Review by Ray Donelick of Tamer et al. "The need for fission track data transparency"

General Comments

- The data presented here are very important and need to be published. My strong preference is to "Accept with Major Revisions" this submission. Lacking major revisions such as suggested here, I urge this paper be Rejected by the Editor, re-written by the Authors, and re-submitted for review.
- Major revision required: The document and its arguments are poorly organized and the English wording and grammar needs to be improved significantly.
- Suggestion regarding re-organization:

  Lines 11-12: "We report a new image-based inter-analyst study to investigate fission-track grain selection and analysis by 13 participants from an image data set that included grains of variable quality."

  From the perspective of the inter-analyst study, organize the data, results, and recommendations around these 6+ "essential ingredients" or essential steps:
  (1) Select suitable apatite grain for age measurement
  (2) Select region of interest for fission track counting
  (3) Count fission tracks intersecting ROI surface
  (4) Measure confined fission track lengths
  (5) Measure Dpar, Cl, or other kinetic parameter
  (6) Measure uranium concentration
  a. EDM – Count induced fission tracks corresponding to ROI surface
  b. LAICPMS – Measure representative area of ROI surface
  +
  (7) Collect, archive, and share digital images to measure and permit re-measurement of fission track data

- I have not reviewed figures, tables, and their captions, pending major revision.
- I have not reviewed supplementary materials, pending major revision.
- I am willing to review a revised manuscript.

Title

- A new title is suggested: The Need for Apatite Fission Track Data Transparency and Sharing

Abstract

- Emphasize this study concerns apatite fission track data.
- Define data transparency and how you propose to implement it.
- Define data sharing and how you propose to implement it.
- Include list of 6+ "essential ingredients" or essential steps here and argue around them

Lines 16-17: "there is a danger of "squeezing the rock" weakening selection criteria." "squeezing the rock" presumably comes from "squeezing blood from a rock" (a saying recognized in USA at least). Please drop this statement. What you mean by "squeezing the rock" is actually more like "filling a line in a table".

Line 18: This statement "Juxtaposing selected regions of interest (ROIs) on the same grains indicates that zoned grains and grains with inclusions and defects yield varying track density estimates, indicating that ROI placement can be an influential factor." is just one of many statements that can be made here. Either list all of them or drop this one.

1 Introduction
- Emphasize this study concerns apatite fission track data.

Line 35: First mention of "apatite", yet the data presented here are from only apatite, and several "essential ingredients" are dominated by apatite studies.

Lines 28-30: I suggest this list of 6+ "essential ingredients"
(1) Select suitable apatite grain for age measurement
(2) Select region of interest for fission track counting
(3) Count fission tracks intersecting ROI surface
(4) Measure confined fission track lengths
(5) Measure Dpar, Cl, or other kinetic parameter
(6) Measure uranium concentration
a. EDM – Count induced fission tracks corresponding to ROI surface
b. LAICPMS – Measure representative area of ROI surface
+
(7) Collect, archive, and share digital images to measure and permit re-measurement of fission track data

Lines 29-30: "(4) average etch pit diameter (Dpar) measurements per grain," re-word to "(4) mean etch figure diameter parallel to c-axis (Dpar; Donelick, 1993; Burtner et al., 1994; Donelick et al., 1999) for each apatite grain," I hate being the guy pushing his own papers, but these papers are appropriate, especially given the >10 years Dr. Gleadow denied the value of this parameter (and its sister Dper).

Donelick, R.A., Ketcham, R.A., and Carlson, W.D., 1999, Variability of apatite fission track annealing kinetics II: Crystallographic orientation effects. American Mineralogist, v. 84, pp. 1224-1234.

Burtner, R.L., Nigrini, A., and Donelick, R.A.,1994, Thermochronology of Lower Cretaceous source rocks in the Idaho-Wyoming thrust belt. American Association of Petroleum Geologists Bulletin, vol. 78, no. 10, pp. 1613-1636.

Donelick, R.A., 1993, A method of fission track analysis utilizing bulk chemical etching of apatite. U.S. Patent Number 5,267,274.

Lines 31-35: "While laser ablation mass spectrometry has become an alternative (Hasebe et al., 2014) to the widely used external detector method (EDM) (Gleadow and Lovering 1977) for uranium content determination, the first four inputs are still largely analyst-driven, although recent developments in image analysis and AI have contributed significant advances in auto-counting and auto-measurement (Gleadow et al., 2009, 2019; Nachtergaele and De Grave 2021; Li et al., 2022; Ren et al., 2023; Boone et al., 2023)." Break this sentence into 2 or more sentences to make these points.

Line 36: "significant variation in measurements for the same samples and even standards" More information is needed here. The abstract in Line 1 leads off introducing your new "inter-analyst study" so you need to compare new to old. Help the reader better understand which of the "6+ essential

ingredients" might be the source of "significant variation" here or there in previous and the current work.

Line 43: "Grains where oily fluids have penetrated" It is not only "oily fluids", but also aqueous fluids. My experiences is that a paper towel alone cannot be guaranteed to remove distilled water in tracks from washing after etching.

Lines 46-47: "cause overestimation of ages." Or underestimation of the presence of defects here and there (say in Durango) may cause the analyst to lean toward "defect" for questionable features.

Lines 52-53: "Whereas the area counted for fission-track density determinations has typically been defined by boxes in an eyepiece reticule,…" Replace "typically" with "historically". "…modern image-based systems allow the user to draw an arbitrarily shaped region of interest." Replace "modern" with "recent". What is a reticule? The original sentence is condescending.

Line 56: "geometry (Donelick et al., 2005)" A better reference would be Fleischer et al. (1975).

Fleischer, R.L., Price, P.B., Walker, R.M., 1975, Nuclear Tracks in Solids: Principles and Techniques. University of California Press, Berkeley, 605 p.

Lines 64-65: "The suggested number of grains for age measurements for igneous-type samples is typically ~20, or more if there is any indication of kinetic variation (Donelick et al., 2005)" Donelick et al. (2005) did not suggest 20 grains ages, but merely stated that it was common practice. The source of 20 grain ages is Dr. Gleadow, with the choice of 20 grain ages being more concerned with making money (minimal work for Geotrack in the 80s) and less concerned with science (getting a pooled age of desired quality and precision).

Line 66: "squeeze the rock" Drop this saying that may make little sense to some people. Instead, focus on the poor data resulting from poor decisions.

2 Materials and Methods
Line 77: "41 grain and 3 graticule images" What do you mean by grain? What do you mean by graticule images? The word apatite does not appear in this paragraph!

Line 82: "grains from UM were etched with 5M HNO3 at 20°C for 20s (Gleadow et al. 1986)." I am pretty sure Dr. Gleadow has been under-etching c-axis-parallel fission tracks in F-rich apatites since the late 1970s. Please put the correct reference here.

Line 88: "and in relevant email lists". I was not included in your list. Who decides who gets invited?

Lines 88-100: Drop the sales pitch and tell us what this software does, and how the "experiment was made possible" with this software. I can do this study and much more with my own software, so I don't need Dr. Gleadow to enable me and my research. This reminds of the Iolite bait-and-switch.

Lines 105-106: "Rather they are simply used as reference values that are probably typical of reasonably experienced analysts." This paragraph is difficult as you do not want to tell everyone that your judgements are correct and those that deviate from yours are incorrect. Give the reader evidence here, right now, that you are qualified to make this decision. Show a zeta calibration with 100 grains of

Durango or something like that! Provide those images too. We should be requiring this routinely, you basically argue this point, and you do not provide the evidence.

Lines 113-114: "was considered unsuitable" By whom? I assume the answer is "L. Chung and… M. Tamer" from Lines 103-104.

Line 119: "borderline-quality grains" Borderline is not defined here. Suitable is not defined here. This whole paragraph needs to be flipped, tell us what makes a grain suitable, what is compromised for borderline, and then finish with what makes a grain absolutely unsuitable.

Line 154: "Graticule images" Really? Images of the graticule in the eyepiece? Perhaps you mean images of a NIST-traceable length calibration grid on a microscope slide?

Line 155: "only five participants reported measuring them." Well, did they get similar results to the default graticule calibration? More info please.

Lines 161-162: "varying region of interest selection, light source preference, and track counting routines" You discuss varying region of interest selection. What do you mean by light source preference and do you have data to back this up? Same question for track counting routines, after telling us what you mean by track counting routines.

Lines 180-181: "This high rate of acknowledgment by the participants supports the soundness of the criteria utilized by the reviewers." Maybe. Who cares. What matters is that there is room to educate each other in this field and to, perhaps, lower variation among labs by abiding by the principles of data transparency and data sharing.

Lines 185-186: "Some of the participants used FastTracks' automatic tools for c-axis orientation and dpar length measurements." This is important information out of nowhere. You need to separate out the effects of these measurements from those who did not use these tools. Also, you should show how well FastTracks reproduced measurements from analyst to analyst that used these capabilities. The ultimate goal is to lower variance among analysts. Did FastTracks succeed or fail here?

Lines 196-197: "Participants 1 and 10 and Participants 8 and 9 are from the same two laboratories and show similarities in their respective track density results." Give numbers here and elsewhere in this discussion that back up your statements. Don't make me search for this information in the tables and figures.

Lines 216-228: The discussion does not give any numbers telling the reader what is meant by "relatable and consistent", "skewed… to lower values", "varying number of measurements", "show similarities", "divergent results". Please use the results to make your case. These generalized statements teach me nothing.

Line 236: "Donelick et al., 2005" A better reference is Fleischer et al., 1975.

Lines 256-257: "However, counting tracks solely in transmitted-light images can cause an underestimation of the track density (Aslanian et al., 2022; Tamer and Ketcham 2023)." The word "can" does not mean "does". And I don't' need Dr. Ketcham telling me how to count. What matters more is that analysts apply the same methods/criteria/data types to unknowns as they do standards. The

argument can he made that data from this whole paper need to be divided by data from the next-inter-analyst study of appropriate age and length calibration standards.

Line 268: "grains etched with the 5.0 M HNO3 20s 20◻C (Gleadow et al. 1986) protocol appear to be under-etched" Because they are. They have been since 1977 or so. They continue to be.

Lines 27-271: "Analysts may consider unsuitable grains and/or conduct invalid confined track length measurements depending on their years of experience, training, and the difficulty in finding sufficient grains to meet analytical goals." This sentence needs to be re-written so that the several points being made are clear to the reader.

Line 271: ""Squeezing the rock"," Drop this saying.

Line 274: "Results of graticule and confined track length calibrations and the identity of the analyst should be stated in publications." In this paper, you offer graticule calibrations somewhere. You do not offer any confined length calibration data. You don't even mention, much less offer, any age calibration data such a zeta calibration standard. I would like to see data here divided by the appropriate calibration data.

Line 280-281: "precise matching of spontaneous and induced track areas in the EDM can also be difficult in some cases." I would love to sort through the decades of mica detectors affixed to under-etched AFT grain mounts at UMelbourne and elsewhere and reveal the staggering percentage of EDM images that are poor due to poor contact – but counted anyway to produce a line in a data table.

Lines 306-307: "Although fission-track data have generally fared well in inter-laboratory age comparisons in recent years" My assessment is just the opposite. The variance among laboratories is increasing, not decreasing, since the 1980s. This is almost certainly due to inconsistent – perhaps even poor at times – training of analysts, at the start and as the years go by. This is made easier by flashy hardware and software products that give the appearance of expertise but do not substitute for it.

Line 316: "encouraging data transparency" Re-write to "encouraging data transparency and sharing".