Reviewer: Willy Guenthner

Our responses are to the reviewer's comments are shown in italics.

The manuscript by Fox and co-authors re-examines certain aspects of a commonly used damage-diffusivity model for the zircon (U-Th)/He system (ZRDAAM). The authors highlight that the model, as originally derived and defined by Guenthner et al. (2013), currently lacks information on kinetic uncertainty, which could in turn influence the thermal history modeling-based outcomes that rely upon ZRDAAM. In particular, the authors argue that the lack of kinetic uncertainty could create over-interpretation of and/or overconfidence in thermal history model output as applied to a recent discussion surrounding the origin of the Great Unconformity erosion surface. I found this contribution to be of importance for three reasons: 1) it stresses the need to incorporate uncertainties into the kinetic models used by many authors and thermal history modeling software packages, 2) it shows how large zircon He datasets (~40 grains) are needed to refine and extract meaningful time-temperature information, and 3) it recognizes that ZRDAAM is a work in progress that could be strengthened by further He diffusion kinetic work. In sum, I think this manuscript is a solid contribution to ongoing discussions surrounding thermal history modeling and thermochronometric data interpretation. However, given the authors stated goals of exploring calibration issues with ZRDAAM, I think the article does not go far enough in investigating multiple aspects and nuances of the current ZRDAAM calibration. My recommendation is major revisions, which should focus on expanding the scope of the current work. I have several general comments that I would like to see the authors address before this article is published.

We thank Willy Guenthner for his detailed review and we have addressed as many points as possible. It is important to note that we are not really interested in producing a new damage model – we are simply interested in propagating uncertainties. Points about changing the end member crystal are useful to consider but this requires a completely different model. Furthermore, the approach we develop here can easily be modified to constrain parameters in any damage model. Our goal is to use the diffusion data directly and not rely on an intermediate regression as this ignores correlations in the data.

It should also be noted that we don't think that the endmember crystal should really ever be a real crystal. What we have done is attempted to fit all the diffusion data with the same model. In this way, all the data contribute to determining the diffusion kinetics of the end member crystals. We think this is preferable to using something like Mud Tank because then the other crystals and diffusion experiments would not be utilised.

First, the authors focus heavily on the theoretically pristine endmember zircon, and the fact that this value is an extrapolation that, correctly, could have a large degree of uncertainty to it. It is difficult in the current preprint to determine how important this endmember is in the context of real datasets. Let me expand. When the ZRDAAM was derived, a proposed "zero-damage" grain was used to make the model setup more straight-forward, and to account for potential scenarios where users may want to model thermal histories at very short time scales. But nearly all grains, especially those used for deep-time problems, will very quickly (within ~5 myr for a grain with 500 ppm eU) accumulate damage levels in excess of the Mud

Tank sample, for which empirical results do exist. So perhaps the challenge with the current ZRDAAM is less, what should we do for a zero-damage endmember, but rather, should we instead be using a sample with known kinetics as our endmember? Would the Mud Tank kinetics therefore be more appropriate here as a realistic endmember, and if so, does that reduce some of the variation in thermal history model output? Secondarily, there are both molecular dynamics (Reich et al., 2007; Saadoune et al., 2009) and empirical results from zircon-like, zero-damage orthophosphates (Farley, 2007) that attempt to place an approximation on He diffusion in a pristine zircon. How do the author's new extrapolations match up to these results as a comparison? Finally, as Ginster et al. (2019) observed, it is very unlikely that a natural zircon would return to a pristine damage state following damage accumulation, so the likelihood of encountering a natural, zero-damage zircon is remote. I would like to see the authors acknowledge some of these points and address these questions. I think the authors have illuminated a valid and important point: that the ZRDAAM as defined by Guenthner et al. 2013 needs to be recalibrated at the low damage end and uncertainty needs to be accounted for. However, as I've hopefully conveyed, I would place the focus more upon using an endmember for which we can obtain kinetic information, rather than (as was done in Guenthner et al., 2013 and is done here) an idealized, extrapolated endmember that does not likely exist in nature.

We have discussed this point above because we feel like this is the main point of the discussion. We have also addressed this in the text by reiterating that we are not aiming to produce a new model. We just want to propagate uncertainty.

Second, the authors highlight a data set from Minnesota as an example with which to test their newly constrained ZRDAAM. My concern with this modeling setup, as articulated in a line specific comment below, is that an apples-to-apples comparison to the McDannell et al. 2022 study would include many more iterations. In that study, 500,000 burn-in, and 500,000 post were used, whereas here it appears that only 100,000 burn-in and 100,000 post are used. That is, the higher amount of uncertainty as to when Neoproterozoic cooling initiates could be that the model was not run for long enough.

This is one reason why the models would be different. But there are many other reasons. We are not comparing the models to the McDannell model, but rather the results with different model parameters.

Third, the authors call for additional diffusion experiments and new data to be added to the ZRDAAM, which I certainly agree with. I would point out though that some of these data have already been collected by Ginster and are published in her PhD dissertation, available through multiple repositories, including the University of Arizona library. This work is yet to be published in a peer-reviewed manuscript, but the data essentially agree with the Guenthner et al. (2013) diffusion data, which suggests that the variation (and therefore the uncertainty) in measurable diffusion kinetics is lower than for extrapolated endmembers. The point should be emphasized that, from an experimental perspective, the ZRDAAM remains on solid footing, although perhaps the endmember kinetics need to be recast, as my previous comment suggests.

Totally agree with this point. Again, we just want to propagate uncertainty. However, we don't think it is appropriate to include the data in the inversion for the diffusion kinetics. These data are unpublished and we would rather wait until Ginster publishes her thesis. Of course, we can redo the calibration again.

Line specific comments:

Line 73: Should acknowledge here that alternatives using an arguably more direct measure of alpha dose (Raman spectroscopy) have been proposed (Ginster et al. 2019) and incorporated into ZRDAAM (Guenthner, 2021). I would also elaborate the discussion here to comment on the different styles of damage that may influence He diffusion kinetics (i.e. alpha ejection, alpha recoil, fission track). The diffusion kinetics are calibrated to alpha dose (essentially alpha recoil), but the mode of annealing is debated and as yet not fully resolved. The Ginster et al. (2019) data and the model demonstration by Guenthner (2021) of these data are particularly salient given my comments above concerning an idealized zero-damage endmember.

We have attempted to keep the paper simple and focus on a single message. Already, this paper is quite complicated.

Line 85: An additional explanation here is that, as McDannell et al. 2022 showed, these models needed to be run for many more path iterations (at least 500,000 pre and post burn-in). In the Thurston et al. (2022) study, we ran models for only 100,000 pre and post burn-in, which was admittedly likely not enough. To this point, we have re-run some of these models with the greater number of paths (currently unpublished) and indeed the earlier portions of the time-temperature history remain under-constrained by zircon He (as discussed in Thurston et al., 2022). However, the late Miocene cooling remains robust.

We have added that parameter space might not have been sampled sufficiently.

Line 99: The phrase "varies by hundreds of millions of years" should be qualified here. As you suggest later in the manuscript, the variation is dependent on the number of grain inputs you have and the spread in date-eU space of those inputs. Moreover, my understanding (this could be clearer, see next comment below at line 238) is that this is for scenarios that incorporate the 2 sigma from the full kinetic distribution.

This has been removed.

Line 238: The focus in this paragraph on fixed endmembers seems out of place with one theme of this manuscript: kinetic uncertainties should be sampled in the rjMCMC approach. Some of this could be my confusion: am I correct that the models were run with Ea and D0 values that represent the 2 sigma of the kinetic distribution? If I'm not correct, then please more thoroughly explain how or why these specific kinetics were selected. If I am correct, then I understand that the authors are perhaps trying to show the worst-case, 2 sigma extremes from their distribution, but why not incorporate the full probability distribution as shown in figure 2 and sample that? The authors mention further below that computation limitations prohibit this exercise, but much of the discussion and implications seem to be

cast in light of the highest possible amount of variation. If the modeling incorporates the full distribution (and samples it) is the situation as dire? I am really curious to see the outcome of a modified MCMC proposal algorithm with a selection statement that samples the kinetic distribution.

It is not clear to us how to do this at present. We have a produced a new version of QTQt that explicitly samples from our inferred model parameters. In this way the model does not attempt to find the best combination of diffusion parameters – the diffusion experiments are the best way to do this. Instead it incorporates this uncertainty into the analysis.

Line 249: For a better apples-to-apples comparison here with the McDannell et al. (2022) study, 500k pre and 500k post burn-in is needed. As we have seen (and learned, see the comment about Thurston up above) these deep-time problems require at a minimum 1,000,000 path iterations.

We tested this point and ran models with more and fewer iterations in the burn in phase. This doesn't seem to be a big problem if the chain is properly sampling parameter space.

Line 294: Are the authors suggesting here that the binning and averaging approach has limitations? The point I'm most struck by, as the authors allude to, is that binning and averaging can eliminate the sensitivity of the whole data set by removing portions of date-eU space.

We agree that this is problematic and would suggest not binning and averaging data. Instead, the distribution of age as a function of eU could be mapped out more completely.

Willy Guenthner

UIUC

Citation: https://doi.org/10.5194/gchron-2022-23-RC5