Review of Ehrenfels et al.: Analytical improvements and machine-learning assessment of age dispersion in laser-ablation (U-Th)/He zircon dating (submitted to Geochronology)

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.

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.

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.

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.

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*.

<u>Detailed line-by-line comments and questions:</u>

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.

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.

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.

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.

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

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.

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.

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.

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.

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

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?

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.

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?

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

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?

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.

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?

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

l. 185: See my comment on line 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.

l. 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”.

l. 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.

l. 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.

Results

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

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

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.

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?

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.

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

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.

l. 339: Do you report these data?

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

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 µm3 to μm^3 (superscript) 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.

l.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?

Discussion

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

l.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?

l.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.

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).

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.

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?

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?

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

l.544: See my comments on lines 78 and 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 “^{4}He” in the caption. As far as I can tell, the corresponding ICP-MS pit sizes are not indicated.

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

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?

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 µ-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.

l. 8: Please define/ briefly explain "σMAD" here.

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

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

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.

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

l. 59: zircon -> zircons

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

l. 88: Remove "a" in "a 30 µm diameter spots".

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.

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

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

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

Table 1: Is "scaled MAD" equivalent to σMAD? Please be consistent with the naming.

Fig. 8: Change µm3 to $µm^3$ (superscript) 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%?

l. 404: 208Pb -> ^{208}Pb (superscript)

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

l. 497: 208Pb -> ^{208}Pb (superscript)

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.

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

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

l. 721: parent concentration -> parent nuclide concentration

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

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

References

Evans, N. J., Mcinnes, B. I. A., Mcdonald, B., & Dani, M. (2015). An in situ technique for (U – Th – Sm)/ He and U – Pb double dating. J. Anal. At. Spectrom, 30, 1636–1645. https://doi.org/10.1039/c5ja00085h

Farley, K. A., Wolf, R. A., & Silver, L. T. (1996). The effects of long alpha-stopping distances on (U-Th)/He ages. Geochimica et Cosmochimica Acta, 60(21), 4223–4229. https://doi.org/10.1016/S0016-7037(96)00193-7

Glotzbach, C., & Ehlers, T. A. (2024). Interpreting cooling dates and histories from laser ablation in situ (U – Th – Sm) / He thermochronometry : a modelling perspective. Geochronology, 6, 697–717. https://doi.org/https://doi.org/10.5194/gchron-6-697-2024

Glotzbach, C., Neely, A., & Ehlers, T. A. (2026). In situ apatite U-Pb , fission track and (U-Th) / He triple dating using a simple embedding approach. Geochronology, 8, 567–588. https://doi.org/https://doi.org/10.5194/gchron-8-567-2026

Hoemann, H., Hueck, M., & Dunkl, I. (2026) Technical Note: Reducing age dispersion in in-situ (U-Th)/He dating: How to optimize ablation pit geometries with OptiPit, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-2590

Maitra, A., Härtel, B., Enkelmann, E., & MacLeod, D. (2026). Implications of uranium and thorium zoning for zircon laser-ablation (U-Th)/ He age interpretation. Chemical Geology, 724. https://doi.org/10.1016/j.chemgeo.2026.123694

Metcalf, J. R., & Flowers, R. M. (2026). Technical note : Design , construction , automation , and first-principles calibration of a low volume He measurement line. Geochronology, 8, 447–462. https://doi.org/https://doi.org/10.5194/gchron-8-447-2026