Review of McDannell and Issler, Geochronology

In this contribution, the authors use synthetic data to demonstrate that a multi-kinetic approach to AFT (and AHe) data can be required for achieving a geologically meaningful thermal history. In their example case, neglect of the multi-kinetic nature of AFT annealing cannot be remedied by adding independent geological constraints, or additional data (AHe). The paper is put forward as a companion paper to an upcoming one in which samples with similar characteristics (very old, multi-kinetic) will be analyzed, and thus a certain amount of what's here may be book-keeping that will be useful for the next paper to refer back to. In general, I'm very supportive of this line of inquiry. The writing is very good, though it does get bogged down or venture off-topic occasionally. I think this paper is worth publishing and appropriate for this journal, but first some errors need to be fixed, and the focus needs to be improved.

This is a tricky kind of study to do, because it's difficult to generalize the problem – in which cases does multi-kinetics matter, and in which can it be neglected without overt penalty? For example, if there is fast cooling, subtle changes in kinetics will not matter too much. Put simply, careful attention to kinetics is likely to be most important in cases of long persistence at, or reheating to, a temperature range that differentiates the thermal responses of the grains present, and thus the ages and lengths recorded. It might be best to state this up front, and then pose the subsequent tests as a demonstration of that principle.

The synthetic data set is a bit over the top, in terms of quality. There are three kinetic populations, all equally represented in terms of grains and tracks. The synthetic t-T path has been designed to just touch into the lower part of the PAZ once for each of the two lower-resistance populations, in the necessary sequence for evidence of each to be preserved. More importantly, the authors say they add an "appropriate level of noise," but don't specify what that is, or how they did it. The age uncertainties on each population are all less than 5%, and all three populations have chi-squared probabilities of 100%, suggesting that the single-grain ages are under-dispersed compared to a true natural sample. This is borne out in the numbers in their Appendix; all grains but one are within 0.5-sigma of the central age; the one exception is almost exactly 1-sigma. This is massively under-dispersed compared to what one would expect with a natural sample (i.e., a random sample from a normal or Poissonian distribution), suggesting that something about how they generated these synthetic data was a bit off. This needs to be redone.

The result is a data set that grabs the viewer by the lapels and shouts "multi-population, multi-kinetic." Perhaps this was the point, but it does not make for effective advertising, as real data will never be this clear. It also doesn't make for a realistic test of the ability of thermal history inversion to read the history. They need to run the test with a normal degree of dispersion, and then it might be fun to run one with some excess dispersion, such as by adding some dispersion into the input kinetics, or adding another couple of small populations at different kinetics that are unidentifiable as populations because there are so few grains.

I think they could also have done more with testing along the lines of what's in Figure 5C, where they deleted a population. First, I disagree with the authors on that result – I think the penalty is surprisingly modest, to the extent of bringing up the question of how important that middle population is. If they had, say, a constraint for the depositional age of the initial sediments of the first burial episode, they could probably do without the second population entirely. This may be foreseeable if the lower part of the PAZ for the most-resistant population overlaps with the upper part of the PAZ of the least-resistant one, thus providing effectively continuous coverage. It also gets to a practical matter – if kinetics appear messy in the data, with lots of overlap (which there usually is), can the major information be extracted by concentrating on the end-members present? This would be a useful question to answer, or at least address. Second, another version of this exercise might allow the authors to run a sub-test – say, omit the highest-resistance population, see what they are missing, and then run the two lower-resistance populations as a single one, and see how that (also) messes up, in a 2-population case. This would be a simpler case to reinforce the general point that reheating to the PAZ is when multi-kinetics gets important.

The authors lost me a bit in the discussion of rmr0; this section needs to be shortened and clarified, while keeping the important parts. I have heard informally that people comparing the 2007 and 1999 models in some situations have noted divergent behavior, and have tended to prefer the 1999 one, but I have not seen a quantitative exploration of why this might be the case. The authors claim that the difference may lie in the 2007 version weighting more resistant, and Cl-rich apatites versus apatites that feature cation substitutions. This may be true, but I don't know; it's not clear that augmenting one aspect of compositional space necessarily diminishes the influence of another. Alternative possibilities include other unknown or unexplored incompatibilities between the Carlson and Barbarand data sets, or that the empirical fitting method is oversimplified. In particular, the 2007 model has a relatively compressed total thermal sensitivity range (maximum Tc = 180C in 1999 (note 2006 erratum), 160C in 2007) because of how the Carlson and Barbarand data sets interacted. It's also worth being very clear that the kinetic meaning of rmr0 differs between annealing models – for example, Durango is 0.827 in the 1999 paper and 0.797 in 2007 (rounding down to 0.79 after applying rmr0+$\kappa$=1.04), but these values lead to the same closure temperature in their respective annealing equations.

The authors could also be a bit more clear when they discuss constraint boxes. To the extent that imposed constraints embody reliable independent geological information, they are *always* proper to add. Similarly, if the geology justifies them, there is no such thing as "excessively tiny" (line 473) – they should be the exact size the geology says they should be. In fact, all paths that go outside those constraints contradict the geology, so why should they even be considered? The two downsides the authors cite seem to some extent like red herrings. First, there is the issue that QTQt seeks the simplest paths, which means it minimizes the number of t-T points. By making a constraint box, they are basically telling one of those few points where it must be, reducing the freedom of the others, and giving the appearance of precision. However, the broader credible interval given by an unconstrained QTQt is not a better reconstruction of the thermal history, because the parts of the envelope that overlap the otherwise neglected part of the true path are based on paths that violate the local geology. (Naturally, the proper solution is to use HeFTy, which will broaden envelopes where the thermal history is poorly constrained by the data ;-) ). Second, the authors caution about opening the door for assumptions [embodied in constraints] "to be heralded as geologic evidence". The simple antidote is to formalize model reporting and document the reason for each constraint, as recommended by Flowers et al. (2015), reinforcing the idea that every constraint should have a reason for being there, and the modeler should be able to state what that reason is. If the evidence underlying a constraint is uncertain, then models can be run with and without it, and the effects evaluated – simple!

The discussion features almost 2 pages on (U-Th)/He kinetics (line 514-574), which is really not the subject of the paper, and is not really further informed by any of the modeling work presented here. It is basically editorializing, and most can be omitted. It's unclear what the authors mean by the “distorted” thermal history from using uniform-rmr0 RDAAM means that a non-FT-kinetics He model should be used (line 527-529). An issue with the current alternatives (Gerin et al., 2017; Willett et al., 2017) is that they do not allow for variation of alpha recoil damage annealing kinetics at all. If these kinetics vary, these other models would not be able to capture them, either. More work needed…

It’s not clear how the authors calculated their rmr0 value for Itambe apatite (line 137); I think it might be that they excluded Si from the “others” category. This should be clarified. It’s not clear what they are trying to say in this part, and going back and forth between the two rmr0’s from 1999 and 2007 is likely to be confusing unless things are very clearly set out.

The authors should make it clear that the 0.882 value of rmr0 for end-member OH-apatite (line 185) is an adaptation of fitted rmr0-$\kappa$ values of HS apatite (0.8559, 0.2206) to the simplification that rmr0 + $\kappa$ = 1, providing approximately the same closure temperature. At least, I think that’s what they did…