Comments	Replies
The manuscript titled "different take on fission-track annealing in apatite" by Jonckheere et al., presents an original and provoking re-interpretation of apatite fission-track annealing. The proposal that compositional effects are primarily expressed through the onset of accelerated shortening rather than through different overall annealing rates is novel and deserves consideration.	We are indebted to the reviewer for her thoughtful and thought-provoking comments, which allow us to broach more general issues, that are perhaps of broader interest than our manuscript itself.
The central assumption of the manuscript is that all apatite compositions share a common gradual annealing trajectory and differ primarily in the onset of accelerated annealing: compositional differences do not primarily control the rate of track shortening; rather, they control the point at which accelerated annealing begins. In this framework, fission tracks in all apatites undergo an initial stage of gradual shortening, followed by a stage of accelerated shortening, with the transition occurring at different mean track lengths depending on apatite composition.	Our main assumption is indeed that different apatites have different thresholds for the transition from gradual to accelerated length reduction (break-up). One can view as it a brittle-failure model to contrast it with the mainstream plastic resilience (r_{mr0}) model. We wish to stress that this assumption reflects a reasonable interpretation of the available data (Figure 1; Figure 2; Figure 6; Supplement Figure S1) just as much as the established mainstream models (Carlson et al., 1999; Barbarand et al., 2003; Ravenhurst et al., 2003).
The core hypothesis of the manuscript is that compositional differences among apatites are more appropriately described by a threshold mean length for the onset of accelerated shortening than by composition-dependent annealing rates as implemented in existing models. However, the procedure used to distinguish data belonging to the gradual and accelerated regimes appears largely empirical. A sensitivity analysis showing how the model parameters vary when different transition thresholds are adopted would considerably strengthen the argument.	The partitioning of the data into gradual- and accelerated-annealing subsets is indeed based on judgement. It is obvious in most cases to which a given mean length must be assigned. In less clear cases, the decision can be based on consideration of the preceding and subsequent annealing stages (Figure 1; Supplement Figure 1). Occasional misassignments of borderline cases, one way and the other, do not invalidate our work. We assume that the proposed sensitivity analysis is meant to establish if there exists an objective partition between the gradual and accelerated stages, and if our subjective partition differs from it to an extent that is significant with respect to the model predictions. That is a difficult proposition. One can think of the following approach (Figure S1): (1) calculate the slope *m* of the regression line to the unannealed data; (2) project all mean lengths parallel to the regression line on the vertical axis (*Dpar* = 0). We assume that the distribution of the projected values will approximate a normal distribution for annealing runs where all the samples remain in the gradual stage, and become more and more left-tailed with accelerated annealing; (3) the lower values can then be trimmed stepwise, until a normal distribution with the higher values remains, with a standard deviation like that of the unannealed samples. It is however not clear that this cut-off is unique. Let's assume that it is nevertheless possible to construct an unbiased algorithm that requires no subjective judgement, and produces the vector of the transition points of all the samples $T(t_1, t_2, \ldots, t_N)$. We can then fit an annealing model to the combined data up to these transition points. That, however, does not answer the question. To establish how sensitive model predictions are to $T(t_1, t_2, \ldots, t_N)$, we must introduce random variation. Resampling $T(t_1, t_2, \ldots, t_N)$ is not enough; we must also consider data on either side of the transition points. So, we either have to explore the complete dataset, or exercise judgement. Each of the countless iterations requires not just redetermining the annealing model parameters, which tell us nothing, but also their effect, e.g., on the closure temperature. We therefore ask that we be spared this enormous and unpromising task. The aim of this manuscript is not to attempt to establish the most accurate annealing model but to show that a different interpretation of the data also produces a

reasonable model.

A critical point is that the high-temperature annealing model fails to reproduce the observed low-temperature behaviour, and a separate logarithmic time-dependent relationship is introduced to reconcile laboratory and geological observations. However, the physical basis for this additional process remains unclear. The introduction of a separate ambient-temperature annealing mechanism appears largely phenomenological.

We understand that introducing a separate process complicates our model. However there is no evidence excluding different (low-T, long-t)- and (high-T, short-t)-kinetics. Most existing annealing models predict longer mean lengths at ambient temperatures than are measured on fossil tracks in geological samples (Wauschkuhn et al., 2015; Figure 12). Differences between fossil and induced tracks have been reported at regular intervals (Price et al., 1973; Gleadow et al., 1983; Durrani and Bull, 1987; Hendriks and Redfield, 2005; McDannell et al., 2019), and are a suspected cause of the late worldwide exhumation, which still manifests itself on occasion in inverse temperature-time models.

This problem was mitigated when annealing models were adjusted to better fit geological benchmarks. This was based on the assumption that one process was responsible fission-track annealing at all time- and temperature scales, albeit one with different kinetics than before. Our work explores the opposite assumption that the high-T model remains the same for lab and geological annealing but a second process accounts for the low-T results.

Ambient-temperature annealing is however more than an assumption; it is an empirical fact. Donelick et al. (1990) and Tamer and Ketcham (2020) both reported shortening of induced tracks at room temperature. There is a semantic issue here: Tamer and Ketcham (2020) incorporated their data in an existing high-T model, which implies that annealing at ambient temperatures is just an extension of conventional annealing, not a separate process, in contrast to our assumption that it is a distinct process, one that also accounts for the excess shortening in geological samples that remained at low temperatures.

Figure 1 illustrates the difference: we selected the data for annealing runs with an internal standard from the Tamer and Ketcham (2020) experiment, and fitted a linear equation in ln(t), ignoring the high-T annealing data and models, but predicting the mean fossil track length in Renfrew apatite to within ~0.3 μm as the result of a separate low-T process. The authors included older low-T data, without internal standard, and fitted the equation for high-T annealing data. The point is that Tamer and Ketcham (2020) made one assumption; we made a different assumption, but they are both assumptions. Neither approach has an adequate physical basis. Since Laslett et al. (1987), it has been accepted that annealing models should not reflect a physical process.

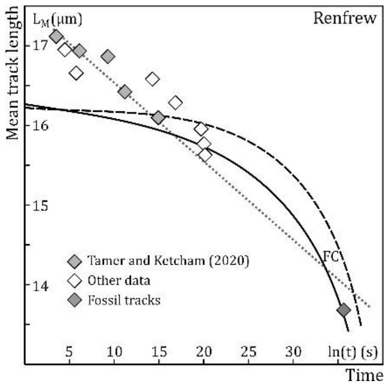

Figure 1. Mean fission track length in Renfrew apatite vs. log annealing time (after Tamer and Ketcham, 2020).

The KTB benchmark represents a key test of the proposed model. But the geological validation remains inconclusive. While the calculated age-depth profile reproduces the observed apatite fission-track ages reasonably well, the predicted mean track lengths show a systematic offset relative to observations. The authors attribute this discrepancy partly to experimental biases and partly to ambient-temperature effects. However,

We consider that the mismatch between the modelled and measured lengths of the KTB samples is a feature worth investigating rather than a failure requiring an excuse. There are several explanations: (1) the measurements are not consistent with the model due to different track selection, etching and measurement protocols (Ketcham et al., 2015; Jonckheere, 2023); (2) if ambient-temperature-annealing is a fact, then the fit with our

mean track lengths constitute one of the principal observables used to calibrate and test annealing models. Therefore, the mismatch should probably receive greater emphasis in the conclusions. At present the manuscript gives the impression of stronger validation than is actually demonstrated.	model prediction is rather good. There is a third, tempting, but less probable explanation: (3) the bias function that weights populations with different lengths is based on a relationship between the densities and mean lengths of annealed tracks (Green, 1988). It is thus more suitable for age corrections than for correcting biases in lengths measurements of confined tracks, which are subject to different selection criteria than surface tracks. It could explain the fact that the age profile is a better fit than the length profile. In our opinion, these alternatives offer a more solid physical basis than *ad hoc* adjustments of high-temperature annealing equations in order to fit certain geological benchmarks.
In conclusion, several key assumptions remain insufficiently demonstrated, particularly regarding the identification of the transition between annealing regimes, the treatment of ambient-temperature annealing, and the geological validation of the model. I recommend a more cautious presentation of the conclusions, clearly distinguishing between observations directly supported by the data and interpretations that remain speculative.	We concede that we have no proof of our assumptions. One can assume that ambient-temperature annealing occurs, or that it doesn't, but they are both assumptions; neither is established in the sense that it has been tried and tested. Neither did we prove that there is a transition between annealing regimes, no-one else has either. It has nonetheless been an uncontested assumption since Green et al. (1986) that tracks do break up at some stage. Ketcham et al. (1999) distinguished high-temperature and low-temperature benchmarks. The former are downhole temperatures at which complete fading of the tracks is observed. As this is based on the fission track ages (Figure 12a) our high-temperature prediction for the KTB fluorapatite samples is good. If we exclude ambient temperature annealing, our model predicts $\lesssim$1 μm shortening in the age standards (Figure 8b; fluorapatite), ~1 μm short of the measured values. In comparison, Figure 1 above shows that, until the low-T data were included in 2020, the established fanning curvilinear model fell ~0.7 μm short of predicting the fossil track shortening in the Renfrew apatite (dashed line). Note, however, that our model has not been adjusted to fit geological benchmarks. Before considering it, we want to exclude measurement factors and ambient-temperature annealing.
I view the proposed model not yet as a replacement for established annealing formulations (the current evidence does not demonstrate that this framework provides a superior explanation compared with established r_{mr0}-based approaches) but rather as an interesting working hypothesis. What is lacking is a quantitative comparison between the predictive performance of the proposed approach and established r_{mr0}-based model. Without such a comparison, it is difficult to assess whether the new parameterization provides a genuine improvement over existing formulations or simply represents an alternative description of the same observations.	Our aim is not to criticize or compete with existing annealing formulations, but to show that a different conceptualization of fission-track annealing also produces reasonable geological predictions, and to invite reflection. The main differences are that our model is the result of a bottom-up process, with no pre-imposed functional form, which resulted in a three-parameter fanning Arrhenius model, with reasonable fanning point co-ordinates ($T = \infty$, $\ln(t) \approx -34$; the values for curvilinear models are $>10^4$ orders of magnitude shorter than the Planck time) and fair geological predictions (KTB: $\ln(t) \approx +34$). In addition, it predicts a partial annealing zone for fluorapatite from ~35 to 119 °C for 10 Ma holding, and a closure temperature of ~105 °C for linear cooling at 5°C/Ma. Excluding ambient temperature annealing, it overestimates the lengths of fossil tracks in geological fluorapatite samples by ~1 μm, i.e., not much worse than the established curvilinear equations. We thank the reviewer for recognizing its potential value as a starting point; we see the mismatch with the mean track lengths of geological samples as an invitation for further investigation, not a sign of failure. We will add the information on the fanning point and the predictions about Tc and the PAZ to the manuscript.
While several aspects of the model remain incompletely validated and require further testing against both experimental and geological datasets, I see clear value in sharing	We acknowledge that our model requires further testing, but point out that one would have to search hard to find a model that has so early been tested to the same extent as

this hypothesis with the fission-track community. The manuscript raises important questions regarding the interpretation of compositional effects on annealing and may stimulate further discussion and investigation of these processes.	ours. The hoped-for future success of our model lies in the fact that it elicits thought about its approach and that of other models and the broader question what to demand of a model. In response to the reviewer's comments, we will formulate our conclusions to highlight the uncertain assumptions on which our model rests in comparison with those of other models.

Sincerely,

R. Jonckheere
F. Trilsch
B. Härtel
T. Nagel