We thank the reviewers for their helpful comments and feedback and address specific concerns below in the table with each individual comment in the left column and our response in the right column.

Comments from reviewer: Marissa Tremblay

Why was the 'accept more complex models' option added when the proposal algorithm for new T-t paths was changed? From reading the manuscript, it seems like the authors think there are limited (or perhaps no) scenarios where this option should be selected	This change was made to allow for comparison with the previous algorithm and it provides visually simpler results. More technically it shows the minimum structure allowed in the thermal history. So instead of filling up the prior range with many time-temperature points when sampling thermal histories in unconstrained parts of the time range, this option will tend to put 1 or 2 points. While this can fill up the prior temperature range, it takes a lot longer to see this. The difference in the algorithms is most evident when changing the model dimension (particularly adding a new time-temp point) The revised algorithm draws new time-temperature points directly from the prior (independent of the current model) whereas the previous algorithm modified (perturbed) existing models (add a new point by perturbing somewhere along the linear segment joining two existing points), so the new model is conditional on the current model. We can add a bit more of this background in section 3.
In the example thermal histories, why model only AFT or AHe synthetic data separately? This choice is never justified, but presumably synthetic ages can be generated for both systems, at least for the first two prescribed T-t histories in the manuscript, and then used to evaluate the inverse models with accept vs. reject complex T-t paths. I suggest this because, in real thermochronology applications, it's generally not recommended (and to my knowledge avoided by most) to have a single sample with just AHe dates like in the second T-t example, unless there is a very large spread in eU that would result in very different thermal sensitivities.	This is a valid point. However, as the focus of this technical note was on showing the difference between accepting and rejecting more complex models. Given that we chose to use simple datasets to avoid the potential confusion with multiple data types impacting different parts of the thermal history (although the AHe dataset has a spread in eU). Also this demonstrates that choice of reject/accept complex is relevant in even simple scenarios. Once aware of the implications of this choice, we feel that users can (should) examine the sensitivity for their own specific datasets
I encourage the authors to consider whether their points could be better illustrated by using real thermochronology data, in addition to synthetic data alone. Are there real datasets that you can highlight, either from your own work or the literature, where you can demonstrate that over-interpretation might occur from accepting more complex paths that do not improve the model fit to the data? I thought these kinds of examples were very effective in the Abbey et al. (2023) paper.	Again, as in the panel above, we are trying to make people aware of the differences in the two options and would prefer to avoid any potential controversy in choosing a specific data set. However, there was a recent article in PNAS by Zhan et al., (2026) and a comment from McDannell et al., (2026) that exemplifies the potential for different interpretations. We will include reference to these in our comparison discussion.

Lines 9-11: This sentence is challenging to read. Suggested revision: Low-temperature thermochronology can provide record near surface thermal perturbations caused by processes like faulting, erosion, intrusion, or hydrothermal circulation. Modelling low-temperature thermochronology data to understand the timing and duration of these geologic processes has become relatively routine.	Thank you for the suggest we are happy to revise.
Line 16: Change 'the accept' to 'to accept'	Will do.
Lines 22-24: Clarify here: 'To verify if the choice to accept or reject complex models plays an important part in model interpretation, we suggest this decision is always tested by running the same model using the same data and input parameters but rejecting the complex models, and comparing the model outputs.' To my mind, rejecting the more complex T-t paths means the model is inherently not the same.	The other reviewer, Peter van der Beek, made the point that we do not clearly state what users should do. This will also be a good place to clarify that we encourage modelers to look at all three output models from both runs accepting and rejecting more complex models. Fortunately, the differences between the two different models is most significant where the data do not have any information to constrain the complexity or the form of the thermal history. Where the data do inform us, we expect the same form of thermal history. However, running both reject/accept complex allows a user to assess which parts of the thermal history are well resolved (informed by the data) or not. This is different to the choice between the Max. likelihood, Max. posterior and Expected models, which needs to be considered irrespective of the choice for accept/reject complexity. As noted above we will try to emphasize this as much as we can throughout the manuscript
Line 40: Suggest revising to say "a user specifies a single set of model parameters".	Will do.
Lines 70-71: I assume that this statement about what many researchers do may be informed by the authors' personal experiences (e.g., running pre-conference modelling workshops and corresponding with authors via email). However, it seems like a statement like this should somehow be backed up, either with examples from the literature or by explaining how the authors arrived at this assessment.	Thank you for pointing this out. The example from Zhan/McDannell is appropriate as this was explicitly addressed in the comment/reply. We will look for additional examples from the literature, but to be honest we are wary of pointing the finger/upsetting individuals.
Lines 80-92: I recommend reorganizing this paragraph to first describe how the original QTQt proposed and made decisions about accepting new T-t paths, and then explaining how the newer versions of QTQt proposes and makes decisions about accepting new T-t paths. Right now the two versions are discussed somewhat in parallel and it is a bit challenging to follow. I also think the authors should emphasize that if the 'reject more complex models' option is selected in newer versions of QTQt, more complex models, i.e., T-t paths with	OK, we can do some re-organizing. As explained in a reply above, the difference is really how new time-temperature points are proposed. The old algorithm uses a perturbation of the current model to choose the new point (so a new point in the proposed model is conditional on the current model). The new algorithm samples a new point directly from the prior (so the new point is independent). Both algorithms will tend to accept a new point if the data fit improves (i.e., the likelihood increases). The

more change points, will be allowed but only if they improve the data fit (if I understand correctly). Currently this point is a bit unclear in this paragraph as written.	new algorithm (as given in Licciardi et al; 2021, and others) can add extra points which have no impact on the likelihood. These more complex models that do not change the data fit can be accepted as the acceptance probability is just the ratio of the likelihoods (which in this case is 1 as the likelihood does not change). The old algorithm (as given in Gallagher JGR 2012) includes a terms that acts as a penalty more directly for adding in points that do not change the data fit. We do make a statement along these lines on line 285, but can bring it in earlier in the manuscript as well.
Line 147: 'The timing is not as accurate' - Be more specific and say the maximum heating occurs later.	Will do.
Lines 154-158: These statements are pretty vague – can you provide a more explicit example?	We will reference the PNAS Zhan/McDannell example here as it is current, concerned with exactly what we are saying here, and will hopefully refrain from "pointing the finger" too much.
Figure 1. This figure is a lot to digest with so many sub-figures. One way to make this easier for the reader to understand would be to combine the top panel in Figure 1A with Figure 2, as they are both showing information about number of T-t points. Then the maximum posterior and maximum likelihood T-t paths could be a separate figure, and the data predictions (current panel B) could be a third figure. In addition, for the top figure in panel A, at first I did not recognize that the scale was different for the left- and right-hand axes. I recommend making the scales the same so that the reader can see that the yellow/green likelihood chain overall has many more T-t points than the teal or magenta chains. For panel B, I recommend adding a line or point to show what the 'true' synthetic age is. I also suggest including the model results for the track length distribution compared to the 'true' synthetic track length distribution, since the models are also attempting to fit these data. The conclusions about model complexity should still hold for the track length data.	Thank you for the suggestion we were also struggling with the best way to present these figures. This is a good way to separate set-up and T-t points from outputs and predictions. We will try splitting into 3 figures. For the scale difference in panel A, we did this because if we put everything on the same scale we see very little structure for the 2 curves relevant to the left side scale. We'd like to keep the different scales but will very explicitly draw attention to the different scale in the caption. We will add a true age line to predictions figure. Yes we can add the TLD predictions – they were effectively the same. To be consistent with the distribution on predicted ages, we will need to add in the range on the predicted length distributions too, which is possible with QTQt. We will then be sure to add text about the comparison with the track length distributions and predictions.
Line 197: RDAAM needs to be defined.	Will do.
Line 199: Since the AHe paths are not actually the same as those in Wolf et al. (1998), it would be much simpler for the reader to just refer to the paths the authors choose for their AHe model as paths 1 and 2 and not try to link them back to comparable/similar paths in earlier publications.	The paths were modified in the cited manuscripts (Murray et al., 2022; Abbey et al., 2023; Stevens-Goddard et al., 2026), and they are all called Path 1 and Path 3 in those three papers, but if it is confusing we are happy to change the naming here and then say that they are the same as paths 1&3 in previous pubs.
Figure 3. This figure needs the different panels to be lettered, and the figure caption needs to refer to this lettering. This will make it much easier to understand which panels in the figure	Will do, thanks for the catch.

are being described by different sentences in the caption.	
Lines 232-233: “skews even more to capture the dearth of possible pre-40 Ma temperature changes” I am not sure what you mean by this, please clarify.	We meant that there is much more range in the temps of the pre-40 Ma history, mainly more cooler temps. This pulls the expected model up to slightly cooler temps before 40Ma and the 95% cred-intervals expand into the cooler temp space pre-40Ma. We can modify the text here to more explicitly point out what we are seeing.
Lines 265-267: The fact that neither the complex nor simple models predicts the synthetic T-t history is really interesting. I recommend expanding the discussion here a bit because this says something about the resolution when using a single chronometer like the AHe system, regardless of the modeling choices that are made. (This may also be a good reason to consider including models that predict ages for more than one chronometer, as mentioned above).	The MaxLike, reject model does a very good job of reproducing the true model. Perhaps we could plot the true input on the two plots with the inversion results to really visually show the comparison. Multiple data types adds more complexity (potential for confusion) so we prefer not to add more complexity (there was a large spread in eU for the AHe data so we can emphasize that). We recommend that users do the same exercise with their own data which may include multiple chronometers.
Lines 269-270: The algorithm details are not actually described here but earlier (lines 80-92).	That is correct. The term “here” meant “in this manuscript”, we can change that to say “above”.
Lines 279-282: I think a more descriptive summary/synthesis of what was learned from the modelling examples is needed here. The rest of this paragraph is more focused on recommendations and should probably be a separate paragraph from such a synthesis/summary of the new model results presented.	We will reiterate what the different examples showed, and refer back to the new work from Zhan and comment from McDannell. Then we can separate the paragraph at the sentence “In reality,…”