Review of Manuscript gchron 2020-31. "Confined fission track revelation in apatite: how it works and why it matters" by Ketcham and Tamer, submitted to Gchron.

Review Accepted 21/10/2020
Review submitted 10/11/20

Paul F. Green,
Director Technical
Geotrack International
10th November 2020

The paper under review (henceforth "the ms") presents results of a series of calculations and models regarding the revelation of fission tracks in apatite, with detailed discussion of the implications. The subject matter falls within the remit of the Journal and would be of interest to a sub-set of readers.

However, I have to say that the ms is written in a style that I found very difficult to follow, and required several readings before I began to understand the overall aims of the paper and the reasoning behind the work described therein. One of the main reasons for this is that the study relies on data and discussion presented in a companion paper (Tamer and Ketcham, 2020) and other works. In addition, or perhaps as a result, many aspects of the work described in the ms are unexplained, or difficult to follow, simply being referenced to previous papers. In fact, I only really began to get a clear idea of the rationale behind the work, together with the experimental design and significance of the analytical results after I had read a series of previous papers going back to Wauschkuhn et al (2015), which appears to form the foundation for these later studies. Needless to say, I regard this as an unnecessary level of effort required to provide a review of a new contribution. The text is also replete with vague expressions which have unclear implications, as well as inaccurate statements or misunderstandings and non-intuitive results that are accepted without question. Some examples are cited below, but this is not an exhaustive list and the authors need to make an effort to explain their procedures and reasoning in a lot more detail.

Perhaps the biggest problem with the ms is that it is founded on a complete misconception of the application of apatite fission track thermochronology, emanating from the paper by Wauschkuhn et al (2015). To quote from the manuscript under review (lines 26 – 28):

> *Measurements of laboratory-annealed spontaneous and induced fission tracks do not agree (Wauschkuhn et al., 2015b), leading to continuing uncertainty on the fidelity of induced tracks annealed in the laboratory as proxies for spontaneous ones annealed at geological conditions over geological time scales.*

The paper by Wauschkuhn et al (2015) casts doubt on the use of kinetic models based on laboratory annealing studies of induced fission tracks in apatite to predict the behaviour of spontaneous tracks in geological conditions. I note that if this doubt were to be well founded, then the very basis of the apatite fission track method, as implemented for example in the HeFTy software of the first author, would be invalid. The Wauschkuhn et al (2015) study is based on the assumption that samples from the KTB borehole in Germany have undergone essentially isothermal annealing since the rock section cooled to temperatures at which tracks were retained. This seems highly unlikely, given both the geological evolution of the region which contains a series of tectonic events during the Cenozoic (discussed in detail by Wauschkuhn et al., 2015) and also a number of previous studies which have revealed a more complex thermal history framework for samples from the borehole. Wauschkuhn et al (2015) claim that the various aspects of the thermal history invoked to explain the apatite data in samples from shallow depths in the borehole cannot be believed, due to an absence of independent evidence. But they then proceed to invoke a novel physical process for which there is also no independent evidence, to explain these data. They thus simply replace one enigmatic process with another. A borehole such as the KTB, where a range of factors may contribute to the thermal history of individual rock samples, cannot and should not be used to provide a geological test of annealing behaviour.

Thankfully, there is ample published evidence showing that the view expressed in the paragraph quoted above is without foundation. In my 1988 paper (cited in the ms) I described a series of experiments which showed that spontaneous tracks anneal in similar fashion to induced tracks once the induced tracks were shortened to a similar degree to the spontaneous tracks (this is not mentioned in the ms). This is an expression of the concept of equivalent time, originally introduced by Duddy et al. (1988) (not cited in the ms). A vitally important fact, that neither the paper under review nor any of the preceding papers have mentioned, is that Duddy et al. (1988) performed a series of careful laboratory experiments to validate this concept. In similar vein, a number of studies have shown that the predictions of fission track annealing models are consistent with measured data in a range of different geological conditions, e.g. Spiegel et al. (2007); Green and Duddy (2018), while other studies (e.g. Green and Duddy, 2010; Japsen et al., 2020; Green et al., 2013, 2018) have reported results from apatite fission track methods which are highly consistent with other paleo-thermal (e.g. vitrinite reflectance) and paleo-burial (e.g. sonic velocity) indicators. None of these studies confirming the equivalence of spontaneous and induced tracks and the validity of apatite fission track methods, have been cited in the ms or its preceding papers. Thus it is my view that the ms is based on a false premise and provides a biased view of the state of the technique. If spontaneous tracks did not behave in similar fashion to induced tracks then surely it would not be possible to use the technique to accurately predict track response in geological conditions. In my view, the failure of this and other papers to cite evidence which oppose their own views represents a serious failure to follow established practices in scientific publication.

The ms suffers from another serious shortcoming is that although not stated explicitly, it suggests that etching of confined tracks is the only consideration in determining appropriate etch times. In practice the etch time is also required to ensure that tracks which intersect the grain surface (i.e. those employed in determining a fission track age) are clearly revealed, to allow accurate counting without too many overlaps which might obscure some tracks. In addition, in routine application involving natural samples, particularly in the case of detrital grains in sandstones, grain mounts invariably contain apatites with a range of bulk etching rates, so that tracks in some grains will be etched to a greater degree than in others. Thus, the typical etch time of 20 seconds represents a balance between the need to sufficiently reveal tracks in grains with the lowest etch rates while not obscuring tracks in grains with higher etch rates. Accepting this view, and given the importance of measuring track lengths and ages in the same grains (although sadly this is often not done), it is difficult to see how the results of this study could find any practical application. Perhaps the authors should give some consideration as to how their work could be implemented in practical application.

Another issue, which the authors appear to discount, is that if measurements (including annealing experiments and data in field samples) are standardised to the same etching and measurement conditions, then the effects that they are discussing in this ms should not be a problem. This has certainly been our philosophy at Geotrack since 1982. In our early days, we performed several standardisation exercises with personnel from other laboratories to ensure that similar measurement protocols were adopted. However in recent years the explosion of laboratories around the word has resulted in a considerable number that have not shared that consistent approach.

Another motivation for the study described in the ms concerns the disappointing degree of reproducibility of track length measurements in inter-laboratory comparisons (lines 59-60). The authors may be better advised to focus on this aspect, rather than any perceived problems with the equivalence of spontaneous and induced tracks. A lack of standardisation between laboratories is likely to be a major feature in causing the observed lack of reproducibility.

Another important issue with one aspect of the analytical procedures adopted in the study concerns the measurement of under-etched tracks. The previous paper (Tamer and Ketcham, 2020b) discusses “a limited number of tracks that remained under-etched at 20 s but were measured anyway to adhere to the experiment design”. This allows a certain degree of arbitrariness into the measurements which makes it difficult to know how to interpret differences from one etch step to the next. This issue is also referred to in the following list in relation to lines 74-77 and 194-196. The question arises of how the ends of under-etched tracks were measured, and what such eaurements mean. The resolution of the track end points becomes a critical issue and where the width of a track is less than 1 µm we are talking about measuring objects with a size close to the wavelength of light. So do these measurements indicate the actual etched length or only some part of that governed by resolution? Thus there is a question of how much the scatter in the measurements is due to problems in resolving track tips and how much is due to real differences in etch rates. This is also an issue with measurement of Dper, all of which are around the wavelength of light or less. This would account for the degree of scatter in the Dper measurements in Figure 5c of Tamer and Ketcham (2020b). For these reasons, some discussion of how the authors dealt with this problem is required.

Specific comments regarding the text:

Line 60: *the region along tracks beyond where most current etching protocols reach*

The meaning of this is unclear.

Lines 44-48: *Virtually all mathematical treatments of track revelation, biasing, and the relationship between confined track length and track density (… … ……… …..) presume that tracks are line segments in space, all etched to their full extents once they are intersected.*

This is inaccurate, at least in terms of those papers involving Galbraith, Laslett and our group. These mathematical treatments are based on the assumption that only tracks that are etched to their full extent are measured. This is not the same as the statement quoted from the ms.

Lines 73-74: *Finally, to be measured a confined track must be etched sufficiently to be observed, the criteria for which will differ depending on the situation.*

What does the second part of that statement mean?

Lines 74-77: *For routine AFT analysis, where the analyst evaluates whether a track is sufficiently etched, the ends of the tracks need to be clearly visible, although this evaluation is analyst-specific. For measuring tracks in early steps of step-etching experiments, the criterion is simply that a track and its tips be visible enough to make a reasonable measurement.*

There is an underlying question here that pervades the ms as well as preceding papers; what criteria should be used in measuring tracks where the ends are not properly visible? This needs to be discussed – at least, the authors should discuss the approach that they used here.

Lines 87-88: *There was no clear indication of v_B varying with track orientation (Tamer and Ketcham, 2020b)*

I find this puzzling. Surely the shape of the track openings in a prismatic surface shows that the bulk etch rate is higher in the direction of the c-axis than perpendicular to it. In this context I am puzzled as to why the authors have adopted the value of 0.022 µm/s for the bulk etch rate, when other samples from the same experiment define different growth rates, and this value is intermediate between the growth rate of Dpar and Dper.

Lines 92-93: In cases where it was difficult to determine if an intersection truly occurred due to interfering features, we conservatively included it.

I would say that a conservative approach (meaning careful, in my view) would be to exclude such values.

Line 94: *Section 4 heading. "4. The model"*

What model? Up to this point there has been no description of what is to be modelled or why. This needs to be explained and the basis of the model should be described. The reader should not have to work this out for him/herself.

Line 121: *Figure 2 illustrates the implied growth curves*

What is meant by "growth curves"? An informed reader can probably work this out but the content of the Figure should be explained more clearly. And in C,D, why are curves not shown for t=20 sec?

What values of VT, VB have been used in constructing this figure?

And in regard to Lines 129-130, It seems unlikely that differing intersection times would contribute significantly to differences between observers if other requirements of measurement (in terms of tip shape) are met, on the basis of this Figure.

Lines 134-135: *First, the relative probability of a track of latent length or half-length L crossing the surface*

Is it latent length or half length? The two terms suggest different meanings.

Line 138: *The semi-track penetration calculation…… ..*
Line 150: *Figure 3 shows examples of penetration and revelation rate.*

Both of these extracts from the text represent one of the most confusing aspects of the paper, at least as far as I am concerned. I cannot work out what Figure 3 is supposed to show, and the text provides no information that will help to work it out. Simply "penetration and revelation rate". These terms should be explained in detail.

Line 165: *Figure 4 shows an example model of 10_7 track intersections*

This Figure doesn't show a model. It shows results from a model. Maybe I'm old fashioned but I think text should be clearly written, describing what is shown accurately and clearly.

Lines 194-196: *For step etching experiments with first steps shorter than 20 s, precise location of the tip is not a prerequisite for track selection, which is instead a matter of simple visibility. In the earliest stages of etching, tracks will be too thin to be observable in visible light. As they grow, they become more efficient at reflecting light, making them more detectable. However, when this occurs in the context of practical fission track analysis is unclear.*

This is not clear to me. To measure a track it is necessary to identify the end, so how did the authors decide where the ends were?

Line 197: *Lacking a physical basis for determining when etched tracks begin to become visible*

This follows on from above. I find this section very confusing. Constructing an empirical bias relationship to allow for this aspect seems a very imprecise solution. Why should the same bias function apply in all cases?

Lines 210-211: *Figure 7 shows the distribution of track lengths and etching times with depth below the surface*

This Figure does not show the distribution of lengths and etch times. It shows model predictions in terms of contours. Once again, precise description is important. Plus more description

Lines 218-219: *Fitting the experimental data consists of posing an etching structure and using it to construct first a semi-track distribution and then a distribution of etched lengths*

What is being fitted to what here? This should be explained clearly.

Lines 221-222: *We also include the option to incorporate some innate variation in latent track length. For the modeling in this paper we chose σ of 0.5 µm, based on the scatter of lengths after **c**-axis projection to remove anisotropy effects*

Wouldn't a value of 0.8 or so be more appropriate, based on the distribution of induced track lengths which show no anisotropy?

It would also be interesting to run models with a single latent length, to investigate how much of the observed distribution of measured track lengths can be attributed simply to etching processes.

Lines 229-232: *The step etch experiments only consist of 3-5 steps, making it difficult to meaningfully constrain models with 2-3 variables defining etching structure, in addition to biasing factors. To increase the amount of data, we paired data sets based on the same or equivalent tracks. Thus, we simultaneously fit data sets SE1 and SE2, both of which analyzed unannealed induced tracks, but with different initial etching steps.*

This is another of the most ambiguous passages in the ms. In what sense were the data sets paired? What does that mean? Why try to fit models with 2-3 variables to data with only 3 to 5 steps? Why not try simpler models or use more etch steps to resolve these problems? And why combine data with different initial etch times? More control on experimental design would ameliorate these issues.

Lines 241-245: *During fitting of the unannealed induced track data (SE1 and SE2), it became apparent that one pair of data points were exerting outsized control on the result. Since the 15-20 s and 20-25 s steps for experiment SE2 feature the very similar mean length reduction and thus almost the same rate, in the context of our model forms assuming a linear rate decrease fits were forced to split this difference to the exclusion of closely fitting the rest of the data for SE1 and SE2. We thus excluded the 15-s mean length for SE2, effectively making the second etching step go from 10 s to 20 s*

I wonder if this presages graver issues within the data set. Why should this dataset be problematical? To simply reject these data seems rather arbitrary.

Line 268: *Figure 9 shows the parameter fits for Constant-core models, and Figure 10 for Linear models*

Here is the biggest single problem that I had with reviewing the ms. What on earth does this mean? This Figure shows a series of cross plots, without any description of what is being plotted or why, and what they mean. What does each dot represent in each plot? I find it totally unacceptable that the authors leave it to the reader to work this out.

Line 287: *It is remarkable that we are able to reproduce the mean length data in each etching step of all experiments simultaneously using these simple models of etching structure*

Give the number of adjustable parameters, each of which is free to vary as required until a fit is obtained, I don't see this as remarkable at all. And regarding the statement that mean length is reproduced, I guess that Figure 8 is supposed to show this, but a plot of measured vs fitted parameters would seem to be a more graphic depiction of this.

Line 329: *Fossil and unannealed induced tracks have slow core etching rates, while all annealed induced tracks have far higher rates.*

It is counter-intuitive that annealing would result in an increase in etch rates within the residual track region, particularly when spontaneous tracks do not show enhanced track rates. Such behaviour would suggest that heating increases the degree of damage, when all previous experience with lattice damage suggests that heating leads to a decrease in damage intensity. This suggests that there is a flaw in either the experimental design or the model design from which this conclusion was derived. The authors seem to accept this observation at face value, but this unusual result deserves further investigation before any firm conclusions can be drawn.

The fitted values of V_{Tmax} and Δx_{Tmax} in Table 2 show no consistent trends that might suggest these results have any meaning. It would be reasonable to expect these values to vary in a consistent manner as the degree of annealing progresses, but for annealing temperatures of 235, 270 and 280°C, V_{Tmax} varies from 4.0 to 1.77 and back to 3.36, while Δx_{Tmax} changes from 0.19 to 11.35 and back to 0.76. It is hard to imagine why the distribution of damage should go through such contortions within a limited range of annealing temperatures. These results suggest that the fitted values are simply empirical numbers without any real meaning. For the Linear model, values of V_{Tmax} change from 4.09 to 3.54 to 3.59 over the same range, compared to the value for induced tracks of 1.70, and again it is difficult to see how these numbers can be accepted as having any physical reality. This I turn raises questions regarding the significance of other aspects of the results.

Incidentally, it would be useful to show error bars in Figure 11, as per normal practice.

Lines 331-332: *This may be responsible for some component of the mismatch in annealing fossil versus induced tracks in laboratory experiments (Wauschkuhn et al., 2015a),*

As mentioned earlier, there is plenty of published evidence that there is no mis-match. The authors appear to have adopted this observation with no critical assessment. At the very least, the authors should mention that evidence to the contrary exists, and preferably also to explain why they discount that evidence.

Line 337: *treatments of fission-track lengths are all based on a line segment model*

Any mention of the line segment model should refer to Rex Galbraith's (2005) book.

Lines 457-457: *The $v_T(x)$ model framework provides a quantitative basis for evaluating whether the etching procedures used today are the most effective at their goal, which is to provide high numbers of reproducible and informative measurements to constrain thermal histories*.

I would dispute that the goal is simply to provide high numbers of measurements. Surely the goal should be to obtain data that provides the most accurate thermal history constraints possible. This depends not only on track length measurements but also on fission track age determinations and also on allowing for different etching rates in apatite grains on a grain mount. The *$v_T(x)$ model* described in the ms focusses on only one of these aspects.

References not cited in the ms under review:

Duddy, I.R., Green, P.F., Laslett, G.M., 1988. Thermal annealing of fission tracks in apatite 3. Variable temperature behaviour. Chem. Geol. 73, 25–38.

Green, P.F. and Duddy, I.R. 2018. Apatite (U-Th-Sm)/He thermochronology on the wrong side of the tracks Chemical Geology 488, 21–33

Green, P.F., J Duddy, I.R. & Japsen, P., 2018. Multiple episodes of regional exhumation and inversion identified in the UK Southern North Sea based on integration of palaeothermal and palaeoburial indicators. In: Bowman, M. & Levell, B. (eds) 2018. Petroleum Geology of NW Europe: 50 Years of Learning – Proceedings of the 8th Petroleum Geology Conference, 47–65.

Spiegel , C, Kohn, B.P., Raza, A., Raiuner, T., Gleadow, A.J.W. 2007. The effect of low temperature exposure on apatite fission track stability: a natural annealing experiment in the deep ocean. Geochimica et Cosmochimica Acta 71:4512- 4537.