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: Peter van der Beek

This manuscript provides some very useful guidance for users of the popular QTQt inverse thermal history modelling code for thermochronometric data, and therefore fits well into the Special Issue of Geochronology to which it was submitted. The authors focus on a (relatively little known?) choice option in the more recent versions of QTQt, to accept or reject more complex thermal histories (i.e., histories characterised by more T,t points) that fit the data similarly to simpler models. In my understanding, earlier versions of QTQt rejected such models by default. However, I actually found the comparison between the different "optimal" models (i.e., the maximum likelihood, maximum posterior and expected models) the most interesting aspect of this contribution. I felt like I learnt something on that aspect from this manuscript (in particular why the expected models so often appear to suggest reheating), whereas the differences between the inversion runs that accept or reject more complex paths are rather subtle and the manuscript does not end up providing very detailed guidance on whether or not (or when) to accept more complex models in inversions.	We recommend looking at both outputs (i.e., accepting and rejecting more complex models). This will help to see where the data have control on the output vs where the introduction of more complexity simply adds more points in areas of the model space that is not controlled by the data. We have suggested this in line 153-154 for example but we will be more explicit in stating this recommendation both in the abstract and in section 5 "choosing to accept or reject…" as well.
I am wondering whether the authors would consider putting more emphasis on the comparison between the different optimal models, at the expense of the comparison between models allowing more or less complexity. I believe this would also serve the community very well, as I have the impression many users of QTQt are often at a loss which of these options to present and use, and I have been wondering myself about the apparent reheating shown by so many QTQt inversion outcomes. I realise some of this material was covered in the Abbey and Gallagher (2023) paper of which this technical note can nearly be considered a "spinoff", but since that was a quite long and involved treatment of many different aspects of QTQt modelling, the message concerning the different representative thermal history paths may not have come through as clearly. I am thus leaving this suggestion here for the authors' consideration.	These differences (between Expected, MaxLike, and MaxPost) were covered in Abbey et al., 2023, but we could perhaps reiterate/emphasize the differences here. For example in section 4 we can highlight more clearly about the differences in the 3 outputs (particularly the predicted ages from the expected model relative to the 95% credible ranges on predicted ages – these show clearly that the predicted ages can be too old due to reheating maximum temperatures in the expected model being underestimated relative to the individual sampled models, due to the smoothing implicit in the averaging process). We can remind readers that these are discussed and exemplified in Abbey et al., 2023 and we can show how they are similarly behaving in our examples here. In section 5 when we suggest looking at both reject & accept complexity models we can also reiterate that the MaxLike, MaxPost, and Expected should all be looked at as was suggested in Abbey et al., 2023).
I have a few general comments on the manuscript and a series of more minor editorial issues. Overall, the writing could be tightened up a bit, as the current manuscript can be a bit confusing in	Thank you, the provided PDF comments are helpful and will indeed help to tighten up and clarify the writing.

places due to imprecise and sometimes somewhat colloquial expressions. I am providing the editorial comments as edits on an attached pdf file of the manuscript and focus here on the more general comments only.	
First, although I realise this is a technical note that is supposed to be short, I do think that some concepts could do with a very short (1-line) explanation or at least a reference to where readers can find more information about them. For instance, the “likelihood” of a model is used throughout but not defined. Also, what is meant by the maximum-likelihood, maximum posterior, and expected models should be made clear somewhere.	In section 2, we will add a few sentences to explain these concepts as well as references to other papers that explain them in more detail
In line 148, it is stated that the maximum posterior model is “possibly the preferred individual best model in Bayesian approaches”, but without justification for this statement. Again, most of this is already in the literature, but it really helps the reader to have a short explanation the first time these terms are used.	Related to the above comment we will add more here and references to explain why this is the preferred model in Basian approaches.
It is argued in several places in the manuscript that, in more complex models, time-temperature points will only be added in the parts of the model space that are unconstrained by the data (e.g., lines 76-77; 273-274). This is not intuitive for me and could therefore do with some justification / explanation. I would agree that model *complexity* is increased in those parts of the model space that are unconstrained by the data (and this is nicely shown in Figs. 1, 3, and 4), but could very well imagine that additional t,T points are added throughout the model space, but do not influence the complexity of the models where these are well constrained by the data (in other words, in those parts of the model space one could add additional t,T points as long as they align with the well-constrained thermal history)	This is a good point. We will clarify this more at lines 76-77; 273-274 as well as any other locations where we refer to this concept. We will annotate the figures pointing out the part of the model constrained by the data vs not. Effectively the range in accepted thermal histories approaches the temperature prior range when the data have no information on the thermal history. As the prior on the number of time-temperature points is uniform, we impose no prior constraint on model complexity (for example compared to a Poisson prior, which would impose a preferred average value for the number of time-temp points). Then if the data have no information (i.e., do not care what happens for temperature over a given time range, e.g., before the time of total resetting for all grains) there is no data constraint on the complexity there.
After reading the manuscript, the one question that lingers is the actual usefulness of the expected model in defining the thermal history. Even for the relatively simple input thermal histories shown in Figs. 1 and 2, it does not capture the thermal histories particularly well, and one wonders whether some other weighting scheme for this model might produce “better” results? As it is, it appears quite heavily influenced by (the probably fairly large number of) models that do not fit the data particularly well. In most cases, the parts of the model space characterised by high relative probability appear to capture the input thermal history significantly better. This ties in to a long-standing discussion within the community about the significance of relative probability and whether/how we should use that as a criterion to select predicted thermal histories.	It is very important to look at the observed data and the equivalent predicted values (including the range of predicted values for all accepted models) when interpreting the output models (but especially the expected model). We will make this point more clearly in the manuscript when describing the differences between the model outputs.