https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


# Technical Note: Remarks on Assessing Complexity in Thermal History Models

Alyssa L. Abbey[1] and Kerry Gallagher[2]

[1]Department of Earth Science, California State University, Long Beach, 1250 Bellflower Boulevard, Long Beach, CA 90840, USA
[2]Géosciences Rennes, UMR6118, CNRS, Université de Rennes, Rennes, France

*Correspondence to*: Alyssa L. Abbey (alyssa.abbey@csulb.edu)

**Abstract.** Modelling low-temperature thermochronology data to understand geological history relating to near-surface thermal perturbations caused by processes like faulting, erosion, intrusion, or hydrothermal circulation, has become relatively routine. However, it is clear that not all modelling efforts include rigorous testing of various modelling decisions. This happens in part because of a lack of understanding about each of the different model parameters and how modifications to those parameters may control different model outputs or predictions. In an effort to reduce ambiguity around how model complexity is dealt with in the modelling program QTQt, we delve into the details behind the algorithm that accepts and/or rejects models with greater complexity (i.e., many time-temperature points within a thermal history), and explore example thermal histories to show the effect of choosing the accept or reject more complex models that do not improve the data fit. Generally, where the data control the model outputs and the data fit is good the model outputs and age predictions are indistinguishable. When the choice is made to accept more complex models, users must be aware that this choice adds more complexity in the areas of the model space that are not controlled by the data and effectively smooths the expected thermal history. Because of this effect, caution should be used when interpreting the expected thermal history from a run that accepts more complex models. To verify if the choice to accept or reject more complex models plays an important part in model interpretation, we suggest this decision is always tested by running the same model rejecting the complex models and comparing the model outputs.

## 1 Introduction

Low-temperature thermochronology is a widely used tool for researchers who are interested in questions about near surface thermal changes on Earth. Such changes could be caused by various geological processes such as faulting, erosion, magmatism, and/or circulation of hydrothermal fluids. Thermochronological data are used to infer the age or timing of such events, rates of temperature change, and long duration (billion year) thermal histories that can incorporate multiple thermal events. To make such inferences, thermochronometric data are modelled with other known information (e.g., independent geochronology), assumptions (e.g., paleo-geothermal gradient), and modelling parameters (e.g., time-temperature space to

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


explore, statistical criteria, diffusion and annealing functions). The decisions made regarding model input and interpretation of model outputs will impact the final geologic interpretation. In light of this, the thermochronological community has recognized the importance of such decisions and has begun creating resources, protocols, and structures for documenting and reporting modelling decisions to ensure more transparency, reproducibility, and continuing education for new and experienced users (Flowers et al., 2015; Ketcham et al., 2022; Murray et al., 2022; Flowers et al., 2023a&b; Abbey et al., 2023).

## 2 Background and motivation

QTQt (Gallagher, 2012) is a popular, user-friendly program that is used to model thermochronometric data. It can be run in two modes:

(1) forward modelling–a user specifies model parameters as single values and tests the predictions with that set of parameters against observations

(2) inverse modelling–a user specifies an input distribution of possible values on the model parameters, and these distributions are sampled and updated based on comparisons of the predictions with the observations.

The inverse modelling method in QTQt is known as Bayesian trans-dimensional Markov Chain Monte Carlo (TDMCMC). The input distributions are known as the *priors* which define the range of values of temperature and time considered possible for the given data. The prior will typically depend on the types of data we have and the temperatures for which each data type is sensitive on geological timescales. The goal is to modify the input distribution, informed by the available data to produce an output, or *posterior,* distribution. This distribution represents a population of thermal histories and as such allows us to assess the uncertainties in the inferred thermal histories. Where the data provide information on the thermal history, we expect the posterior to differ from the prior, while for those parts that are not constrained by the data, the posterior is close or even equal to the prior. This latter phrase means that the data tell us nothing more about that part of the thermal history than what was already defined in the prior.

As with any modeling approach QTQt requires various decisions to be made concerning input and control parameters. Ideally, these decisions should be systematically assessed for the impact one decision may have on the model outputs. Abbey et al. (2023) exemplified this approach by doing sensitivity tests (impact on the inferred thermal history) considering the role of different data types (e.g., apatite (U-Th-Sm)/He vs. apatite fission track), modifying the uncertainty on input data, adding a starting point constraint (e.g., a stratigraphic age for a sediment, or the formation age of an igneous rock) and changing the range of time-temperature priors. These tests show that while making different decisions can produce different model outputs, overall, it is the data that controls the important or well resolved parts of the output.

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


The examples presented by Abbey et al. (2023) addressed only a few of the decisions that can be made when modeling in QTQt. However, performing sensitivity tests on every individual decision made for every modeling project is not practical. Furthermore, although the need for such testing is evident, both new and experienced researchers may find it difficult to know which decisions should be tested. The sensitivity testing presented in Murray et al. (2022) and Abbey et al. (2023) helps provide a guide for researchers to recognize in which scenarios certain decision tests will be most useful. Additionally, there are several publications that describe the different nuances and specialized modifications that have been added to QTQt for researchers to cater their modeling to more specific geologic questions (e.g., Gallagher, 2012; Cogné et al., 2011; Georgieva et al., 2019). These examples also act as useful guides to determine which decisions may be more important for different research questions. However, not all aspects of QTQt are fully described in the published literature, and many researchers still model data with default parameters and little exploration into each of the parameters that can be modified.

## 3 Model choices regarding model complexity

Here we focus on the choice to accept or reject more complex models that do not improve data fit. This is made when setting the MCMC parameters before a model run. Model complexity here is defined by the number of discrete time-temperature (t-T) points used to approximate the continuous geological thermal history. The choice a user can make in this case is to allow a potentially large number of time-temperature points (i.e., accept complex models), which typically come from the parts of the model space that are unconstrained by the data, or to reject such models if the additional complexity make no difference in the data fit (i.e., reject complex models).

The option to make such a choice first appeared in QTQt around 2019 (post QTQt version 5.7) which implemented a similar algorithm to that presented in Licciardi et al. (2020) and see also Mosegaard and Tarantola (1995) and Agostinett and Malinverno (2010) for more details. This approach is simpler than that implemented in the original version of QTQt (Gallagher, 2012). In this simpler algorithm, a new time-temperature (t-T) point is proposed simply by drawing it directly from the prior distributions and so is independent of the current model. The algorithm in Gallagher (2012) proposes a new t-T point based on a transformation or perturbation of the current model t-T points. In this case, the proposed model is conditional, or dependent, on the current model. In both cases, the prior on the number of model parameters is defined to be uniform between a minimum and maximum number of t-T points (2 to 50 by default). This means that we are starting with an assumption that a thermal history with 50 points is as probable as one with 2, thus, we have no prior preference on the complexity and are happy to accept models from this range of complexity. However, the original algorithm in QTQt penalizes more complex models during the sampling if there is no improvement in the fit to the data. In other words, if the data neither require nor justify extra complexity a simpler model is preferred (see Mackay, 2003 and Sambridge et al., 2006 for explanations of how this operates in general).

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


In contrast, the modified (newer) algorithm is formulated so that model rejection does not occur in the same way. This becomes most obvious in the unconstrained parts of the model space. There the estimated posterior will equal the prior and there are no constraints on the model complexity. In inverse theory jargon, this is known as the model null space such that one can add model parameters that make no difference to the predictions.

A simple example case is

$$A = B + C$$

where A is the data and B and C are the unknown model parameters. If we say A = 5, there is an infinite number of values of B and C that can be added to give a sum equal to A. Null models have parameter values that make no difference to the predictions. So, they can always be added into the model, but the data do not give information about them.

So, if B = 3 and C = 2, null parameters can be added as

$$A = B + C + B_{null} + C_{null}$$

If $B_{null} = -C_{null}$, their contribution to the sum is zero and the predicted value of A does not change.

Therefore, if the choice is made to accept more complex models (if the data fit does not change), there can be numerous t-T points that are not required (i.e., from the null space) and make no difference to the data fit. This behavior results partly by the fact that a uniform prior is assumed on the number of t-T points, meaning that we are happy to accept the most complex models as mentioned earlier. The examples presented in Green (1995) and Jasra et al. (2005) use a more targeted prior on the number of model parameters (e.g., Poisson distribution). This implicitly assigns different penalties on complexity based directly on the number of model parameters through the prior. Concerning the accept/reject complex models with a uniform prior, it is expected that the parts of the model space that are constrained, or that the data contain information about, will be the same (or similar) for both assumptions.

Here, we explore this expectation that both approaches (accepting or rejecting more complex models) will give the same result for the parts of the thermal history that are constrained by the data. We use synthetic data produced from forward modeling a simple heating and cooling scenario with AFT data and examples with apatite (U-Th-Sm)/He (AHe) data exploring both a simple and a more complex thermal history (modified from Wolf et al., 1998; Abbey et al., 2023).

## 4 Examples of choosing to include the complex paths or not

As mentioned above, predictions are not influenced by adding additional null model parameters and so these parameters are not constrained by the data. In thermochronology, this can occur for example, when there are temperatures that are much lower than a later maximum temperature (see Fig. 1a). This is due to the dominant effect of maximum temperatures on both diffusion and annealing. When choosing to accept more complex models for which the likelihood does not change, the acceptance probability for a proposed model with null parameters is always 1. This means we will always accept these more

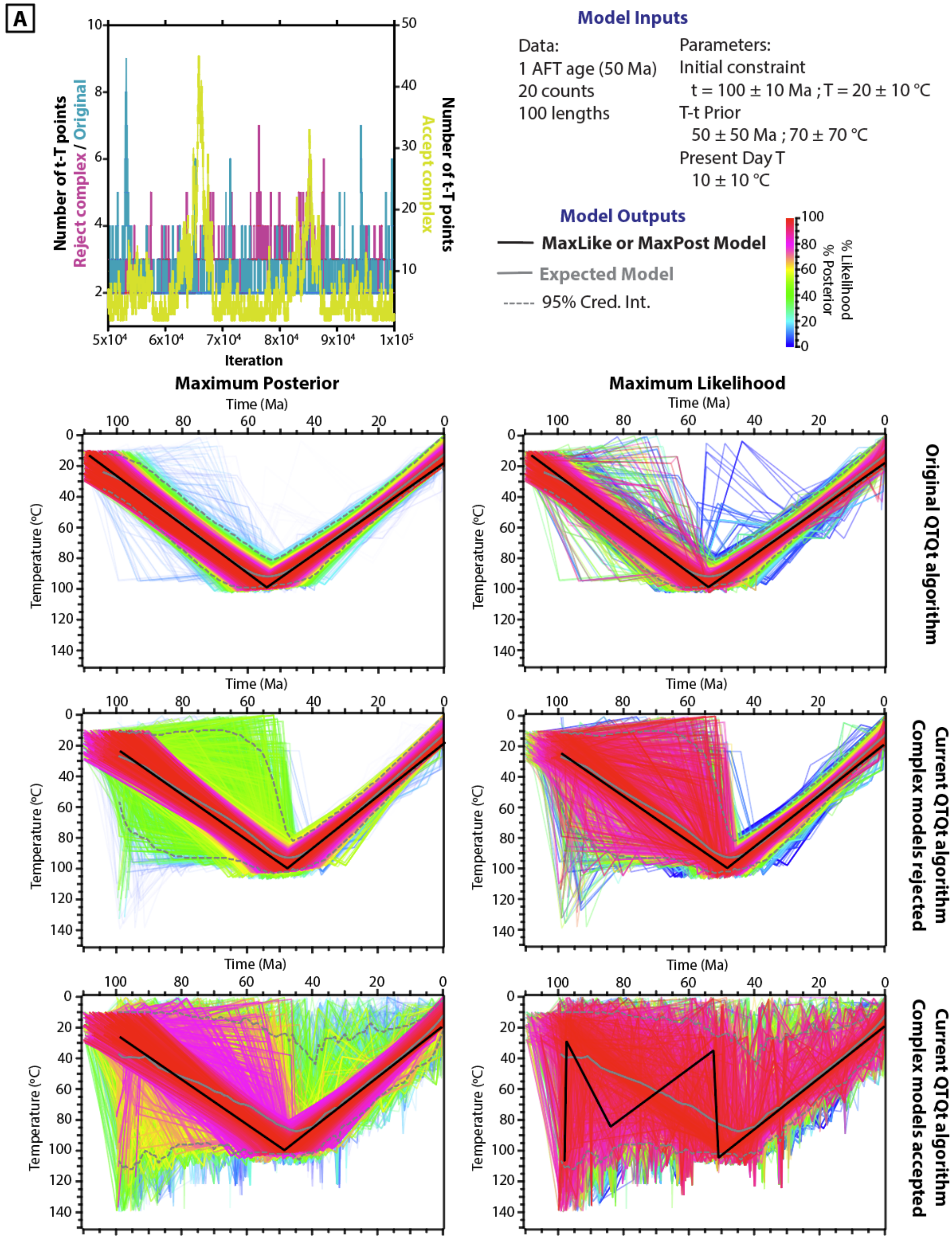
https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


complex models, even though the predictions of the data are not influenced by the extra complexity. In the inverse modelling sense, the data have no information about these unimportant parameters, or in the forward modelling sense, the unimportant parameters do not change how the data is predicted. Either way, this does not help obtain simpler models in those unconstrained parts of the model space. However, the posterior distribution can help focus on which parts of the model space actually are constrained, or what structure in the thermal history is required to explain the data. This is because the posterior distribution in unconstrained parts of the model space will be the same as the prior. The posterior can be thought of as an updated version of the prior, the prior being updated by the information contained in the data. If there is no information in the data about certain parts of the model space, the prior and posterior will be the same there.

### 4.1 Example 1: simple re-heating and cooling with AFT data

To illustrate this point, consider a simple heating-cooling thermal history, starting at 20 °C (100 Ma) heating to a maximum temperature of 100 °C (50 Ma) and then cooling to 20 °C at the present day. We produced synthetic apatite fission track data from this thermal history, and used these synthetic data to run inverse models at 100,000 iterations each with 50,000 post-burn-in. We used an old version of QTQt (v. 5.7.0) with the original algorithm from Gallagher (2012) and the current version 5.9.0) with the option to choose to reject or accept more complex models that do not change the likelihood (Figs. 1 & 2). While we expect the post-maximum temperature cooling to be well constrained in all models, both the original version and the reject complex models should prefer models that go straight from an initial low temperature to the maximum temperature. Both can allow some variation in the complexity before the time of the maximum temperature. Accepting more complex (but potentially unconstrained) models, will tend to fill up the time-temperature space before the time of the maximum temperature up to (or perhaps just a little lower) than the maximum temperature (Fig. 1). We also see that, for the accept complex models option, the maximum likelihood model (i.e., the best data fitting model) has artifacts, or structure not in the true model, and overestimates the maximum temperature (due to the rapid heating just before). The expected model, in this case, tends to underestimate the maximum temperature and its timing is not as accurate as the other two approaches (Fig. 1a). The maximum posterior models, possibly the preferred individual best model in Bayesian approaches, are similar in all 3 cases However, one clear effect of the simpler algorithm relative to the original is to expand the uncertainty, to approach the width of the prior, seen by the 95% credible intervals (although not exactly as they are the 95% credible intervals rather than 100%)(Fig. 1a). This is expected to occur before the maximum temperature is reached, as suggested earlier, but accepting complex models also shows this effect to some extent in parts of the model space that we expect to be well constrained (i.e., the cooling to the present day). In practice, looking at results from both models, and their predictions is a good idea to assess the resolution of the thermal history overall (Figs. 1 & 2). This is particularly recommended for those parts of the thermal history that prefer simpler models that may be overinterpreted in terms of geological significance if relying just on the thermochronometric data. In practice, this can be important for very long (deep time) duration thermal histories based on limited data or data that cannot reliably provide information on the detail of the thermal history structure over such long timescales.

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026

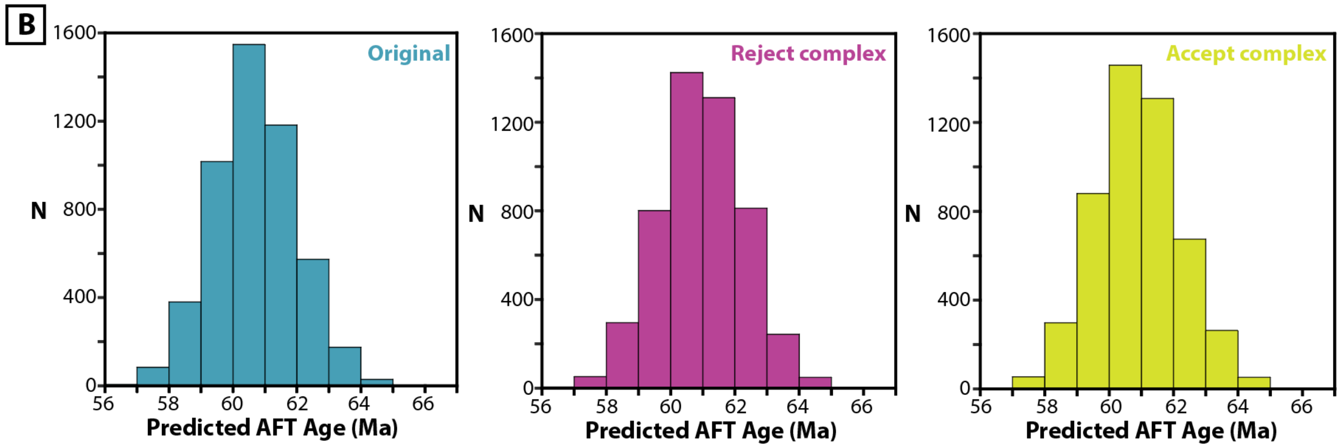




**Figure 1: (A) Simple example with a heating/cooling thermal history, starting at 20 °C at 100 Ma, a maximum temperature of 100 °C at 50 Ma and present-day temperature of 20 °C. This predicts an AFT age of around 62 Ma and a bimodal length distribution with a mean length of 11.6 microns. Here we compare results of running 100,000 iterations (including 50,000 burn-in) with the original QTQt algorithm (Gallagher, 2012) and with the modified algorithm both rejecting and accepting more complex paths. Top left panel: comparison between the likelihood chains illustrating the increase in t-T points per iteration when the more complex paths are accepted (right y-axis). Time-temperature history outputs are plotted for both the Maximum Likelihood (right column) and Maximum Posterior (left column) with paths color-coded by likelihood or posterior respectively scaled to range from 0 (= minimum value) to 100% (= maximum value). The expected cooling history and 95% credible intervals are overlaid (grey solid line and grey dashed lines). (B) Data predictions for the inverse models run using the simple example in Fig. 1a with the original QTQt algorithm (Gallagher, 2012) and with the modified algorithm both rejecting and accepting more complex paths. The predicted age distributions are all effectively the same, showing that the complexity does not matter for age predictions.**

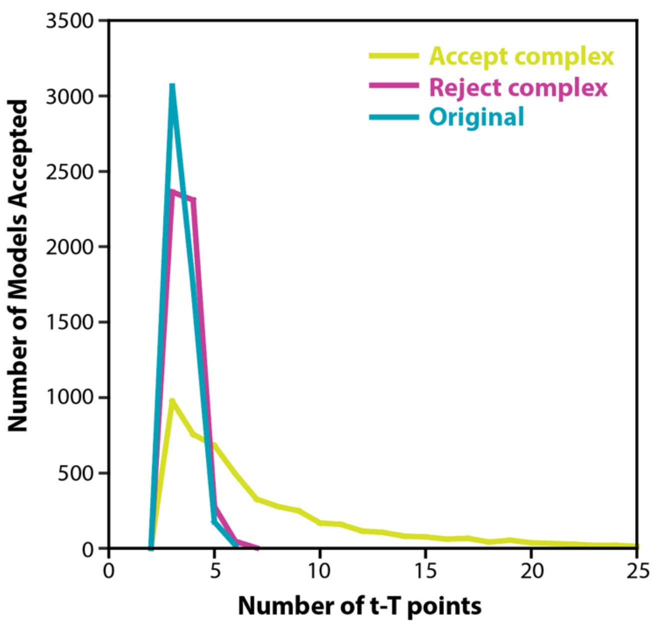



**Figure 2. The distribution on the number of time-temperature points for the different model runs in Fig. 1. All have the same peak, but the run that accepts more complex models has a long tail to more model parameters (which goes to > 40, although this plot is truncated at 25 t-T points).**

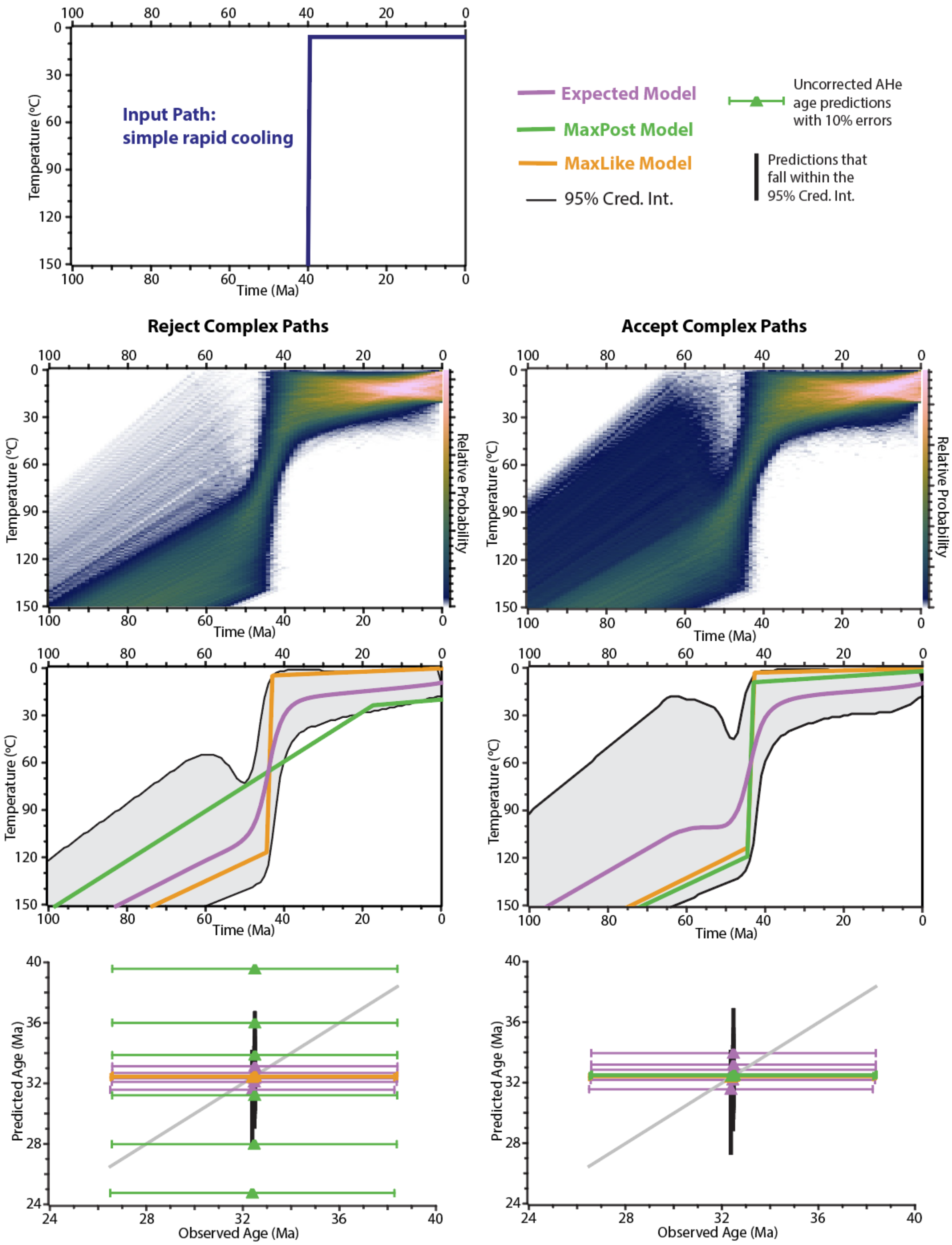
https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


### 4.2 Examples 2 & 3: rapid cooling and isothermal holding with AHe data

To further illustrate the effect of the choice to include complex paths or not in the current version of QTQt, we use two cooling history scenarios to predict synthetic multigrain apatite (U-Th-Sm)/He (AHe) data. The first is a simple cooling history–starting hot and cooling rapidly to surface temperatures then remaining at surface temperatures until present, and the second is a slightly more complex thermal history beginning with some cooling, residence in the AHe partial retention zone (PRZ), then more rapid cooling to surface temperatures. These two cooling paths are modified versions of the cooling paths adopted by Wolf et al. (1998). The initial publication by Wolf et al. (1998) illustrated five different cooling paths that yielded a 40 Ma AHe age. However, these cooling histories and resultant ages were not created with radiation damage in mind. The sensitivity tests we show here use the Flowers et al. (2009) RDAAM, so to obtain a 40 Ma age for these two scenarios we use modified paths to obtain a 40 Ma age prediction based on the kinetics implicit in this model (Murray et al., 2022; Abbey et al., 2023; Stevens-Goddard et al., 2026). Thus, the simple cooling path (Path 1) begins at 200 °C at 40 Ma, cools to 5 °C at 39.9 Ma, and remains at 5 °C until 0 Ma. The more complex cooling path (Path 3) begins at 90 °C at 100 Ma, cools to 61 °C by 21 Ma, then cools to 5 °C by 19 Ma, and remains at 5 °C to 0 Ma (Stevens-Goddard et al., 2026). Furthermore, instead of predicting a single age for a 60 μm crystal with 60 μg/g [eU] we include predictions for 6 different 60 μm crystals with [eU] ranging from 10-300 μg/g. We include six crystals and a range of [eU] values to mimic current practices of selecting multiple single-grain aliquots per sample and to add complexity to the data as a way to see how data complexity plays with model complexity. In all tests, we include an initial constraint of 200 ± 5 °C at 200 ± 5 Ma to give the model a geologically reasonable starting point rather than letting model outputs "appear out of thin air" in the model space. As we know the forward histories used to generate the synthetic data for these tests began at 40 Ma (Path 1) and 100 Ma (Path 3) respectively, we ignore any part of the inverse modeled thermal history before the true start times i.e., before the data can potentially constrain the model (Abbey et al., 2023).

When modeling the simple cooling path (Path 1) and rejecting more complex models we see that there are a range of cooling scenarios that predict the observed ages. However, the range is narrow and mostly the cooling paths follow the known input (Fig. 3). However, because of the averaging involved in calculating the expected output model, this model is smoothed so that the rapid cooling at 40 Ma is less rapid – beginning earlier and ending later, implying a slower cooling rate (Fig. 3). The maximum likelihood model shows an almost perfect match, but the maximum posterior simplifies the cooling history to a constant cooling history at the same rate (something similar to "path 2" in Wolf et al., 1998; Murray et al., 2022; Abbey et al., 2023; Stevens-Goddard et al., 2026; Fig. 3). Looking at the fits between the observed data and the model predictions we see that the expected model and maximum likelihood model fit well, but the simplification of the posterior model produces a poor fit to the individual grain ages but probably captures the mean age adequately (Fig. 3).

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026

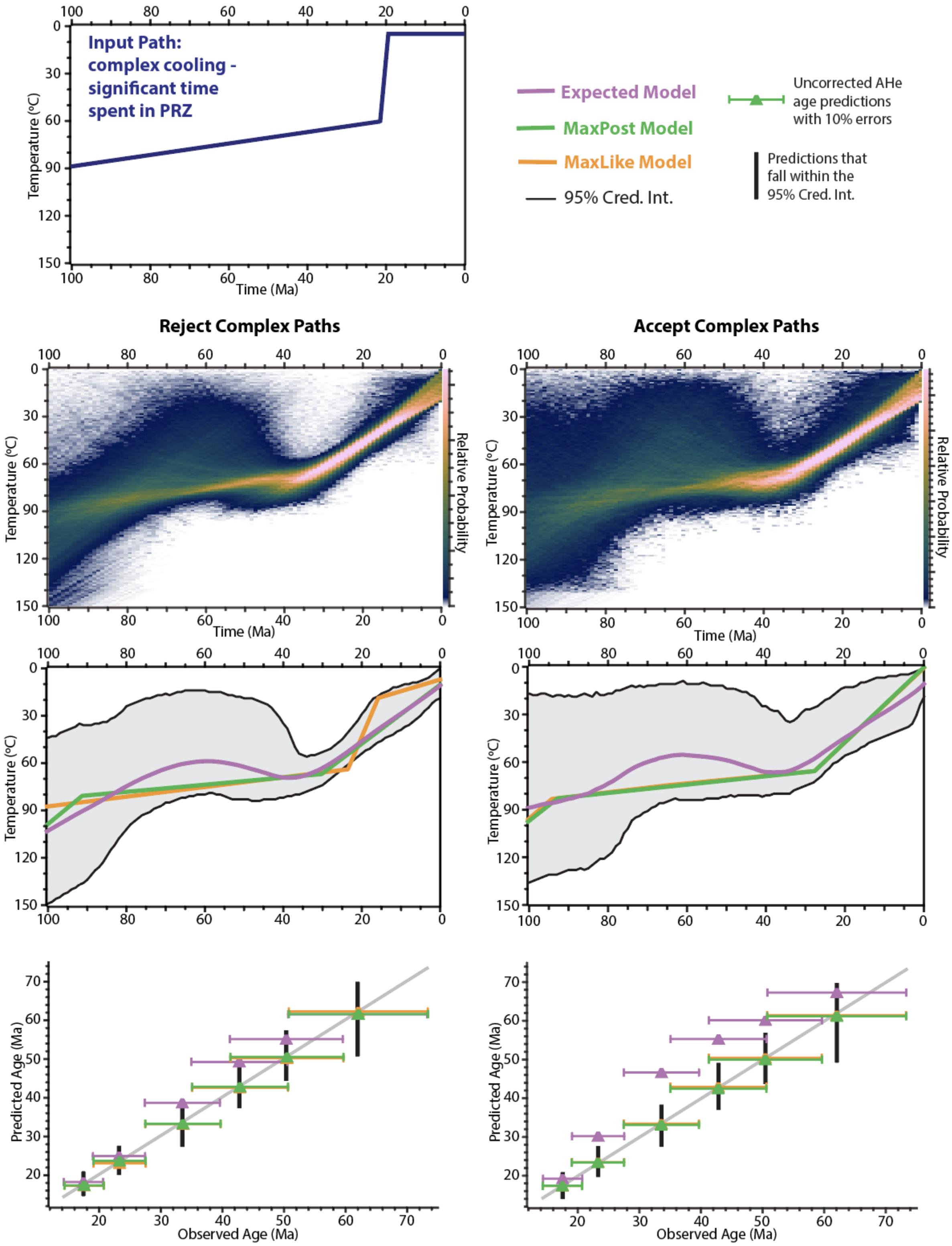
https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


**Figure 3: Inverse modeling of a simple cooling history scenario (path 1). The input forward model used to generate the synthetic data to model is a modified version of path 1 (Wolf et al., 1998) to account for radiation damage. Note: each damage model requires different modification to obtain a 40 Ma age (Stevens-Goddard et al., 2026). We compare model outputs from an inversion that rejects more complex models that do not improve the data fit (left column) and model outputs from an inversion that accepts more complex models (right column). Grey shaded region in T-t history plots encompasses the entire 95% credible interval space. Grey line in bottom plots is a 1:1 comparison of uncorrected predicted ages and uncorrected observed ages, color coded by the thermal history model. Black vertical bars represent 95% of all the predicted ages for a given sample that fall in that range.**

When we run the exact same inverse model test and select the option to accept more complex paths even if they do not improve the fit, we see much the same result as the test rejecting the more complex models. In this case, however, the range of pre-rapid cooling histories is greatly increased as seen by the wide 95% credible envelope before 40 Ma (Fig. 3). The expected model is smoothed out as before but skews even more to capture the dearth of possible pre-40 Ma temperature changes. The maximum likelihood and maximum posterior align almost perfectly with the input model. The predictions show a near perfect fit for the maximum likelihood and maximum posterior outputs, but the skewing of the expected model produces a looser fit between the observed and predicted ages (Fig. 3). As we know that the input model had no information before 40 Ma, we can easily discount the modeled pre-40 Ma history; however, we acknowledge that in nearly all real case studies, we do not know the true history, and hence the need to model the data. If the data do not constrain parts of the model, choosing to accept more complex models will add uninterpretable complexity to the cooling history but also increase the 95% credible intervals highlighting that the data do not provide much information on the model structure there.

Testing a slightly more complex cooling history, we see a similar result (Fig. 4). The option to reject more complex t-T histories again produces a range of acceptable models and the averaging to produce the expected model produces implied cooling from 100-60 Ma, reheating from 60-40 Ma, then cooling from 35-0 Ma. The maximum likelihood model does a better job of capturing the true history as seen visually, while the maximum posterior model does not capture the more rapid cooling around 20 Ma. These last two models (maximum likelihood and maximum posterior) make similar predictions that correspond well with the observations, while the expected model shows a tendency to have older predicted ages (falling above the 1:1 line) as the thermal history is at lower temperatures than the other two prior to the cooling. This again reflects smoothing inherent in the averaging process and the fact that the distribution of temperatures at a given time tends to be asymmetrical, typically skewed to lower temperatures. We see this also in the simple AFT example where the maximum temperature around 50 Ma in the expected model is lower than that in the maximum likelihood and posterior models (Fig. 1a). This is because of the non-linear dependence of diffusion (and annealing) on temperature, such that increasing temperature linearly leads to progressively faster diffusion/annealing and greater sensitivity (until the temperature required for total diffusive loss of He or annealing of fission tracks is exceeded).

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


**Figure 4: Inverse modeling of a slightly more complex cooling history (path 3). The input forward model used to generate the synthetic data is a modified version of path 3 (Wolf et al., 1998) to account for radiation damage (Stevens-Goddard et al., 2026). We compare model outputs from an inversion that rejects more complex models that do not improve the data fit (left column) and model outputs from an inversion that accepts more complex models (right column). Grey shaded region in T-t history plots encompasses the entire 95% credible interval space. Grey lines in bottom plots are a 1:1 comparison of uncorrected predicted ages and uncorrected observed ages, color coded by the thermal history model. Black vertical bars represent 95% of all the predicted ages for a given sample that fall in that range.**

When more complex paths are accepted, the 95% credible intervals widen, as in the previous example (Figs. 3 & 4). The expected model is skewed to even lower temperatures (e.g., between 80 and 40 Ma) and we see the effect of this in the predicted ages which tend to be systematically too old. The maximum likelihood model no longer reproduces the input thermal history as well and is effectively the same as the maximum posterior. This reflects some of the randomness of MCMC sampling, but irrespective, both of these models can adequately predict the observed data, yet produce a different thermal history than the true input thermal history (Fig. 4).

## 5 Choosing to accept or reject more complex models that do not improve the data fit

Here, we elaborate on the details of the QTQt inverse model algorithm associated with model complexity (i.e., number of t-T points in a thermal history). It is important to note that only QTQt versions post 5.7.0 have the option to accept or reject more complex models, which is a choice made when setting the MCMC parameters. In the examples presented here we show that predicted ages for the maximum likelihood paths do not differ with any significance whether the choice is made to accept or reject more complex models. The choice to accept more complex models has the largest effect on the parts of the thermal history that are not constrained by the data. When accepting more complex models there are a lot more solutions and variation within the modeled space; thus the expected model becomes more smoothed out from averaging effects (e.g., Figs. 3 & 4).

The examples we explore here are admittedly fairly simple thermal history scenarios and cover fairly short geologic timescales (<100 Ma). In these examples, the model outputs and data predictions are essentially the same for the model runs that accept or reject more complex models, when evaluating paths that fit the data the best or when comparing the parts of the paths that are constrained by the data and "known". Resolution is lost and data fits may be less good for the expected model that includes more complex paths (accept). In reality, we almost never know the true thermal history. So, we, as researchers, need to think critically about what geologic scenarios make sense and then perform many sensitivity tests to find models that can fit the data and also that make geologic sense. Assessment of uncertainty is critical and the two approaches– rejecting or accepting complex models can allow us to do that, albeit in a relatively simple fashion. We expect that the two populations of thermal histories will be similar where the data provide information but can differ where the data do not. We have noted also that the expected model can be unrepresentative of the individual models in the final distribution due to the smoothing involved in averaging and this will be exacerbated when the timing of heating events is not well constrained.

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


Therefore, it is important to also assess individual sampled models, such as the maximum likelihood and maximum posterior models. In practice, it is likely that the further back in time we try to extract thermal history information, the less certain the inferences will be, although this clearly depends on the nature of the data we are using. Thus, it may perhaps make sense to choose to accept more complex models when exploring longer timescales (deep-time questions) when there may be several known or unknown (but suspected) geological events that should be accounted for with a more complex t-T history. However, any features or structure in a preferred thermal history need to be evaluated to determine if they are required to explain the data or simply do not contradict the data (i.e., the predictions are not sensitive to the structure). In the latter case, geological arguments can be valuable to justify the inclusion of such a structure, but these also need to undergo robust assessment. Therefore, it is important to test model decisions that may control model outputs and data predictions to determine if the data are constraining the outputs or if it is a model parameter that controls the output (Murray et al., 2022; Abbey et al., 2023; Stevens Goddard et al., 2026). This can be addressed by targeted forward modelling, targeting features in the thermal history models, and perhaps removing or changing them and assessing the change in the data predictions for all data. Furthermore, each modeling choice that is made (e.g., what data to include, geological assumptions, model parameters) should be clearly articulated and accompanied with reasoning for why the choice was made. We suggest following reporting templates provided in Flowers et al., 2015; Murray et al., 2022; Flowers et al., 2023a; Abbey et al., 2023).

## Data availability

The data used in the models presented here is all synthetic data from running forward models in QTQt. These synthetic data are used at model inputs in the form of text files. The text files are provided in the supplemental file and all model choices are outlined in tables in the supplemental file.

## Supplement link

The link to the supplement will be included by Copernicus, if applicable.

## Author contributions

ALA and KG contributed to project design, modeling, and manuscript preparation.

## Competing interests

None.

**Acknowledgements**

We thank Kendra E. Murray, Andrea L. Stevens Goddard, Mark Wildman, Alyssa L. Abbey, and the GChron editors for creating this special issue to record and house new progress in thermal history modeling methods and also Kalin McDannell for exchanges on this subject.

**Financial support**

None.

## References

Abbey, A. L., Wildman, M., Stevens Goddard, A. L., & Murray, K. E. (2023). Thermal history modeling techniques and interpretation strategies: Applications using QTQt. *Geosphere*, *19*(2), 493-530. https://doi.org/10.1130/GES02528.1

Agostinetti, N. P., & Malinverno, A. (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, *181*(2), 858-872. https://doi.org/10.1111/j.1365-246X.2010.04530.x

Cogné, N., Gallagher, K., & Cobbold, P. R. (2011). Post-rift reactivation of the onshore margin of southeast Brazil: Evidence from apatite (U–Th)/He and fission-track data. *Earth and Planetary Science Letters*, *309*(1-2), 118-130. https://doi.org/10.1016/j.epsl.2011.06.025

Flowers, R. M., Farley, K. A., & Ketcham, R. A. (2015). A reporting protocol for thermochronologic modeling illustrated with data from the Grand Canyon. *Earth and Planetary Science Letters*, *432*, 425-435.

Flowers, R. M., Ketcham, R. A., Enkelmann, E., Gautheron, C., Reiners, P. W., Metcalf, J. R., ... & Brown, R. W. (2023a). (U-Th)/He chronology: Part 2. Considerations for evaluating, integrating, and interpreting conventional individual aliquot data. *Bulletin*, *135*(1-2), 137-161. https://doi.org/10.1016/j.epsl.2015.09.053

Flowers, R. M., Ketcham, R. A., Shuster, D. L., & Farley, K. A. (2009). Apatite (U–Th)/He thermochronometry using a radiation damage accumulation and annealing model. *Geochimica et Cosmochimica acta*, *73*(8), 2347-2365. https://doi.org/10.1016/j.gca.2009.01.015

Flowers, R. M., Zeitler, P. K., Danišík, M., Reiners, P. W., Gautheron, C., Ketcham, R. A., ... & Brown, R. W. (2023b). (U-Th)/He chronology: Part 1. Data, uncertainty, and reporting. *Bulletin*, *135*(1-2), 104-136. https://doi.org/10.1130/B36266.1

Gallagher, K. (2012). Transdimensional inverse thermal history modeling for quantitative thermochronology. *Journal of Geophysical Research: Solid Earth*, *117*(B2). https://doi.org/10.1029/2011JB008825Digital Object Identifier (DOI)

Georgieva, V., Gallagher, K., Sobczyk, A., Sobel, E. R., Schildgen, T. F., Ehlers, T. A., & Strecker, M. R. (2019). Effects of slab-window, alkaline volcanism, and glaciation on thermochronometer cooling histories, Patagonian Andes. *Earth and Planetary Science Letters*, *511*, 164-176. https://doi.org/10.1016/j.epsl.2019.01.030

https://doi.org/10.5194/egusphere-2026-2328
Preprint. Discussion started: 29 May 2026


Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711-732. https://doi.org/10.1093/biomet/82.4.711

Jasra, A., Stephens, D.A., Gallagher, K. and Holmes, C.C., (2005) Analysis of geochronological data with measurement error using Bayesian mixtures. *Mathematical Geology*, 38(3), 269-300.

Ketcham, R. A., Tremblay, M., Abbey, A., Baughman, J., Cooperdock, E., Jepson, G., Murray, K., Odlum, M., Stanley, J., & Thurston, O. (2022). Report from the 17th International Conference on Thermochronology. *Authorea Preprints*. https://doi.org/10.1002/essoar.10511082.1

Licciardi, A., Gallagher, K., Clark, S.A., (2020) A Bayesian approach for thermal history reconstruction in basin modelling. *Journal of Geophysical Research*, *125*, e2020JB019384. https://doi.org/ 10.1029/2020JB019384

MacKay, D. J. (2003). Information theory, inference and learning algorithms. *Cambridge university press*.

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431-12447. https://doi.org/10.1029/94JB03097

Murray, K. E., Stevens Goddard, A. L. , Abbey, A. L., & Wildman, M. (2022). Thermal history modeling techniques and interpretation strategies: Applications using HeFTy. *Geosphere*, *18*(5), 1622-1642. https://doi.org/10.1130/GES02500.1

Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, *167*(2), 528-542. https://doi.org/10.1111/j.1365-246X.2006.03155.x

Stevens Goddard, A. L., Murray, K. E., Abbey, A. L., & Wildman, M. (2026). Benchmark time-temperature paths provide a shared framework for evaluating and communicating thermochronologic data interpretation. *Authorea Preprints*. https://doi.org/10.31223/X5876K

Wolf, R. A., Farley, K. A., & Kass, D. M. (1998). Modeling of the temperature sensitivity of the apatite (U–Th)/He thermochronometer. *Chemical Geology*, *148*(1-2), 105-114. https://doi.org/10.1016/S0009-2541(98)00024-2