

We thank the reviewer for the detailed and constructive review. Several comments overlap with those of Reviewer 1. In these cases, we refer to our reply to Reviewer 1.

The manuscript presents the detailed methodology of the UTChron laboratory at the University of Texas for the acquisition of laser-ablation (U-Th)/He data in zircons, which includes a new approach to calculating parent nuclide concentrations via oxide normalisation and a method for computing depth-weighted parent nuclide concentrations as analytical improvements to existing protocols. Furthermore, the authors apply a machine learning approach to conduct a detailed investigation into the sources of age dispersion in laser-ablation (U-Th)/He dating and identify the most important contributors in a dataset of 811 measurements from the Fish Canyon Tuff (FCT). These measurements are supplemented by 100 measurements in the LGC-3 zircon standard.

The manuscript is a valuable contribution to the literature that excellently fits within the recent effort in the thermochronology community to make measurement protocols and instrument setups openly available (e.g. Glotzbach et al., 2026; Metcalf & Flowers, 2026). It is also a welcome contribution to the ongoing investigations into the sources of dispersion in LA-(U-Th)/He data and efforts made towards analytical improvements (e.g. Maitra et al., 2026; Hoemann et al., in prep.). Personally, I particularly appreciate the authors' approach to addressing the redistribution of alpha particles and its effect on in situ dating, because I deem this problem to be undervalued in many current LA-(U-Th)/He applications. I find the parent-source depth weighting method presented an elegant approach to the problem.

In my view, the manuscript is well-written and has the potential to be a key resource for a laser ablation (U-Th)/He workflow owing to the detailed descriptions of the equipment and methodology provided in the main text and appendices and the shared code for data reduction. Yet, there are several issues I list below that I'd like the authors to address because they impede the clarity of the methodology as presented in its current form and potentially impact the conclusions drawn from the machine learning approach.

We thank the reviewer for the positive assessment.

General comments

1. Overall, the manuscript presents a commendably thorough and detailed explanation of the analytical setup and steps of zircon LA-(U-Th)/He dating in the main text and appendices. However, crucial details, especially pertaining to the choice of He and U-Th ablation pit sizes, geometries and configuration, are imprecise and not sufficiently discussed, preventing the presented workflow from being fully reproducible. I have pointed out the specific instances in my detailed line-by-line comments.

We agree. We will state all pit dimensions used for the FCT and LGC-3 data (He pit 20 μm in diameter, 100 pulses, median depth 7.4 μm ; ICP-MS pit 50 μm in diameter, median depth 39 μm) and add a short paragraph on the choice of pit geometry. Details are given in our replies to the line-by-line comments.

2. The manuscript presents a general LA-(U-Th)/He workflow while also detailing the acquisition of their extensive FCT and LGC-3 datasets. While both are crucial, their overlap poses a challenge for the reader to distinguish whether the settings described in the measurement procedure are specific to the analysed samples or general practices. This impedes the clarity of the methodology presented. I have highlighted relevant instances in the detailed line-to-line comments. Furthermore, the traceability of analytical steps by the reader is hindered by the unavailability of the analysed datasets. I recommend adding them to the supplement.

We agree. We will separate the general workflow from the settings used for the FCT and LGC-3 data, which we will summarize in one place in Appendix A. As mentioned in our reply to Reviewer 1, the full FCT and LGC-3 datasets, including all analytical parameters, will be added to the supplement.

3. The setup and parameter choices for the applied machine learning model are insufficiently described. As the parameter choices potentially have a strong impact on the interpretation of the modelling results (see my detailed line-by-line comments), they need to be elaborated on to provide the reader with necessary context and prevent

misinterpretation. Furthermore, the full potential of the machine learning approach and the extensive acquired datasets is not exploited, as the machine learning analysis and the evaluation of the effect of the presented workflow improvements was only performed on one of the datasets, FCT, but not LGC-3. This is surprising to me, since analysing both would strengthen the interpretation and conclusions drawn, and would presumably allow greater insight into the effect of workflow-specific factors vs grain-specific factors on the age dispersion. The manuscript would benefit from performing and presenting the analyses for LGC-3, if possible.

We will describe the model setup in more detail in Sect. 2.6 and Appendix D (see our replies to L. 234, 247, 366, and 818). We did not apply the model to LGC-3 for two reasons. First, 100 spots are too few to train a stable model with 33 predictors. Second, there is no excess dispersion to explain. The 100 LGC-3 spots scatter by only 1.2% (σ_{MAD}), less than their median analytical uncertainty of 2.8%, and their MSWD is below one. Instead, we compare the four data-reduction schemes for both LGC-3 and FCT (new second row in Fig. 6; see our reply to Reviewer 1). Oxide normalization improves the accuracy of both materials, whereas depth weighting reduces the dispersion only in the zoned FCT grains.

In summary, the paper is overall well-written and constructed but requires clarification of several methodological aspects. After these issues are addressed, I deem this manuscript a significant and valuable contribution to Geochronology.

Detailed line-by-line comments and questions

Abstract

l. 14: I am conflicted about this statement, as it strongly suggests that zonation generally has little effect on LA-(U-Th)/He dating, which your results don't indicate. As you point out later in the manuscript (line 425), the lack of a strong predictive signal from the degree of zonation is most likely a positive effect of your specific data-reduction workflow and not a general feature of LA-(U-Th)/He. I suggest putting more emphasis on this strength of your parent-source depth-weighting method in the abstract.

We agree and will rephrase this sentence. A model run on ages without depth weighting shows that depth weighting removes more than half of the predictive signal of the measured zonation (see our reply to L. 420). The revised abstract will emphasize this, and that zonation in the polished-away part of the grain remains a major source of dispersion.

Introduction

l. 43: Please consider elaborating on how these aspects can affect LA-(U-Th)/He dating. This would not only provide important background information but also situate your study in the context of current research and literature.

We agree and will expand this paragraph, as also requested by Reviewer 1 (L. 43).

Improved analytical workflow

l. 58: I am curious how many different grains you analysed for the 811 spots in FCT. Did you measure several spots in the same grain? If so, it would be interesting to determine whether the ages you obtained in a grain were consistent. Please also consider mentioning the separate dataset you describe in Appendix C in this introductory paragraph, as those data are displayed in Fig. 5, but not described or mentioned in the main text.

Each grain contained one or two spots, but we did not count the total number of grains. As our dataset does not record which spots share a grain, we cannot test the consistency of ages within grains here. We will state this and mention the μ -CT benchmark dataset (Appendix C) in this paragraph.

l. 74: I suggest adding the duration of heating and the force used for hot-pressing here and not only in the caption of Fig. 1.

Added (80 s at 320 °C and 0.6 MPa, followed by a second pressing step of 60 s after the grains are placed).

l. 78: Please add information on the force used for polishing.

Added (15 N).

l. 79: Do you check for uneven abrasion of the mount during polishing? This would be helpful information for readers using your manuscript as a guide for analytical procedures, because in my experience, mounts tend to abrade unevenly, especially when grains of different diameters are embedded.

We did not quantify this systematically, but we observed that the polished depth can vary by up to ~10 µm between the center and the rim of a mount, mainly because the PFA disc warps during hot pressing. We will mention this in Sect. 2.1 and Appendix A. This variation adds to the grain-to-grain differences in the polished-away volume discussed in Sect. 4.4.

l. 88: Could you elaborate on your choice of a 30 µm diameter spot for the He measurements? Is a 30 µm spot diameter your default choice for LA-(U-Th)/He measurements in zircons, or do you determine optimal sample-specific spot diameters by testing several spot sizes? Did you use this 30 µm spot diameter for all FCT and LGC-3 measurements? Furthermore, you give a wide range of total laser pulses. Did you use this range to deliberately generate a spread of pit depths in your measurements? The manuscript would benefit from this information.

A 20 µm spot is our default for zircon, and we used it with 100 pulses for all FCT and LGC-3 He pits. The 30 µm in L. 88 was an error, and the pulse range describes our general routine; it was not used to create a spread in pit depths (see also our reply to Reviewer 1, L. 88). For FCT, this setting yields a median ⁴He amount ~16 times the limit of quantification. We will correct the text.

l. 89: Could you please explain how you determined the ablation rate? Also, does the pit depth range given here apply to both FCT and LGC-3? This is not clear as a reader.

The ablation rate is the interferometrically measured pit depth divided by the number of pulses. For the FCT He pits (100 pulses), the median depth was 7.4 µm (interquartile range 7.0–7.8 µm), i.e., ~0.07 µm per pulse. We will give the pit depths of both FCT and LGC-3 in the text and in the data table.

Fig. 2: Consider adjusting the x- and y-axis ranges for pit diameters and eU concentrations to values one realistically encounters when conducting zircon LA-(U-Th)/He measurements. This would make the figure more informative. Further, I suggest including the limit of quantification of your QMS line in the figure caption for context.

We rescaled the axes (see also our reply to Reviewer 1) and will add the limit of quantification ($\sim 0.58 \times 10^{-16}$ mol) to the caption.

l. 113: Please define for the reader what you mean by “clean” spots.

We will define clean spots as pits with a regular, cylindrical shape and without melt droplets (popcorn), ejecta, or cracks.

l. 118: Are these ~300 ablation pit measurements independent from the 100 LGC-3 measurements or is there overlap? Could you provide the data in the supplement?

No. The µ-CT test pits were drilled into a different Sri Lankan megacryst, which is not used as a reference material. We will clarify this and add the data to the supplement.

l. 126: Not all analytical protocols place the LA-ICP-MS spot over the He pit; some place a smaller LA-ICP-MS spot inside the He pit (e.g. Evans et al., 2015; Glotzbach & Ehlers, 2024). I think mentioning this and briefly elaborating on the benefits and drawbacks of both configurations would provide readers who consult this manuscript as a resource for a complete analytical LA-(U-Th)/He workflow with valuable context.

We agree and will briefly describe both configurations. A smaller ICP-MS pit inside the He pit (e.g., Vermeesch et al., 2012; Evans et al., 2015; Glotzbach and Ehlers, 2024) requires the He to be measured last. Since by then the mount has been under a high-He atmosphere (due to the He carrier gas in the ICP-MS), which significantly increased its blank in tests we conducted.

l. 127: You state that you increase the spot size for the ICP-MS measurements by 30 to 40 μm . The manuscript would benefit from an explanation for the reason for this variation and what your spot size choice depends on. Further, could you please include the U-Th pit diameter in the supplementary data table?

In this study, the ICP-MS spot was 50 μm for all analyses, i.e., 30 μm wider than the He pit (see our reply to Reviewer 1, L. 127). We will replace the range in the text with this value and add the diameter to the data table.

l. 128: Please provide the repetition rate and the number of pulses and/or the duration of ablation.

Added (10 Hz for 40 s, i.e., 400 pulses, after 20 s of background measurement).

l. 129: Please double-check that 50 μm is the minimum required zircon size. If you polish away 20 μm of the grain, ablate a 15 μm He pit, and add a U-Th pit that is 25 μm deeper than the He pit, don't you need at least a 60 μm grain to avoid ablating into the embedding material?

We agree. Besides the lateral constraint discussed in our reply to Reviewer 1 (L. 129), the grain must be thick enough to contain the ICP-MS pit. With $\sim 20 \mu\text{m}$ removed by polishing and ICP-MS pits of $\sim 39 \mu\text{m}$ depth in this study, grains need to be at least $\sim 60 \mu\text{m}$ thick. We will correct the minimum grain size.

l. 133: According to lines 211 and 212, you measure ^{147}Sm and ^{235}U . Please include them here in the list of parent isotopes.

Added (see also our reply to Reviewer 1).

l. 135: For workflow reproducibility, it would be useful if you could provide an example of your ablation sequence in the appendix or supplement. How often and where in the sequence do you measure the NIST 610 standard, the zircon standards and line blanks?

NIST SRM 610 and GJ-1 were each measured in blocks of four spots every 50 analyses. For the ICP-MS, the background is measured at the start of each analysis before the laser fires, so no separate blanks are run. For the He line, a blank was measured after every five analyses (see our reply to Reviewer 1, L. 594). We will describe the sequence in Appendix B.

l. 137: Which zircon standard(s) did you use for your LGC-3 measurements? Please also mention the GJ-1 zircon.

LGC-3 was analyzed as an unknown in the same sequences as the FCT spots, with the same standards (NIST SRM 610 and GJ-1). We will add GJ-1 here (see also our reply to Reviewer 1, L. 222).

l. 185: See my comment on line 126.

See our reply to the comment on L. 126.

Fig. 4: I am surprised by your choice to compute the probability clouds for a pit diameter of 20 μm , even though you used 30 μm for the FCT measurements. Displaying results for the pit geometry used in your measurements would be more intuitive and thus beneficial to the reader.

Fig. 4 already uses the 20 μm diameter of our He pits; the 30 μm in L. 88 was an error (see our reply to L. 88). We will change the depth from 5 μm to the median FCT He pit depth of 7 μm , as also requested by Reviewer 1.

I. 225: You only mention performing the machine-learning analysis on your FCT dataset. Did you also perform the analysis for LGC-3? If not, why? In my view, comparing results from FCT and LGC-3 would strengthen your conclusions about factors influencing the age dispersion. Presumably, it would allow you to better distinguish between grain-specific effects and “workflow effects”.

See our reply to the third general comment.

I. 234: I understand that you exclude the He, U and Th concentrations from the predictor set because they connect to the age equation. However, I don't understand why you then include Sm in the predictor set. I also wonder whether the overcompensation problem that concerns the parent and daughter nuclide concentrations applies to the He pit volume as well. As the He concentration depends on the He pit volume, could the model, indirectly through the pit volumes, still be compensating for the dispersion via the age equation? Could this explain the strong He-pit-volume signal? A more detailed explanation would be beneficial to the clarity of the manuscript and the interpretability of the machine learning model results.

Sm is kept because its contribution to the He budget of zircon is negligible, so the model cannot use it to correct ages through the age equation. In the FCT data, Sm is below detection in 60% of the spots, and ^{147}Sm produces at most 0.2% of the ^4He (median 0%). It serves only as a chemical indicator of inclusions. Regarding pit volume, see our reply to Reviewer 1 (third general comment). Unlike He, U, and Th, pit volume alone does not define the age, so it can only predict the age residual if volume-related errors exist. A model without pit volume and depth shifts part of their information to a correlated proxy (time of He analysis) but does not reveal other predictors. We will explain this in Sect. 2.6.

I. 247: Please elaborate on the choice of the 33 parameters you tested, as it is not clear to the reader. Why did you choose this selection? Further, you mention that several predictors correlate strongly with each other. I wonder whether, or rather how, this influences your results. Elaboration on these questions would provide valuable context.

The 33 parameters are the per-spot quantities from our data reduction that describe pit geometry, ICP-MS chemistry, mineral phases, zonation, U-Pb context, the He measurement, and the time of analysis, i.e., the factors proposed as causes of dispersion, excluding the quantities that define the age. Correlated predictors share information. Shuffling one of them then underestimates its permutation importance, because the information remains available in the others, and SHAP values may split or shift credit between them. For this reason, we also evaluate whole predictor groups and tested the effect of removing the strongest predictor (see our reply to Reviewer 1, third general comment). We will add this to Sect. 2.6.

Results

I. 254: Please mention which of the datasets you are referring to here for clarity.

We will specify that this refers to the $\mu\text{-CT}$ benchmark dataset (Appendix C).

I. 255: What pit size range (in μm) do you mean by “range of pit sizes commonly used for LA-(U-Th)/He”? Please specify.

About 10–50 μm in diameter. We will specify this.

Fig. 5: I assume the dataset plotted corresponds to the measurements detailed in Appendix C. However, the range of pit sizes plotted here does not match the range of pit diameters and depths given in Appendix C. Could you please check? Additionally, please consider changing the colour scale as it makes the figure difficult to interpret. Specifically, I find it difficult to distinguish similar-sized He spots in the upper plot and to correlate pit diameters with their corresponding depths in the lower subplot. Lastly, please specify whether “spot size” in the upper plot refers to the radius or the diameter.

Yes, Fig. 5 shows the Appendix C dataset; we will correct the ranges given in Appendix C. We will also use a color scale with more distinct colors and state that the spot size is the diameter.

l. 259: Did you only analyse the pit configurations mentioned here? In Appendix C, you describe ablating a wider range of pit diameters and depths. Were these not analysed? Also, could you elaborate, where fitting in the manuscript, on why you chose to analyse 40 μm pits instead of 30 μm pits, as used in your FCT measurements?

All pits shown in Fig. 5 were analyzed (total σ_{MAD} 4.7%); L. 259 lists only the three largest populations. The $\mu\text{-CT}$ measurements were made before we settled on the pit size for dating. We chose 40 μm to be comparable to our ablations but large enough to be resolved by $\mu\text{-CT}$. We will add this to Sect. 3.1.

Fig. 6: I suggest presenting the effects of the different reduction methods for LGC-3 as you do for FCT. Comparing the effects of the different reduction methods on zoned and pristine material would strengthen your analysis and argumentation, especially regarding the effect of your parent nuclide depth-weighting approach. Also, please indicate in the figure which is the median and which is the σ -normalised median, and consider adding a legend or label for the hashed lines and shaded areas that presumably indicate the reference age and its uncertainty.

We will add a second row with the LGC-3 ages for all four reduction schemes (see our reply to Reviewer 1, Fig. 6). We will also label the median and σ_{MAD} values and add a legend for the dashed lines and shaded bands, which show the reference ages and their uncertainties.

l. 267: Could you elaborate on why you use 2% as global systematic uncertainty and not 5.5-6%?

The 2% describes the precision of the interferometric measurement itself, based on repeated measurements of the same pits. The 5.5–6% also includes the uncertainty of the $\mu\text{-CT}$ volumes and pit-to-pit differences. On the flat, homogeneous LGC-3 crystal, 2% is sufficient, as the 100 spots scatter by only 1.2%. For natural grains, we recommend an additional external uncertainty of ~5.5–6% (see our replies to Reviewer 1, L. 270 and L. 371).

Fig. 7: Consider presenting the same compilation for LGC-3. In my opinion, it would aid your discussion of dispersion (or lack thereof) induced by the analytical workflow versus grain- or sample-specific properties. Please add units for the U and Th concentrations.

We will add the units ($\mu\text{g/g}$). LGC-3 is a single homogeneous crystal, so its concentrations and zonation metrics vary little. We will summarize them in one sentence in Sect. 3.3 instead of adding a figure.

l. 339: Do you report these data?

Yes. The out-of-fold predictions and residuals of every spot are part of the model output and will be included in the supplement.

l. 354: See my earlier comment on line 234. Could this be due to the model design?

See our reply to the comment on L. 234.

Fig. 9: Please consider colouring the KDE areas and the right y-axis differently from the PDP curve and the corresponding left y-axis to improve plot readability. If possible, change μm^3 to μm^3 for the He Pit Volume unit in the upper-left plot and consider relabelling the y-axis or clearly defining “resid_singed_ref_ma” in the caption.

We will change the colors and the unit and relabel the y axis as “Predicted age residual (Ma)”.

I. 366: I wonder whether Rutile%, Apatite%, Al Phase % and Fe Phase % correlate too strongly with Ti49 conc., Al27 conc., Ca44 conc. and Fe57 conc. and therefore are “ignored” by the model, leading to misleading outcomes. Could you explain how including strongly correlated features influences the SHAP analysis?

Each phase fraction is calculated from one of these concentrations (we measure ^{42}Ca), so the two are perfectly rank-correlated (Spearman $r_s = 1.00$). The trees then use only one of them, here the concentration, and the phase fraction receives zero importance in both the permutation and the SHAP analysis. The zero values therefore reflect redundancy, not missing information. The information itself is small, as the concentrations also have low importance and removing the whole mineralogy group does not change the model performance. We will explain this in Sect. 3.4.

Discussion

I. 393: Please explain to the reader what you mean by “popcorn”.

We will explain the term here (melt droplets at the bottom of the pit; Appendix A2).

I. 394: It is unclear which dataset you refer to here. Is this a dataset different from the 100 LA-(U-Th)/He measurements in LGC-3 and the 291 measurements mentioned in Appendix C, or is there overlap?

It refers to the 100 LGC-3 LA-(U-Th)/He spots shown in Fig. 6, measured as a 10×10 grid on one polished crystal (see also our reply to Reviewer 1, L. 394). We will clarify this.

I. 420: I don't think this is unexpected, but rather an expected result of your analytical approach, as you mention later in line 425. Have you tested your machine learning approach on a dataset where you did not apply your parent-source depth weighting and compared the results? It would be interesting to see if the zonation group's impact is larger in the dataset not using parent-source depth weighting.

We agree and will rephrase. Following the suggestion, we repeated the analysis with ages calculated without depth weighting, using the same predictors and partitions. Without depth weighting, the zonation metrics carry about twice as much information (summed permutation importance 0.099 vs. 0.046; Figure R1), and removing them increases the unexplained σ_{MAD} by 0.28 ± 0.11 Ma, compared with 0.10 ± 0.13 Ma with depth weighting. Depth weighting therefore removes more than half of the zonation signal, as the reviewer expected. In both cases, removing the zonation metrics lowers the explained variance by 7–8 percentage points, so the remaining signal mainly concerns the largest residuals (see our reply to Reviewer 1, third general comment). ^{208}Pb also gains importance without depth weighting, possibly because inclusions in the deeper part of the ICP-MS profile receive full weight in a simple average. We will add this comparison to Sect. 4.3.

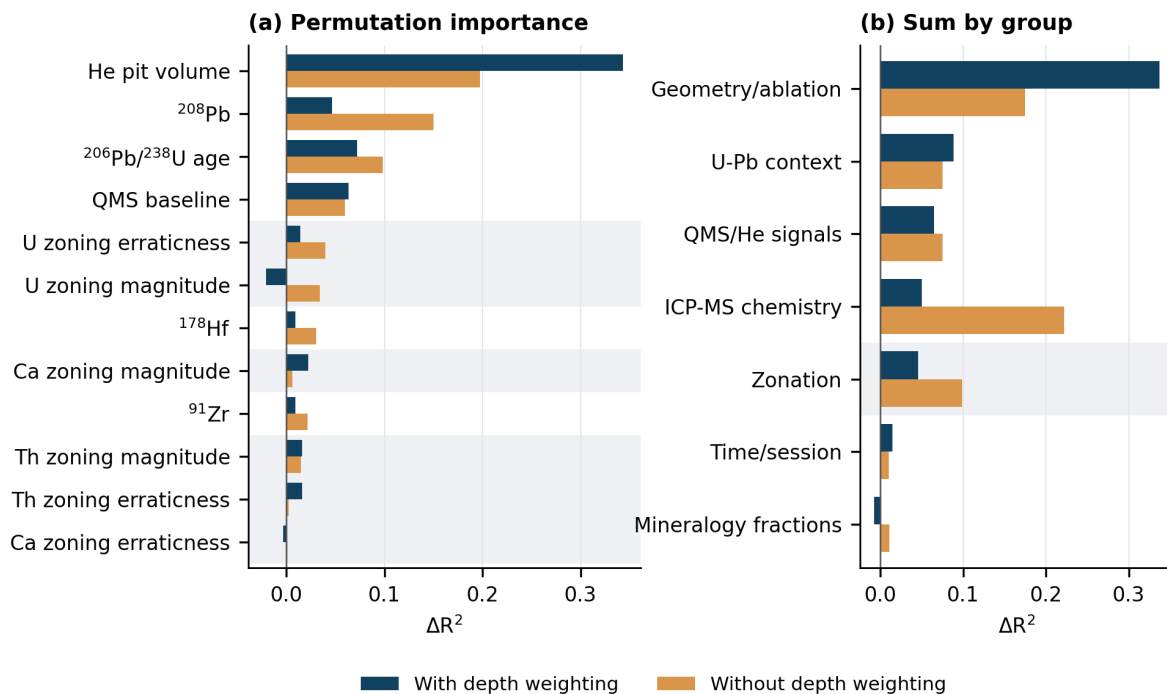


Figure R1. Permutation importance of the main predictors (a) and summed by predictor group (b) for models trained on FCT ages with (blue) and without (orange) parent-source depth weighting. Both models use the same predictors and cross-validation partitions. The zonation predictors are shaded.

l. 458: I presume this effect also depends on the width of the pits and the configuration of the He pit and the LA-ICP-MS pit. Also, please mention that your modelling does not account for the effects of He diffusion in the grains. While it arguably does not have much of an effect on FCT specifically, it is still relevant information as diffusion leads, among other factors, to a dependence of the measured in situ age on the ablation spot location in the grain (e.g. Glotzbach and Ehlers, 2024).

We agree. Under the layered-cake assumption, the pit diameters do not affect the result (see our reply to Reviewer 1, L. 451), but with lateral zonation in real grains they do; we will mention this. We will also state that the model does not include He diffusion, which is negligible for the rapidly cooled FCT but can make in situ ages depend on the spot position in slowly cooled samples (Glotzbach and Ehlers, 2024).

l. 464: Do the pit dimensions in the modelling exercise correspond to the pit dimensions used for the FCT measurements? The pit diameters are not indicated.

Approximately. The model uses a 5 μm He pit and a 25 μm ICP-MS pit (both measured from the polished surface), whereas the FCT pits have median depths of 7.4 μm and 39 μm. The diameters do not affect the model under the layered-cake assumption (see our reply to Reviewer 1, L. 451). We will state this in the text.

Fig. 11: Figure caption: Is the ICP-MS pit depth measured from the He-pit bottom or the grain surface? What are the He and ICP-MS pit diameters?

From the polished surface. The pit diameters do not affect the model under the layered-cake assumption (see our reply to Reviewer 1, L. 451). We will add both points to the caption.

Appendices

l. 531: Do you try to match the size of the glass spheres to the size of the grains you embed, or do you use the same glass sphere size for all samples?

We use the same 177.5 μm spheres for all mounts. The polished depth is calculated from the exposed sphere diameter, which changes fastest with depth near the top of the sphere, so larger spheres give a more accurate depth. Spheres much larger than the grains, however, are not fully pressed into the PFA. The 177.5 μm spheres are a compromise between the two. We will add this to Appendix A.

l. 533: Have you quantified the proportion of He potentially lost/degassed during hot-pressing?

We did not measure it directly. With the zircon He diffusion kinetics of ZRDAAM (Guenther et al., 2013) and the alpha doses of the FCT grains ($\sim 10^{16}$ to 10^{17} α/g), pressing at 320 °C for 60 s causes a He loss of at most $\sim 0.3\%$ (median $\sim 0.05\%$), limited to the outer ~ 30 nm of each grain. This rim is removed by polishing. The agreement of the FCT ages with the reference age also argues against a significant loss.

l. 544: See my comments on lines 78 and 79.

See our replies to L. 78 and L. 79.

Fig. B2: Are those pit depths you would recommend/ encounter in typical analyses, or is it more of a hypothetical modelling exercise? A more detailed discussion of the benefits and drawbacks of the presented pit geometries would be beneficial to the manuscript. Other notes: Please change “ 4He ” to “ ^4He ” in the caption. As far as I can tell, the corresponding ICP-MS pit sizes are not indicated.

Fig. B2 is a modelling exercise over a range of He pit geometries; the geometry used for our data (20 μm diameter, ~ 7 μm depth) is shown in Fig. 4. We will add a few sentences on the trade-offs. Narrow, deep pits lose less of the parent-source region to polishing but yield less He and are harder to measure by interferometry, whereas wide pits yield more He but lose more of the parent-source region. We corrected the superscript. Under the layered-cake assumption, the weighting curves do not depend on the ICP-MS pit diameter, which we will state in the caption.

l. 778: Please provide details for the acquisition, specifically the laser ablation settings.

The test pits were ablated with the same laser settings as the He pits (Sect. 2.2), except for the spot diameter. We will add this.

l. 818: Please explain the choice of parameters (tree depth, learning rate and boosting iterations). Are these default values or are they specifically tuned for your FCT dataset?

The settings were chosen after limited manual testing on the FCT data. To show that the results do not depend on this choice, we reran the model with the scikit-learn default settings (100 iterations, learning rate 0.1, unlimited depth), with shallower trees (depth 3), and with a higher learning rate and fewer iterations (0.1 and 300). The OOF variance explained ranges from 27.7% to 30.0% (28.7% with our settings), and the four strongest predictors (He pit volume, $^{206}\text{Pb}/^{238}\text{U}$ age, QMS baseline, and ^{208}Pb) are the same in all runs. We will add this to Appendix D.

Technical corrections

General comments: 1. Please define the abbreviations used in the abstract. 2. Please be consistent in the use of “ α ” or “alpha” with respect to α particles, stopping distances and ejection and please also be consistent in the use of micro-CT or $\mu\text{-CT}$. 3. Please always define the variables used in your equations. 4. Please ensure that the font size in each figure is large enough to be readable.

We will define all abbreviations, use α and $\mu\text{-CT}$ consistently, define all variables, and increase the font sizes where needed.

l. 8: Please define/ briefly explain “ σ_{MAD} ” here.

Added.

l. 21: “emission of α -decay” is odd wording. Consider changing this to “emission of α -particles” or “emission of α -radiation”

Changed to “emission of α -particles”.

l. 24: Consider adding a reference for α -ejection correction (e.g. Farley et al., 1996).

Added.

l. 47: Please define or spell out “UHV”. You define it later, in line 69, but ideally abbreviations are defined the first time they appear.

Added.

Fig. 1: Please consider labelling the equipment depicted. Change ^3He to ^3He in the figure caption.

We will label the main components and correct the superscript.

l. 59: zircon -> zircons

Corrected.

l. 80: Please be consistent with using lowercase or capital letters when referring to figure subplots. You use a capital letter in line 480.

Corrected.

l. 88: Remove “a” in “a 30 μm diameter spots”.

Corrected.

l. 120: “voxel resolution between 1.0 μm and 3.4 μm ” implies you also used resolutions other than 1.0 μm and 3.4 μm . However, as I understand Appendix C, you only used those two. Consider rephrasing to avoid confusion.

Rephrased to “at voxel resolutions of 1.02 μm and 3.41 μm ”.

l. 130: Please define HR-ICP-MS.

Added.

l. 143: Please check if “gas time-resolved gas evolution” is what you wanted to write here.

Corrected to “by monitoring the time-resolved gas evolution”.

From l. 283: This section is very discussion-heavy and, in my opinion, better suited for the discussion section than the results.

We agree and will move this part to the discussion (see our reply to Reviewer 1, L. 283–295).

Table 1: Is “scaled MAD” equivalent to σ_{MAD} ? Please be consistent with the naming.

Yes. We will use σ_{MAD} throughout.

Fig. 8: Change μm^3 to μm^3 for “He Pit Volume”, and please be consistent with writing the atomic number in front of or after the chemical symbol. Does % as a unit for “Hf178 Conc”, “Zr91 Conc” and “Si29 Conc” correspond to wt%?

We will correct the unit and the isotope notation. Hf and Si are given in wt%; the Zr values are in $\mu\text{g/g}$ and were mislabeled (see also our reply to Reviewer 1).

l. 404: ^{208}Pb -> ^{208}Pb

Corrected.

Fig. 10: Presumably, pit depth and diameter refer to the He pit (and not ICP-MS pit). Please be specific.

Yes. We corrected the axis titles (see also our reply to Reviewer 1).

l. 497: ^{208}Pb -> ^{208}Pb

Corrected.

Fig. A1: “Items are for ultra-high vacuum use; please minimize the use of oil” Please add the hyphen and the semicolon to the text as indicated.

Corrected.

Fig. A2: Are the dimensions given in mixed units (inches and mm)? If so, please unify.

The flange and tube sizes are standard designations in inches; all other dimensions are in mm. We will state this in the caption.

l. 651: Do you mean “main pit bottom depth” or “mean pit bottom depth”? Please check.

“Main” is correct. It refers to the most frequent depth of the pit bottom (Appendix B2, step 9), which we will clarify.

l. 721: parent concentration -> parent nuclide concentration

Corrected.

Fig. B1: Some labels overlap. Please ensure that they are all readable.

We will fix the overlapping labels.

Figs. C1 & C2: Please add a legend or labels, and add a scale to the CT image.

We will add labels, a legend, and a scale bar.