Final author response on manuscript Confined fission track revelation in apatite: how it works and why it matters

Richard Ketcham and Murat Tamer

13 February 2021

The main thrust of the combined comments is that we need to spend more time setting the stage for the problem, including presenting the Tamer and Ketcham (2020) data and discussing more previous work on etching rates and why they are thought to vary, including etching rate being linked to the energy loss rate of the fission particle. We also need to more thoroughly describe how our model works, including a summary figure. In our initial submission we tried to go quickly through these initial parts to keep the paper relatively short, but evidently too quickly and not carefully enough. Other good suggestions are photos of developing track tips, and exploring model predictions on the dip distribution of confined tracks. All told, we anticipate this will require considerable new text, and 4-5 new figures, resulting in a substantially longer contribution, but we believe a more accessible and understandable one.

We have responded to the initial comments by Jonckheere and Green previously. We have transferred all subsequent comments below (except the final one by Jonckheere and Wauschkuhn (RC7), which is addressed to Green's remarks and not our contribution), and our answers are in *red italics*.

Answers to final comments by Raymond Jonckheere and Bastian Wauschkuhn (RC3)

Our essential comments are contained in our review (RC2), to which the authors have replied (AC2). The manuscript discussion forum invites us, or at least allows us, to react to these re-plies. In the absence of comments from other scientist, this could result in an endless back and forth between authors and reviewers until a consensus is reached or the deadline expires and the editors must arbitrate. It appears to us that this could be an exhausting and unproductive process, as it is most improbable that agreement will ever be reached on what is, or is thought to be, right or wrong. Since our scientific differences are not the real issue, we *recommend that the manuscript be considered for publication*, with such corrections as the authors and editors think necessary.

We are gratified that the referees recommend consideration for publication after corrections.

Nevertheless, we one last time wish to draw attention to some points, unrelated to scientific content. Various issues contribute to the fact that the manuscript makes an unfavourable impression on an attentive scientist. There is an apparent disdain for the reader, who is expected to guess at the intended meanings of word strings that appear to have been made up on the run, with indifference. This could be avoided by rewriting the manuscript with the aim to be understood.

It is inevitable that readers will evaluate the approach followed in this manuscript as a quick fix. The claim that the goal is to determine track etch rates is unconvincing. Price et al. (1973) and Green et al. (1978) performed *actual detailed measurements of v_T* along ion tracks in minerals more difficult to investigate than apatite. Furthermore, suspicious emphasis is placed on incidental geometrical observations. The road to v_T is paved with precarious assumptions and precipitous leaps. The most unlikely and limited data (the mean lengths of etched confined tracks accessed via surface-intersecting host tracks and an intervening apatite segment) are used for estimating the model parameters. The resulting v_T profiles have at best the most tenuous of connections with real fission-track or apatite properties. It cannot be otherwise, considering how they are arrived at. The computer-generated v_T-models are used to calculate etched-track geometries, of which a subset is selected for calculating the mean track lengths. Instead of that, the calculated geometries could have been compared with countless images of step-etched tracks, eliminating the need for inventing selection criteria. Within the current concept, it would in fact be most efficient to start from observed track geometries, and trace back their individual v_T profiles, i.e. perform the reverse calculation from that illustrated in Figure 5. The authors are doubtless aware that it does not work. This should however also alert them to the weaker links in their approach.

We are also grateful for their comments on pertinent prior work on documenting etching rates (Price et al. 1973; Green et al. 1978), and will include this work in a revised introductory section that will better set the stage for the work we present. It appears that the reviewers are referring to approaches their research group is using, which we continue to see as complementary.

The reader requires more detailed, unambiguous information. E*xactly* which cores does Table 2 describe? Fission tracks are created by nuclides with different masses, charges and energies, resulting in variable and anisotropic lengths. Are all tracks in one sample assumed to be the same? If not, are, e.g., the listed core lengths averages over all tracks in a single run? Must a set of related cores be assumed in order to allow for length variation, or is it an automatic outcome of the separation between the host and confined tracks and the locus of their intersections? How does length anisotropy emerge in an isotropic model? Are the cores perhaps scaled to different latent track lengths at the start; if so, how? Is the latent track envisaged as the traditional line segment of finite length, with variable v_T, but without structure, etching threshold or range deficit? Or are the v_T models in Figure 1 somehow also compatible with recent observations of unetched tracks?

Our revised introduction/background will put our linear and constant-core models into the context of variable particle energies and stopping characteristics, which tend to feature first gradual and then faster decreases in energy loss rate of the fission particles, which have been hypothesized (e.g. Price et al 1973) as being linked to along-track etch rate.

How is it that a model predicting step-etch data cannot be applied to single-step experiments? Most datasets indeed do not present lengths for consecutive etch times, but the model should nevertheless fit the data. If, on the other hand, the fits depend on the observation function of one analyst to the point that all other data are excluded from consideration, then what is the significance of four digit etch-rate estimates? How are the step-etch experiments dealt with? Are the emersion, rinsing of the etchant and etch products, and re-immersion in fresh etchant without effect?

Our model can be, and is, applied to single-step experiments (e.g. experiments TK20; EAE1,2,3), but single-step experiments do not contain enough information to constrain the model, unless perhaps one had a lot of different single-step experiments etched for different times. We already discuss the experimental uncertainties, and report appropriate confidence intervals.

The authors are modest in their replies *(this is a first step; all models are wrong but some are useful),* but this does not help their manuscript at all. To abuse their image: yesterday we stood at the edge of the precipice, but today we made a great leap forward. It would greatly benefit the reception of this manuscript if the authors addressed the questions that all trackers will puzzle about.

The authors are right that this is the first calculation of its kind. However, in our opinion, no-one will be interested unless they can bring themselves to address the readers' obvious concerns. This could turn this manuscript from a wild shot into a considered and considerate paper that professionals will *want to read* and will appreciate for its *thoughtful assessment* of its methods and results.

Replies

1. The *conceptual model* fitted to the Laslett et al. (1984; Figure 8) step-etch data (albeit not to strictly identical confined track samples) corresponds, in terms of the present manuscript, to an average constant-core model with Δx_{Tmax} = 11.1 μm, $\Delta x_{Tmax\text{-}B}$ = 2.1 μm, V_{Tmax} = 1.67 μm/s, and V_B = 0.04 μm/s.

We don't believe these parameters strictly match the Laslett et al. (1984) model; we plotted up our model with these parameters assuming instantaneous access of the acid to the tracks (which we think they are assuming), and the result does not resemble the figure (VTmax=1.1 μm/s works better). There are a number of factors that make the Laslett data difficult to model, the largest one being that they measured TINCLEs, making it difficult to estimate when the acid would reach the track – undoubtedly faster than for TINTs, but still requiring an estimate of V-cleavage. The Laslett data may also include tracks in later etching steps that were unobservable during the early ones. As for the conceptual model, we believe that the reviewers are over-interpreting that they connected their data points with a curved line, while ignoring what is actually stated in the text.

2. With some hesitation, the authors concur with the first reviewer that our geological context of the KTB is open to question. We wish that once, just once, a critic would come up with a single scientific fact.

3. Our last comment concerning the need for an appendix with model equations is not as bizarre as the authors think. More than 30 years ago one of us wrote a track-etch simulation program to investigate if the Dartyge et al. (1981) track model could account for some properties of etched confined tracks (it can). There was no host track and no distance to bridge to the confined track. The confined track itself was represented by a vector, with each position x corresponding to a fraction of a micron along the track that could be etched at a cost ~$1/v_T(x)$. This approach could handle discontinuous $v_T(x)$-profiles resulting from the local densities of extended and point defects, much more complex than those in the present manuscript, without needing to solve equations.

These replies are conversational in nature, and do not require responses, at least in the context of evaluating this manuscript. We do note that we included in this submission the complete computer codes for generating our results.

Answers to comments by Andrew Gleadow (RC4)

General Comments

This paper reports on new variable VT fission track etching models for apatite and considers in detail its implications for confined fission track length measurements and their application in thermochronology. The models involve a number of clearly acknowledged simplifying assumptions, such as omitting the effects of anisotropic etching and annealing, but nonetheless provides a number of important insights into track etching behaviour and explanations for a range of previously observed phenomena. The title is an apt description of the content. The models are then applied to an existing empirical track length data set and achieve considerable success in explaining their characteristics. In my opinion the paper represents an important advance in our understanding of fission track etching behaviour with significant implications for practical applications in fission track thermochronology.

We are grateful for the reviewer's endorsement.

In essence, the paper explores the consequences of varying the track etching rate, VT, along a fission track rather than being constant along its entire length. It does not purport to be a fully comprehensive model of the detailed form and distribution of etched tracks in apatite, along the lines of the models being developed by Raymond Jonckheere and his colleagues, but rather studies whether variable VT produces first-order predictions that can be usefully compared with observations. In this, I think it has been remarkably successful and has made a strong case that such variability is both real and has a significant influence on measurements. The particular form of the simplified models considered are somewhat arbitrary but conform broadly to what is known about track structure in apatite and other minerals. In fact, because the models do not include the specific crystallographic characteristics of apatite, the implications could probably also be generalised to track etching in other minerals.

We again are grateful for the comment, and particularly for pointing out the benefits of our general approach for applying to other minerals.

In order to understand the revelation of sub-surface confined tracks, the paper also considers the etching characteristics of surface semi-tracks from which they are progressively intersected in TINT measurements. The study represents the first quantitative exploration of the implications of this two-stage etching process, which has previously been understood only in the vaguest terms. This semi-track modelling component of the paper leads to its own conclusions, independently of their role in revealing confined tracks, such as the estimates made in 6.7 of the track counting efficiency for apatite. Doubtless these estimates could be refined by including crystallographic anisotropy in more complex models but the first-order agreement with independent estimates is very encouraging.

Another outcome of modelling the joint development of semi-tracks and confined tracks is to show that the great majority of confined tracks are intersected at shallow depths (section 4.3, Fig 4), with the consequence that most confined tracks will necessarily lie at low dip angles in order to avoid intersecting the surface. This explains why the usual restriction to measuring only 'horizontal' confined tracks, usually taken to mean dips of <10°, actually includes a high proportion of all the available tracks.

We thank the reviewer for this important observation as well.

It is noted in line 158 that a confined track dip of 25° is close to the maximum observed in the Tamer and Ketcham 2020b data set and a similar observation is a significant feature of our own dip measurements on confined tracks in a range of samples (Li et al, 2018 Amer. Mineral. 103, 430) which show that ~70% of confined tracks in a range of apatites have dips <10° and virtually all are <30°. The model also explains the observation of Li et al. that confined tracks with the greatest dips have slightly shorter mean lengths, probably due to a greater average intersection depth and later etching start times.

These are interesting and pertinent points, but we note that, in the models we present, we actually limit the maximum dip to 25°, and we do not explore the relative preservation of tracks as a function of dip angle. In the revision we can explore this a little further, but are doubtful that we can fully explain this empirical observation, because there is likely to be an analytical bias against looking for, finding, and being able to confidently measure confined tracks at high dips.

The first major conclusion of the paper that VT is indeed variable along the length of fission tracks is, I think, well-made and unlikely to be controversial, indeed such variability has long been accepted by many, but without any clarity as to the implications. The second group of conclusions about laboratory annealed tracks etching faster than either fossil tracks or freshly induced tracks is more surprising and unlikely to be readily accepted at face value. Given that these are based solely on the modelling of a very limited data set, and the simplifying assumptions involved, I would suggest at least a more cautious and tentative statement of these conclusions acknowledging such uncertainties. The result is certainly interesting and points to the need for further study to be certain that this is a general phenomenon and not an experimental or modelling artefact.

We believe we were cautious in our wording, but we can revisit it. However, we do note that the pattern of faster etching for annealed tracks does hold up over four unannealed and six annealed track experiments, each of which is essentially an independent test.

One very important conclusion is that the very poor reproducibility of track length measurements in previous interlaboratory comparisons is probably centred on just one factor – the criteria used by analysts to determine which tracks are acceptably etched for measurement. Previously, the observed discrepancies could be attributed to a range of possible analytical, calibration and training factors, compounded by the difficulty of standardisation in track length measurements. The new model results, however, suggest that attention should be focused primarily on standardising track selection criteria in order to reach a much-needed concordance in length measurements across different laboratories.

We agree.

Having said all this, I actually found the paper quite complex and surprisingly difficult to read in many places. I spent considerable time trying to understand some sections and some of the diagrams, despite considerable familiarity with the subject material. I think the overall readability needs to be improved if the paper is to be accessible to a wider audience. I make some specific suggestions below, but I think there are some general principles that could help.

First, a number of unfamiliar terms are used based on the geometry of the tracks which would be greatly clarified by an additional diagram where these are labelled. These terms include 'Impingement point' in Fig 2 (maybe 'Intersection point' would be clearer), the 'Relative semi-track' penetration and 'Confined track penetration' in Fig 3 and 'Intersection depth' (Fig 4). While these are defined at various places in the text, a diagram that showed them all in one place would greatly assist the reader in understanding the following diagrams. Perhaps this new key diagram might include a latent and etched semi-track and one or more confined tracks intersected at different stages of development, annotated with all the parameters used in the subsequent discussion.

We can add such a figure.

Second, some basic information is simply missing such as the etchant used – presumably this was 5.5M HNO3, but this is not actually mentioned anywhere in the paper. It is not reasonable for the reader to have to look up the original source paper to know this, especially when etching conditions are actually part of the discussion and the etching times are central to the text. There is also no indication of how the track densities might have differed between the different experiments and which ones were Cf-irradiated, which must exert a major control on some of the parameters, such as the number of intersections per track.

We can transfer this information from our previous paper, as also requested by other reviewers.

Third, the paper also assumes that the experiment codes (SE1, SE2 etc) are sufficient to alert the reader to the particular track characteristics involved. These are summarized in Table 1, but even there they do not appear to be complete. For example, only one of the experiments (SE3) in Table 1 is indicated to have involved Cf-irradiation, but it is apparent from the text that several others at least also involved this procedure. It is very hard to keep all of the different experimental details in mind when looking at the results, making them more difficult to comprehend. A good example is Figure 12, where histograms of the intersection depths are given for each experiment. There are clear differences between these, but it requires considerable work of detection to figure out how these relate to the track characteristics. Perhaps these histograms could be grouped under sub-headings or annotated with 'spontaneous', 'unannealed induced', 'annealed induced' so that commonalities in the different groups would be obvious. More explanatory and clearer figure captions would also help.

We can include this information.

Another example is Figure 13 B and D, which show the etching start time for the modelled confined tracks, but the discussion arising from this diagram is about the track etching time thereby requiring the reader to make a mental inversion to make sense of the diagram. Why not simply plot the inverted track etching time (20 sec, minus the start time).

We can consider this, but there is a reason for them being plotted this way which we could also make more clear: these diagrams are extensions of the set begun in Fig 4A,B and Fig 7A, in which each represents a subset of the previous one showing which intersected tracks remain valid and detectable for measurement.

In summary, I believe that this paper is an important contribution and provides a number of significant insights into track etching behaviour that have important implication for the practice of fission track thermochronology. I recommend acceptance for publication after significant revision.

Specific Comments:

Most of these are wording changes that we can address during revisions. We respond to the longer comments.

Line 105: 'Only simple modes are justifiable at …'

Line 125: '…shows the development of total confined track length…' (is this the intention?)

Line 139: '…of lengths, dips, and …'

Line 201: Is this 'power-law increase' arbitrary, based on the model, or derived from the actual measurements?

It is the result of a rough fitting process, which we can describe further.

Line 206: '…corresponding case for standard track selection'. Also, what does 'standard track selection' mean here?

Track selection as it would be done when attempting to measure "fully etched tracks"

Line 207: I see how a selection criterion of VT/VB<=12 can be applied to the modelling results, but how can such a criterion be applied in practice in anisotropic crystals, where the form of the track tips are determined largely by fast and slow etching facets (Jonckheere et al 2019), rather than the idealised tip shapes in Fig 5.

This consideration is why we explicitly and repeatedly leave the door open for a better summary parameter to be determined. However, that being said, although the tip forms might be "largely" determined by fast and slow etching directions, they will also be determined by vT(x) in a way not characterized by Jonckheere et al. (2019), who only depict the tips of fully-etched tracks. Essentially, vT(x) will determine when those facets at the track tips can start to develop.

Line 210: '…also predicts the intersection depth and etching…'

Line 211: '…distribution of the modelled track lengths…', also 'It is clear that…'

Line 218: '…first a semi-track length distribution…'

Line 219: '…then a distribution of confined track lengths…'

Line 221: '…function is the reduced chi-squared…' (or 'a' – still a clumsy sentence)

Line 241-2: '…were exerting a disproportionate control…'

Line 242: '…for experiment SE2 feature a very similar mean…'

Line 243: The expression of this whole line is clumsy and difficult to understand.

Line 278-285: I am really not sure what this paragraph is trying to say.

We can work on it. This is another case of first-ever-of-its-type data, and we can't fully explain it, so we want to leave a marker.

Line 299: '…impingement happened first for each track, where more than one is present.'

Line 344: 'both revealed using Cf semi-tracks' – this is not indicated in Table 1.

Line 454-5: I have no problem in considering a change in etching protocol and I think achieving standardisation in this area should be a community-wide objective, but the aim of 'allowing tracks at all levels of annealing to etch more completely' is probably impossible to achieve in practice and the concept of 'fully etched tracks' likely to remain a mirage. The disparity in compositionally controlled etching rates is so great between grains in many samples that it is essentially impossible to achieve an identical degree of etching in each. I think standardised procedures and calibration of the degree of etching is a more achievable goal.

We agree and disagree. Etching "more completely" is not an impossibility – "more" is quite relative and achievable. The idea, at its simplest, is to have a higher percentage of tracks be more fully etched, which our model indicates is achievable as expressed in some of our discussion about Figure 13.

Answers to comments by Edward Sobel (RC5)

General comments I have read through both the manuscript and the 3 presently existing comments and replies (no reply to RC4 at this moment). I will try not to reiterate what the other reviewers have noted. I think that they have done a good job in addressing weaknesses in the ms, although not always in a manner which makes it easy to see how to revise the text. In general, the study has the potential to be an important work examining how confined track length and termination shape evolves as a function of etch time. It is clearly not the final answer to this topic, nor do the authors claim that it is. However, it is a complicated problem which has been studied for over 35 years, so why expect a quick solution :-) Yes, progressive steps forward are useful. In this case, technology has made it possible to track track growth in ways that previous generations could not measure - this has lead to the ability to improve our understanding. We all (try to) build upon previous work - that is the goal! I certainly learned a great deal about how and where confined tracks form and lengthen - and how few of them are actually measurable. The discussion of how track tips become more defined as they become better etched and how defining these shapes may lead to better inter-operator agreement is an important point in the study.

We are grateful for the endorsement

Specific comments In general, I concur with the other reviewers that the text is challenging to read and needs significant rewriting to make the points easier to understand. It would help if the introduction provided a better roadmap of what will be presented in the ms. Renaming the sample numbers with abbreviations matching the experimental conditions might help the reader to keep track of what is being done. E.g., SE1 could be renamed IU1 (Induced Unannealed).

A good suggestion that we will implement.

Transition sentences explaining what is about to be presented - such as 'observed data will be compared to numerical results' – would help the reader to follow the text. Some figures would benefit from captions that stated whether the plots depict observed or modeled data. In general, the captions are too short and do not sufficiently explain what is being shown (and why). Better labelling of figures would also help. Rather than just writing text in the caption, place information on the individual panels as well. For instance - add "randomly oriented unannealed induced tracks" and 'Cf tracks' to the individual panels in fig. 3. I agree with Gleadow's comment that a figure showing all geometries and terms is required. A table defining all terms would also help. The study of annealing of fission tracks has been advancing in fits and starts for over 55 years. A fair number of references from this evolution are included. Most fission trackers have probably not read most of these papers nor are they deeply attuned to the unresolved problems. Yes, they should be aware; however, a more robust introduction to the open problems would be useful for many readers. The text is written in a very compact form, with the authors apparently assuming that the readers are quite familiar with the topic. I encourage them to expand the text. This is a paper which can at present only be read by a specialist. Perhaps this is true of most scientific papers, and this is not a bad thing. However, this ms requires a large amount of prior knowledge because many points are taken for granted rather than being explained. The introduction should clearly explain what tracks are and how they etch (damage 1st, then bulk etching in 3D). A sentence about how tracks form wouldn't hurt either. At the moment, I think that many non-specialist readers wouldn't even finish reading the introduction. Note that by non-specialist, I mean a fission-tracker who isn't deeply into methodology. Repeating methodologic information from the experimental paper (Tamer and Ketcham, 2020b) would be useful - how exactly was the step-etching done? This is pretty fundamental for understanding this dataset. And as many institutions do not subscribe to Elsevier journals, it is not always a 1 minute task to get this reference.

The discussion is much easier to follow than the preceding sections.

All of these points are well taken. Among our concerns when writing this paper was that it was getting to be too long, imparting a different kind of length bias against people reading it, which is why we economized on some of our introduction and description. We also note that we received critiques both ways; the first comment by Jonckheere and Wauschkuhn took us to task for spending too much time explaining important angular biasing effects when we could have just given an *equation and referenced Dakowski (1978), an important but under-read paper. In our revision we will err on the side of explaining more.*

59-60 More recent work has documented enhanced but continuously diminishing etching velocity in the region along tracks beyond where most current etching protocols reach (Jonckheere et al., 2017). Could it be that as etchant travels along a longer/deeper track, that the strength of the acid and hence the etch rate decreases? This is suggested by the difference in measured track length of 20 sec etching of SE3 (3 steps, 14.89 microns) versus TK20 (1 step, 14.43 microns) and SE2 (2 steps, 20 sec, 16.19 microns) versus SE1 (1 step, 20 sec, 15.77 microns).

This is an interesting idea, on which we are aware of no relevant data. We don't think this is a significant effect, however, because (a) we can and do explain all of these data without such a mechanism, as an outcome of track selection; (b) we see different etching rates within tracks of similar length but different origin (SE3 unannealed spontaneous versus SE6 lightly annealed induced); (c) our EAE experiments measure the same bulk etch rate for a 5-second versus 10-second etch.

A parallel question - does the strength of all possible etchants (e.g., 1.6 vs 5.5 mol nitric acid) have the same relationship to both bulk etching rate and track revelation rate? Or is there a difference between these rates as the acid is changed? This is relevant for both slow versus fast etch acid recipes as well as changes in acid strength as it penetrates farther into the crystal. These questions likely reveal my ignorance about the etching literature; however, answers or guesses about these questions might be helpful for improving the ms, particularly as the average reader is probably as ignorant as I am about this topic. The 2nd point is partly addressed in 413-423.

We believe that etchant strength does affect etching anisotropy (Dpar/Dper), and thus expect that it would also affect the relationship between bulk etch rate and along-track rate. However, we are not aware of a published study that has this information for apatite. We can add a cautionary note, but prefer not to speculate too much.

87-88 There was no clear indication of vB varying with track orientation (Tamer and Ketcham, 2020b). This was commented on by other reviewers. Please add a comment in the text about this observation, similar to the response to reviewer.

We will do this.

92 In cases where it was difficult to determine if an intersection truly occurred due to interfering features, we conservatively included it. Following Green's comment and your reply, it would be useful to explain your explanation of conservative here.

We will do this.

129-130 Variation in impingement point alone is likely responsible for some component of the observed variation in track lengths. How many tracks would one have to measure for such variability to become irrelevant. I.e., is this concern real and important for actual data collection or is this just an issue for interpreting this data set?

We don't really think it's an issue so much as an inevitable part of the data. No number of measurements would make it go away, unless, say, one screened tracks based on impingement point (which we don't think would be worth it).

180-185 Adding photos of actual track tips would help to explain this useful concept. Naturally, anisotropy is quite important here, but this approach provides a way to go forward for better defining criteria about when a track can be measured. The argument that operator's track identification choice is responsible for much of the length variability observed in the inter-lab comparison is a key result of this study.

We will compile such a figure.

207-208 The tracks have a range of tip development (Fig. 6D), and only selecting those with $vT/vB < 12$ (Fig. 6E) results in an excellent match to the measured data (Fig 6F). Excellent seems overstated. Adding text to the caption to point out what one is supposed to notice (this agreement) would help.

One does need to mentally rescale the dark blue histogram bars, but we think the match is pretty darn good…

480 efficiency suggests that a carefully controlled preheating step could greatly increase confined track numbers, potentially without affecting lengths and thus paleothermal information Potentially is the key word in this sentence. If it did affect lengths, wouldn't this be fatal? It seems unlikely that the length reduction could be sufficiently well-constrained, particularly for detrital apatites, which are more likely to exhibit kinetic variability.

The idea is that a heating treatment could possibly be found that results in no length reduction but a different etching rate. We don't think investigating this idea would be too difficult, as it's not hard to conduct experiments where one aliquot undergoes slight heating and the other does not. For example in the experiment from Wauschkuhn et al. (2015) that's already been discussed in this review process, heating at 1h for 175°C results in no reduction in mean spontaneous track length in Durango apatite. Might it be enough to accelerate etching? Potentially…

fig 2 caption - please add definition of xint in caption to help reader understand the figure. It is not defined in the text. It appears to mean the position where the track began to etch / was intersected by a semi-track. The reader shouldn't have to interpret how a parameter is defined - it must be clearly stated. Ah - it is defined on line 159 – far too late. It is quite hard to see the difference between the 2 sets of figures; therefore, the difference should be described in the figure caption. If there isn't a real difference, then say so in the caption so that the reader doesn't get frustrated.

OK; actually, we can incorporate this definition into the figure requested by Andy Gleadow schematically illustrating the entire model.

Fig. 8 is quite hard to understand. Write a useful caption! Fig. 9 shows 15 plots. The caption says nothing. Do you really expect the reader to look at this and understand your point? Ok, you have plotted 5 different data sets, each with the same 3 cross-plots. Conclusions? What is significant? Yes, if I have lots of time, I can try to figure out your point, and remind myself what are the differences between these 5 sets. However, if you want readers to understand your points, it would be better if you guide them. Otherwise, many people will give up and get nothing from at least this part of the paper. And ultimately, articles which demand too much patience by the reader are not yet ready to be published. Please increase your font sizes. Look at the text below the color scale on this figure. Do you really think that it can be read (including the subscript) without strong magnification?

Admittedly, with those we are counting on the fact that this is being published in electronic format, where zooming is trivial; the possibly too-ambitious goal was to fit everything on one page. We'd be interest to hear form the journal editor on their recommendations in this respect.

Technical corrections 105 Only simple models justifiable at this point because our data consist of a very limited number of experiments. Add 'are' 150 Figure 3 shows examples of penetration and revelation rate. add 'calculated' 178 which will develop as a function etching velocity add 'of the' 199-200 Etching time must be greater than 3 s, after which track observability the probability of selecting a track is represented as ((L - 4.5)/5.5)3; A word or 2 are missing here 220 The merit function is reduced chi-squared () This is too brief -for instance: We used a reduced chi-squared () value for the merit function. 230 Thus, we simultaneously fit data sets SE1 and SE2, At a minimum, point to table 1, so the reader can be reminded what SE1 and SE2 are. 233-234 Similarly, we co-fit SE3 with a single-step 20s measurement of unannealed fossil tracks from Tamer and Ketcham (2020a). Please name this dataset here (TK20). 236 After several trials, we settled on a vT/vB of 12, why? 242-244 adding commas to this sentence might help to make it understandable on the 1st reading. 270 add micron after ~0-0.2 Table 1 notes need to define more column headers. Fig. 11 - Explain what velocity gradient means in the caption. What is the database here? Fig. 13 - What do lighter and darker bars on histogram mean? Presumably the same as a similar fig. 6. State that - don't just expect the reader to remember (or to have read the entire ms).

We can implement these corrections.

Answers to further comments by Paul Green (RC6)

The comments by different Reviewers of this ms show a high degree of consistency in highlighting a large number of issues, particularly emphasising the poor standard of presentation associated with this manuscript in terms of the lack of explanation in both the text and Figure captions. It should not be necessary for a reviewer to read a number of other papers in order to review a new paper, but in this case it is not possible to follow the work from the ms alone. I spent way too much time in reviewing this ms and I'm sure the other reviewers did too. I particularly congratulate Ed Sobel on his thorough and thoughtful comments.

We also note that Dr. Sobel was supportive of the importance of this work.

My overall response to the ms is to ask what practical consequence this work could have. I understand that any progress in understanding the fundamentals of track revelation is useful in principle, but at the end of the day, in practical application we have a grain mount containing various apatite species with different etch rates, all etched together, for one etch time and with one etchant. If analytical data in unknowns and age standards and calibration data for kinetic models are generated using the same conditions, then reliable results should be obtained.

This depends on what one means by "reliable" and "should." Recent studies have documented quite thoroughly that using the same conditions does not necessarily lead to the same results, and our work provides a testable answer as to why, which is a very useful first step in solving the problem – a very practical consequence. Another practical consequence is providing a potential quantitative basis for optimizing etching, rather than assuming that we learned everything we need to know 40 years ago. A final practical consequence is to improve our ability to construct a physically based, rather than empirical, understanding of annealing. As Andy Gleadow cogently pointed out in his review, confined track revelation "has previously been understood only in the vaguest terms," and a vague understanding of the principal observational data is not conducive to developing a rigorous understanding of what those data say.

This raises a huge number of practical questions, but they are not directly relevant to the present discussion. One issue that still puzzles me is the repeated assertion, in the paper under review and in earlier papers in the chain, that VB is not anisotropic. Surely the etch figures in a prismatic surface show that the etch rate is higher along the c-axis than perpendicular to it. Can anyone explain this conundrum?

In their response (RC7), Jonckheere and Wauschkuhn provide an explanation based on etching theory. For our part, we could object to what appears to be a repeated willful misrepresentation – we never claim that VB is not anisotropic, we just state that our experiment showed no clear indication of anisotropy.

I see little point in responding to most of the authors' comments on my review. The consistent opinion of all reviewers is that this is an extremely poorly presented manuscript, and requires significant revision before it could be considered acceptable for publication. Perhaps if the authors take all the comments into account the final product may eventually be easier to understand and appreciate. However I would like to respond to one comment: In their response to my Review, Ketcham and Tamer suggest that I was misdirected in my criticism of the Wauschkuhn et al. (2015) study. I was not. In my review I discussed the basic motivation of the Wauschkuhn et al. study, in terms of their ideas concerning the KTB borehole (as expressed in the title of their paper). I was not discussing the particular aspect that Ketcham and Tamer refer to in their comments, as illustrated in Figure 15 of Wauschkuhn et al., casting doubt on the principle of equivalent time. As I was reviewing the ms by Ketcham and Tamer I did not see the need to provide additional comments on any other aspect of the Wauschkuhn et al. study, apart from their basic premise. In regard to the evidence in Figure 15 of Wauschkuhn et al. (2015), the data presented there appear to fully support the validity of equivalent time. My reading of that Figure is that the induced tracks that were pre-annealed do not begin to start shortening again until heated at a temperature above that used in the initial treatment. At higher temperatures, both induced and pre-annealed induced populations give similar track lengths, which is just what is predicted from equivalent time. Regarding the comparison with spontaneous tracks, Wauschkuhn et al. (2015) acknowledge that they cannot be compared directly with induced tracks because "the fossil track population is not a single-length population, and contains somewhat shorter and somewhat longer tracks than a population of induced tracks pre-annealed to the same mean length". Duddy et al. (1988) provided clear experimental evidence confirming the validity of equivalent time, based on measurements of mean track length in apatites that had undergone various variable temperature annealing treatments. In common with Wauschkuhn et al. (2015), Ketcham and Tamer in the paper under review fail to mention this and proceed to cast doubt on the concept. In their response to my review, Ketcham and Tamer state "The experiment in Wauschkuhn was designed very specifically to test the equivalent time hypothesis in a way Green (1988) did not." This comment is either deliberately misleading or betrays a basic lack of understanding. The Green (1988) study was not designed to test equivalent time. The Duddy et al. (1988) study was designed to test equivalent time and the concept passed with flying colours. It is unacceptable to cast doubt on the validity of equivalent time without citing published evidence that clearly validates the concept.

We find this critique unwarranted. Not all tests are by design (e.g., "she was tested in battle"). As we said in our earlier response, and as Jonckheere and Wauschkuhn say in their follow-up comment, Duddy et al. (1988) only tested the equivalent time hypothesis on induced tracks, not spontaneous ones. We did not imply that the experiments in Green (1988) were designed to test equivalent time, but they were the first experimental cross-validation of confined length annealing between spontaneous and induced tracks, so they were a test all the same. Moreover, it is a test the hypothesis appeared to pass, as the outcome led to the recommendation by Green (1988) to *normalize spontaneous lengths with unannealed induced lengths, supporting the idea that the two types of tracks can be considered interchangeably. Wauschkuhn et al. (2015) provided a more direct test, and as discussed in the J&W follow-up reply (RC7), the equivalent time concept as originally stated did not pass it. We are aware of one other unpublished study that repeats the Wauschkuhn et al. (2015) result, and our present study suggests a new distinction between fresh induced tracks annealed in the laboratory and those that have evolved over millions of years. We concur with J&W that further follow-up is warranted.*

The fact that they discussed Duddy et al. (1988) in a previous paper, as noted by Ketcham and Tamer in their Response, is irrelevant to the present discussion, and only highlights the lack of full documentation in the ms under review.

Comment on remarks by Jonckheere and Wauschkuhn (RC7)

These remarks concern Green's review, not our paper, and so we do not see a need to respond, other than to agree with them, and to thank them for saving us the effort with their thorough response.