.0%)	450.8 ± 7.8 (1.7%)	-5.4%
Si internal standard	26.7 ± 6.2 (23.2%)	463.3 ± 6.5 (1.4%)	-2.8%
Oxide normalization	27.3 ± 4.9 (18.0%)	468.3 ± 6.7 (1.4%)	-1.7%
Oxide normalization and depth weighting	28.3 ± 3.9 (13.6%)	467.8 ± 5.7 (1.2%)	-1.8%

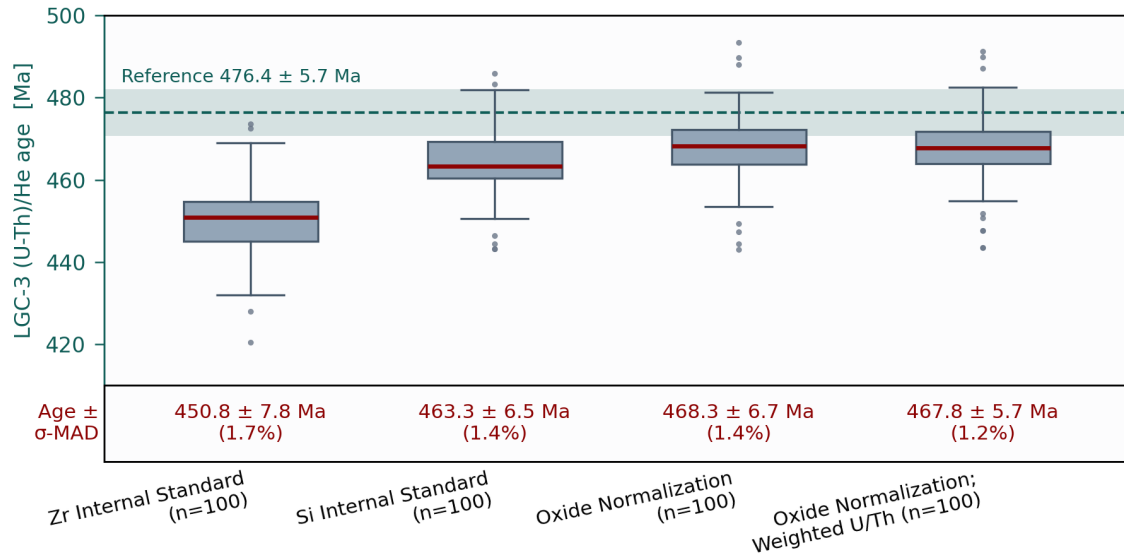


Figure R1. LA-(U-Th)/He ages of the 100 LGC-3 spots reduced with the four data-reduction schemes, shown in the format of Fig. 6. Boxes show the interquartile range, whiskers extend to 1.5 times the interquartile range, and the values below the boxes give the median $\pm \sigma_{MAD}$. The shaded band and dashed line show the reference age of 476.4 ± 5.7 Ma (Tian et al., 2017). This panel will be added as a second row to Fig. 6.

Third, the rationale for the machine-learning dispersion model needs to be clarified. I understand why the authors deliberately excluded the He, U, and Th concentrations because of their direct appearance in the age equation. However, I wonder whether including the pit volume, which also affects the (U-Th)/He age directly, introduces a similar correlation with the model's target. In particular, could the direct relationship between pit volume and age obscure correlations with other variables (see comments on L. 234, L. 371)? At least, I would appreciate a discussion of the practical implications, e.g., if a more conservative treatment of pit-volume uncertainties is necessary.

Unlike ^4He , U , and Th , which together define the age, pit volume alone does not. For correctly measured volumes, a larger pit simply contains proportionally more He , so pit volume can only predict the age residual if volume-related errors exist. A positive or negative correlation of age and pit volume, which the ML model aims to find, would not exist for a “perfect” volume measurement. To test whether including the pit volume in the model masks other variables, we reran the model without He pit volume and depth (Table R2, Figure R2). The OOF variance explained decreases from $28.7 \pm 1.8\%$ to $20.8 \pm 1.7\%$. Part of this information is taken over by the time of He analysis, which correlates with pit volume ($r_s = -0.76$) but carries almost no information when pit volume is included. Apart from the QMS baseline, which also correlates with time and volume, the other predictors keep similar importances, so pit volume does not mask them.

When rerunning the model for this reply, we noticed that the leave-one-group-out values in Table 2 were based on a single cross-validation partition, which came from an earlier version of the code. Recomputed on repeated partitions, the leading groups are unchanged, but removing the zonation group lowers the OOF variance explained by 7.4 ± 1.1 percentage points (8.1 ± 1.5 without pit volume), although its effect on σ_{MAD} remains small ($+0.10 \pm 0.13$ Ma). The zonation metrics therefore help to predict the largest residuals, independent of pit volume. We will update Table 2 and the related text, and add the second model run to the supplement. Practical implications for the volume uncertainty are addressed in our reply to L. 371.

Table R2. Dispersion model with and without He pit volume and He pit depth. Performance values are mean $\pm$ SD over 10×10 repeated cross-validation. LOGO values are mean $\pm$ SD of paired changes over five repeated partitions. PI is the permutation importance (ΔR^2).

Metric	With pit volume and depth	Without
OOF variance explained	$28.7 \pm 1.8\%$	$20.8 \pm 1.7\%$
Unexplained σ_{MAD} (Ma)	3.02 ± 0.05	3.33 ± 0.07
Null model, OOF variance explained	$-21.1 \pm 3.0\%$	$-20.3 \pm 3.4\%$
PI, He pit volume	0.343	n/a
PI, time of He analysis	0.003	0.204
PI, ^{208}Pb	0.047	0.079
PI, $^{206}\text{Pb}/^{238}\text{U}$ age	0.072	0.070
LOGO zonation, loss of OOF variance explained	7.4 ± 1.1 points	8.1 ± 1.5 points
LOGO zonation, $\Delta\sigma_{\text{MAD}}$ (Ma)	$+0.10 \pm 0.13$	$+0.17 \pm 0.14$

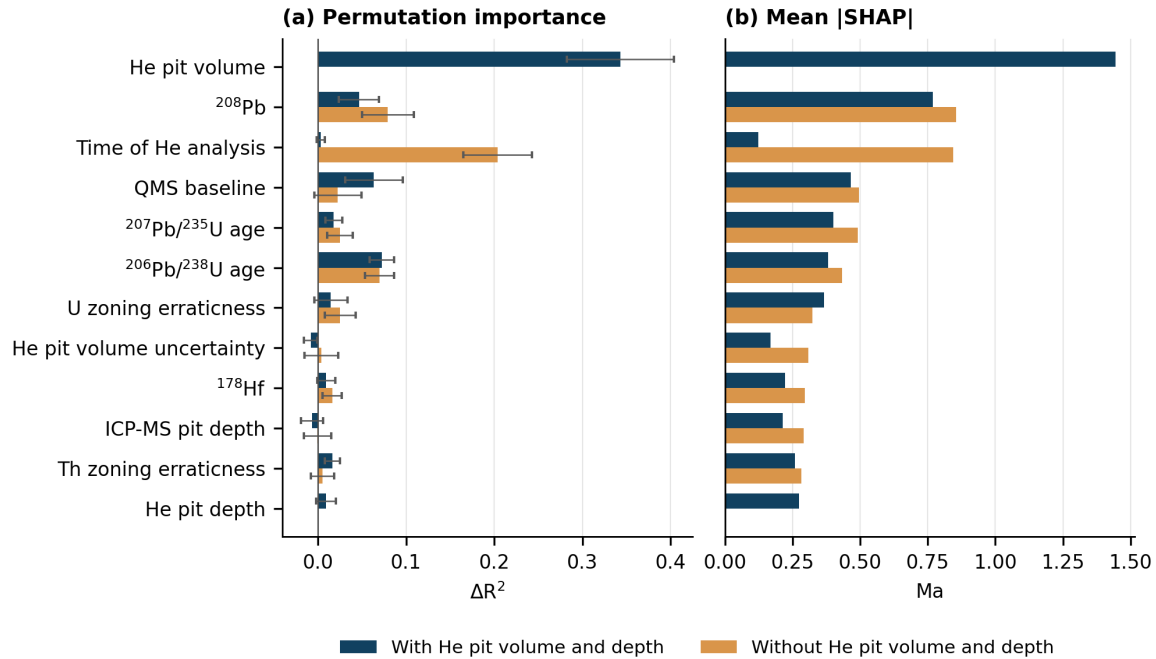


Figure R2. Permutation importance (a) and mean absolute SHAP contribution (b) of the 12 most important predictors in the dispersion model with (blue) and without (orange) He pit volume and He pit depth. Error bars in (a) show the standard deviation over permutation repeats. The 21 predictors that were left out of this figure have a mean $|\text{SHAP}|$ of at most 0.25 Ma in both runs.

Specific comments:

L. 43: I encourage the authors to develop these factors a little more for LA-based (U-Th)/He dating and include the recent literature dealing with these issues. For example, we can expect the effects of α -ejection bias mentioned for whole-grain dating to be mostly negligible. However, other factors such as inclusions or zoning would hit the small, mis-matched ablation volumes harder. Please include references to the articles that discuss the problems related to zoning in laser-ablation (U-Th)/He dating such as Danišik et al. (2017), Léger et al. (2023), and Vermeesch et al. (2023). The recent preprint by Hoemann et al. (in review) discussed and even quantified the effect of lateral zoning data with respect to different ablation pits.

There is some argument to be made that the effects of α -ejection bias can be present in LA-(U-Th)/He datasets if the polished depth is not well calibrated, or grains of different sizes are mounted and polished together. We agree that this and other factors (such as zoning and inclusions) are worthy of a longer discussion in the introduction chapter.

L. 62: I agree with the authors that LGC seems to be a nearly-ideal material to test accuracy and precision of the presented workflow. I thus wonder how (besides the final result for LGC presented in L. 301) the step-wise modifications of the analytical procedure affected the LGC data relative to the heterogeneous and inclusion-bearing FCT. I would predict that the oxide normalization probably would have improved accuracy, while parent-source weighting would leave the data from the homogeneous crystals mostly unaffected. I suggest presenting these data in parallel to the FCT data in Fig. 6 and section 3.2.

We agree, and the LGC data confirm the reviewer's prediction (see our reply to the second general comment, Table R1, Figure R1). Oxide normalization reduces the offset from the reference age from -5.4% (Zr) and -2.8% (Si) to -1.7% , whereas depth weighting leaves the ages essentially unchanged (median per-spot change -0.1%). We will add these results to Sect. 3.2 and Fig. 6.

L. 70: Consider citing Dunkl et al. (2024) here, who presented an alternative way of mounting in epoxy resin.

Added.

L. 86: Is this the nominal fluence corresponding to the 7 mJ? Please clarify.

The energy of 7 mJ and the fluence of 11.38 J cm^{-2} are set on the laser. The final energy of $\sim 0.75 \text{ mJ}$ was measured with an energy meter on the sample.

L. 88: The variation in the number of pulses is not explained in the subsequent text. Do these different numbers of pulses refer to different acquisition sessions for the FCT data? This broad range of pulses is surprising to me because it is not reflected in the pit depth given in the supplementary material.

All ^4He pits of the samples of this study were ablated with 100 pulses and a spot diameter of $20 \mu\text{m}$. We will update the text.

L. 127: The description of the LA-ICP-MS analysis requires some more detail. Where the main analyses of this study on the FCT grains done with one specific diameter or with different ones? If these varied between sessions, this needs to be documented properly in the Supplementary Material.

The diameter of the laser spot was constant for all samples, at $50 \mu\text{m}$.

L. 128: I assume the $25 \mu\text{m}$ pit depth for the LA-ICP-MS refers to the added depth on top of the He-extraction pit? Were the reference materials (AusZ2, LGC-3, NIST SRM 610) ablated prior to LA-ICP-MS to obtain the same ablation geometry as for the unknowns?

Correct. No, the reference materials were not pre-ablated.

L. 133: I suggest listing ^{235}U and ^{147}Sm with the other parent nuclides, although ^{235}U was not directly used for age calculation. Also, list the other trace elements of interest that were analyzed (i.e., ^{27}Al , ^{57}Fe , etc.).

The chapter describes the general workflow for LA-(U-Th)/He dating in our lab. We will update the specific parameters for the samples of this study.

L. 185: Vermeesch et al. (2012) and Evans et al. (2015) describe small-diameter LA-ICP-MS pits placed within their He-extraction pits.

We will adjust the wording of that sentence.

L. 222: Add the primary U/Pb reference material GJ-1 to the methods section.

We will add a reference to GJ-1 in 2.4.

L. 234: I understand the decision of excluding parent and daughter nuclides to avoid the dominance of inherent correlations, but what about the pit volume, which is also part of the age calculation (see L. 209)? If the measured He concentration (ncc) co-varies more strongly with U and Th than with the pit volume, I would expect the pit volume to strongly affect the apparent (U-Th)/He age and thereby mask the effect of other variables.

See our reply to the third general comment. Because pit volume alone does not define the age, it can only predict the residual through volume-related errors. Removing pit volume and depth from the model shifts part of their information to a correlated proxy (time of He analysis) but does not reveal other predictors. We will add this test and the rationale for including pit volume to the next version of the manuscript.

L. 270: I wonder what the implications of the pit-volume comparison are for the uncertainties of the FCT analyses in the subsequent discussion. Looking at the variation Fig. 5, I wonder whether applying the 5.5–6% from the 40/8 μm pits may be more appropriate for the FCT data.

We were hesitant to apply an uncertainty to our dataset that is not purely a measured analytical one. The results of our ML analysis do suggest that the analytical uncertainty of our pit volume measurement is at least underestimated. The magnitude of the true uncertainty (including all external uncertainties) is indeed probably closer to 5 – 6 %.

L. 271: Please cite an overall range of single-grain uncertainties for the (U-Th)/He ages to compare to the age range in Fig. 6 and the uncertainties mentioned in the text.

The data reduction scheme used for single-grain (dissolution) (U-Th)/He ages applies a uniform 8% uncertainty to all grains.

L. 283–295: This paragraph adds interpretation to the results presented in this section. While interpreting part of the results here is not a bad idea, I would like to see a dedicated discussion section on how the analytical modifications affected the results for the FCT dataset. This improvement represents one of the main goals of the manuscript and is strongly supported by the bulk of the data, so I see the risk that focusing only on FCT age dispersion undersells the strongest aspect of the manuscript. I also suggest using this discussion section to further explore: (1) the bias in U and Th concentrations due to substitution and extrapolation, e.g., by discussing concentrations in the secondary reference materials; (2) the effect of the modifications on the well-behaved reference material LGC; and (3) comparing the results to existing laser-ablation (U-Th)/He datasets from FCT zircon (e.g., Horne et al., 2016; Léger et al., 2023), which used different pit-size combinations and did not reduce the LA-ICP-MS data by oxide normalization.

We agree and will move this interpretation into a short dedicated discussion section (new Sect. 4.1) that covers the three points.

(1) For AusZ2, the three normalization schemes agree within 3.4% (Table R3). The oxide-normalized U and Th concentrations are 4.0% and 2.3% higher than the LA-ICP-MS values of Kennedy et al. (2014), within the range of matrix effects expected for calibration against NIST SRM 610 (Sylvester, 2008), and the Th/U ratio is identical for all schemes. Substitution and extrapolation effects are therefore small for inclusion-free zircon. The offset of Zr normalization is material-dependent (0.8% lower U for AusZ2, ~4% higher eU for LGC, ~14% higher median eU for FCT), which indicates that the large FCT offset is caused by inclusions and non-stoichiometric ablation volumes. We will revise L. 287–291 accordingly.

(2) See our reply to L. 62.

(3) We recalculated the statistics of the FCT datasets of Horne et al. (2016) and Léger et al. (2023) from their published data (Table R4, Figure R3). All medians lie between 28.0 and 29.0 Ma. Our dispersion (13.6%) is similar to that of FCT-A of Léger et al. (2023; 13.8%) and lower than that of their batches FCT-B and FCT-C (20.3% and 24.6%), which were reduced with a Si internal standard. Our own Si-normalized reduction yields 23.2%. The smaller, CL-screened dataset of Horne et al. (2016; n = 44), which used a wider U-Th pit (65 vs. 50 μm), has a lower dispersion of 10.0%, consistent with the inclusion- and zonation-related effects discussed in Sects. 4.2–4.4. We will add this comparison as one paragraph. It should also be noted that we did not bias our grain selection in any way, for example by not picking grains that showed visible inclusions, broken grains, grains of the same size etc. So the three datasets are not entirely comparable.

Table R3. U and Th concentrations of AusZ2 (mean $\pm$ 1 SD, n = 5) obtained with the three normalization schemes. Deviations are relative to the LA-ICP-MS values of Kennedy et al. (2014; U 244 ± 3 $\mu\text{g/g}$, Th 102 ± 1 $\mu\text{g/g}$, 1 σ standard error of the mean).

Normalization	U ($\mu\text{g/g}$)	Th ($\mu\text{g/g}$)	ΔU	ΔTh
Zr internal standard	251.8 ± 5.6	103.5 ± 1.5	+3.2%	+1.4%
Si internal standard	260.2 ± 2.0	106.9 ± 3.9	+6.7%	+4.8%

Normalization	U (μg/g)	Th (μg/g)	ΔU	ΔTh
Oxide normalization	253.8 ± 3.5	104.3 ± 2.3	+4.0%	+2.3%

Table R4. Laser-ablation (U-Th)/He datasets of FCT zircon. Statistics were recalculated from the published data tables ($\sigma_{\text{MAD}} = 1.4826 \times \text{MAD}$; MSWD calculated with 1σ uncertainties). The He pit depth given for Horne et al. (2016) is the equivalent cylinder depth of their median pit volume.

Study	n	He pit (diameter / depth)	U-Th pit diameter	Parent quantification	Median (Ma)	σ_{MAD} (Ma)	MSWD
Horne et al. (2016)	44	25 μm / ~8 μm	65 μm	Pit volume (SynZircon)	28.8	2.9 (10.0%)	4.6
Léger et al. (2023)	249	30 μm / 15 μm	60 μm	^{29}Si internal standard	28.3	5.1 (18.1%)	9.2
This study, Si internal standard	800	20 μm / ~7 μm	50 μm	^{29}Si internal standard	26.7	6.2 (23.2%)	n/a
This study, final reduction	800	20 μm / ~7 μm	50 μm	Oxide normalization and depth weighting	28.35	3.85 (13.6%)	8.9

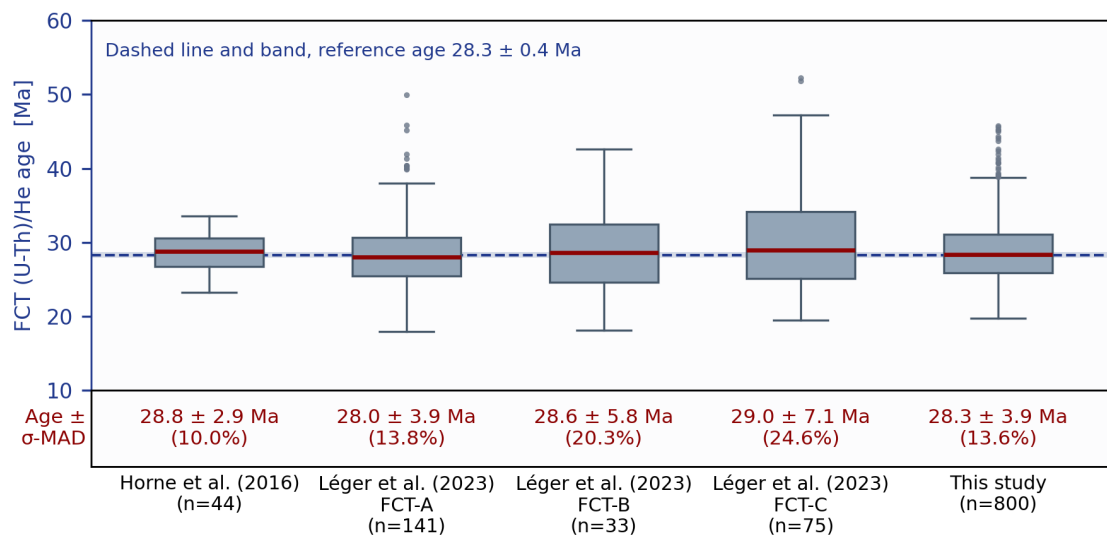


Figure R3. LA-(U-Th)/He ages of FCT zircon from Horne et al. (2016), the three batches of Léger et al. (2023), and this study (final reduction), shown in the format of Fig. 6. Statistics were recalculated from the published data tables. The dashed line and band show the reference age of 28.3 ± 0.4 Ma (Gleadow et al., 2015).

L. 289: Replace 15.3 wt% by 50 wt%, the approximate stoichiometric Zr content in zircon. Please check.

Corrected.

L. 371: I like the discussion about the influence of pit volume beyond its obvious contribution to the measured age. I think the mismatch between variation in pit volume and pit-volume uncertainty is another important observation (L. 393). As mentioned in the general comments, I wonder if its inclusion in the machine-learning models may either mask effects of other variables and what the direct implications of this analysis are for treating measurement uncertainties on the pit volume that include both the instrument capacities and pit-specific factors.

On the flat, homogeneous LGC megacryst, the scatter of 100 spots (σ_{MAD} 1.2%, SD 1.9%) is smaller than the median propagated age uncertainty (2.8%). The interferometric measurement itself is therefore accurately described by the bootstrap uncertainty plus the 2% systematic term, and the larger volume-related scatter in FCT reflects grain-specific factors such as tilted or rough surfaces, inclusions, and focusing. For grain-by-grain analyses of natural

crystals, a more conservative external volume uncertainty of ~5.5–6% is warranted, based on the μ -CT comparison (see our reply to L. 270). Propagating 6% instead of the analytical value reduces the MSWD of the FCT dataset from 8.9 to 4.0, so it explains part, but not all, of the excess dispersion. We will add this to Sect. 4.1, while keeping the analytical uncertainties in the data table unchanged. Regarding masking, see our reply to the third general comment.

L. 394: Which exact analysis does this 10x10 grid refer to, the LGC data presented in section 3.2 or measurements for volume benchmarking? Please specify.

It refers to the spots used for LA-(U-Th)/He on LGC. A grid of 10 X 10 spots was measured on a single polished megacryst. This is unlike the FCT data, which was measured on hundreds of individual crystals.

L. 415: There are two interesting aspects here that would merit a short note each (2–3 sentences). First, is there a mechanism that would explain why low- ^{208}Pb grains have a positive PDP (Fig. 9)? Is this just thorium's contribution to the age equation? Second, what are the practical implications besides using U/Pb age and ^{208}Pb as proxies for disturbance? I think we can safely assume that the likelihood of hitting inclusions that disturb the isotopic systems increases with the LA-ICP-MS pit volume. I'd then expect that large spot diameters compromise the main advantages of laser-based zircon (U-Th)/He dating.

It is likely not related to the contribution of Th to the age equation, as ^{208}Pb is predominantly “common”, and a low ^{208}Pb concentration simply reflects the absence of common lead carrying inclusions. Beyond that, we have no explanation for the cause of this correlation.

Unfortunately, we did not measure the effect of a larger or smaller spot size on the dispersion. There are likely two competing effects here: Larger spot volumes do increase the likelihood of hitting an inclusion, which might perturb the measured age. But on the other hand, a larger volume also averages out many of the effects that cause dispersion in the first place, such as erratic zonation patterns and simple analytical uncertainties. For practical purposes, the number of spots with an extremely high ^{208}Pb concentration is low and therefore does not significantly contribute to the overall dispersion of the entire dataset. It could serve as a good marker to identify outliers though.

L. 419: Please consider restructuring the discussion here by combining sections 4.3 and 4.4. From L. 435 on, the discussion here already includes the “ghost grain” volume, and section 4.4 is rather the quantitative extension.

We agree and will improve the flow of these two chapters.

L. 424: Again, I would add a reference to Léger et al. (2023) here.

Added.

L. 431: I appreciate the critical discussion of the 1D vs. 3D characterization. This is especially important for the “over-coring” geometry of laser-ablation pits as material from below the He pit is mixed with the annular region around the He pit (see discussion in Härtel et al., 2026). This geometry further complicates the relationship between vertical zoning and time-resolved ICP-MS signal.

We have observed this effect but found it to be largely negligible. In tests where we drilled through an entire grain, the signal on all elements goes to close to background once the other side is reached. We have not observed any significant mixing of shells or zones in our LA-ICP-MS profiles that would come from the walls of the pits. In regard to the volume below the He pit, the ICP-MS pit seems to even out its pit bottom fairly quickly - after a few μm . We did not observe any remnant He pit topography after the ICP-MS ablation. So mixing is likely minimal.

L. 433: I would expect sub-stopping-distance zoning to be substantially averaged out if it is not super-imposed on a larger-scale U-Th enrichment or depletion trend.

While this is correct, our point here is that any lateral variation in U/Th, no matter the amplitude, will go undetected since we have to rely on a simple 1D profile.

L. 445: I think it would be helpful for the reader to also include the He pit depths here.

We will mention the average He pit depth here (~7.2 μm).

L. 451: Please indicate the diameters of the He and ICP-MS pit assumed for the model.

Since we assume “layered-cake” zoning, meaning the U-Th concentration only changes in one dimension, our weighting model curve does not change with the diameter of either pit.

L. 480: I agree that without U-Th zonation from the polished-away part of the grain, we cannot be sure how much our eU estimates are in error. I’d appreciate if references to our discussion in Härtel et al. (2026) and that in Hoemann et al. (in review) on robust pit-size configurations could be included here. I also suggest adding a reference to Evans et al. (2015) regarding pit geometry and accurate volume measurement.

We will include a reference to those relevant studies.

L. 498: The statement about the one-dimensional U and Th zoning needs to be rephrased because it omits the fact that the parent-source weighting largely takes care of the vertical zoning. This is an important positive result that should be emphasized.

We will rephrase to “By contrast, measured one-dimensional U-Th zonation and mineral-phase proportions show no statistically significant correlation with age residuals after applying a U-Th depth-weighting approach and using a stoichiometric sum normalization to account for inclusions.”

L. 594: Please indicate the interval of instrumental-blank measurements; are these repeated after every ablation?

A blank was measured after every 5 analyses.

L. 660: Was the σ -scaled or the raw 5x MAD used for filtering?

The raw 5xMAD was used.

L. 742: I see that the x and y coordinates are ignored for the calculation of the parent-source weights. What does that mean for the accuracy of the correction for different ICP-MS pit sizes?

See our reply to the comment on L. 451.

Figure 4: I suggest adapting the figure to the actual ablation-pit parameters used for the FCT analyses.

We will adapt it to the average He parameters (20 μm diameter and 7 μm depth).

Figure 6: Consider adding another row with boxplots for samples LGC to have a comparison of the improvements for a zoned and an unzoned reference material.

We will add a second row showing the LGC-3 ages for all four reduction schemes (Figure R1), using the same layout and statistics as the FCT row.

Figure B2: Please indicate that the lower left plot corresponds to the He-pit configuration used for the data acquired in this study.

The lower left panel shows the results for a He pit of 30 μm diameters. For our FCT analyses, we used a diameter of 20 μm .

Technical corrections

L. 7: I would rephrase the second part of the sentence to make it more understandable that the depth-weighting applies to LA-ICP-MS.

We changed to “parent-source depth weighting applied to the measured U-Th profiles”

L. 9: Age variation – “dispersion” may make people think of σ in the radial plot.

We disagree. “Age variation” suggests that the ages differ for a reason, such as a geologic or systematic one. “Dispersion” is the more established term in geochronology.

L. 11: He pit volume emerges (singular as in the rest of the phrase).

Corrected.

L. 40: F_T with a subscript.

Corrected.

L. 49: I suggest to rephrase: “The goal of this study is to (1) present an improved routine workflow for LA-(U-Th)/He dating and (2) identify key factors contributing to the analytical uncertainties using a machine-learning approach.”

Corrected.

L. 99 and throughout (optional): Consider using mass-concentration units such as $\mu\text{g/g}$ instead of ppm.

We agree. It is better to use SI units, and we will change this throughout.

L. 129: I think the minimum grain size (diameter) for a He pit of 30 μm and a 30–40 μm broader LA-ICPMS pit should rather be in the 70–80 μm range. Please check.

Our point here is that without a complicated alpha-ejection correction on our ablated volume, the minimal grain size is $\sim 50 \mu\text{m}$, assuming our He pit size is negligibly small. We will clarify.

L. 220: Rephrase: This represents an advantage over conventional dissolution-based (U-Th)/He dating.

Corrected.

L. 225: I only counted 800 analyses in the Supplementary data sheet, please double-check.

See our reply to the first general comment.

L. 255: Please briefly introduce the mean bias.

Added.

L. 275: Consider using “variation” instead of “precision” here; I think most readers would link “precision” with single-grain uncertainties.

We would avoid using “variation” in this context, for the reason given in our reply to the comment on L. 9.

L. 277: Replace “reducing” by “processing”.

Corrected.

L. 307: Please indicate if the variations are given as σ_{MAD} .

They are. Corrected.

L. 404: Superscript on ^{208}Pb .

Corrected.

L. 406: Missing period.

Corrected.

L. 420: I am not sure why this result is unexpected, since the downhole variation in U and Th was corrected for (see L. 425). Please rephrase.

“corrected for” is the wrong expression in this context, since the depth weighting is not able to fully correct the effect of zonation in the entire grain. It assumes a very simple zonation model (“layered-cake”) to weight the 1-D ICP-MS profile. It is furthermore quite surprising that despite the still large dispersion in the dataset, no apparent correlation with zonation metrics exists. This is despite the well-established link between the two from previous studies.

L. 464: Consider “5 μm pit depth” in the parentheses, as the reader may think of a 5- μm -diameter pit.

Corrected.

L. 497: Superscript on ^{208}Pb .

Corrected.

L. 697: Missing “to” after “assigned”.

Corrected.

Figure 2: I suggest to rescale the axes to the typical spot sizes used in zircon (U-Th)/He studies and to eU values more relevant to typical zircon (U-Th)/He applications.

Corrected.

Figure 5: Introduce the abbreviation WLI in the caption.

Corrected.

Figure 7: Specify the meaning of the annotations below each boxplot, I assume these are again medians and σ_{MAD} .

It is specified to the left, next to the y axis.

Figure 10: Please add to the caption or axis labels that the axes refer to the He pit. Also rescale the axes to depth and diameter ranges that are typically used (see Forte et al., 2022 or Zawacki et al., 2022).

We fixed the x axis title. The full range shown, however, is necessary for our discussion.

Figure 11: The annotations in the grey sketches at the top are difficult to read and should have a larger font size. Consider using a legend for the eU concentrations and the pits to reduce the clutter with larger fonts. Please also indicate the diameters of the He and ICP-MS pits.

The diameter of the pits is irrelevant here, since we again assume a layered-cake zonation model. We will fix the clutter in the top of the figure.

In B, I suggest using the same scales for the y axis in B to show the differences in eU error more intuitively.

Fixed.

Figure A2: Period missing at the end of the caption.

Fixed.

Figure B2: The legends are difficult to read, consider increasing the font sizes. Please also declutter the overlapping x-axis labels in the upper plots. I suggest adding index letters for easier reference to the figure.

Fixed.

Supplement Table zFCT_AGES_RESULTS.xlsx: Zr concentrations are labeled as %, but the numbers suggest they are in $\mu\text{g/g}$.

Fixed.