



This paper is published under the terms of the CC-BY-NC license.

© 2023 The Authors

# Thermal history modeling techniques and interpretation strategies: Applications using QTQt

Alyssa L. Abbey<sup>1</sup>, Mark Wildman<sup>2</sup>, Andrea L. Stevens Goddard<sup>3</sup>, and Kendra E. Murray<sup>4</sup>

<sup>1</sup>Department of Geological Sciences, California State University, Long Beach, 1250 Bellflower Boulevard, Long Beach, California 90840, USA

<sup>2</sup>School of Geographical and Earth Sciences, University of Glasgow, Lilybank Gardens, G12 8RZ Glasgow, UK

<sup>3</sup>Department of Earth and Atmospheric Sciences, Indiana University, 1001 E. 10<sup>th</sup> Street, Bloomington, Indiana 47408, USA

<sup>4</sup>Department of Geosciences, Idaho State University, 921 South 8th Avenue, Pocatello, Idaho 83209, USA

## ABSTRACT

**Advances in low-temperature thermochronology have made it applicable to a plethora of geoscience investigations. The development of modeling programs (e.g., QTQt and HeFTy) that extract thermal histories from thermochronologic data has facilitated growth of this field. However, the increasingly wide range of scientists who apply these tools requires an accessible entry point to thermal history modeling and how these models develop our understanding of complex geological processes. This contribution offers a discussion of modeling strategies, using QTQt, including making decisions about model design, data input, kinetic parameters, and other factors that may influence the model output. We present a suite of synthetic data sets derived from known thermal histories with accompanying tutorial exercises in the Supplemental Material<sup>1</sup>. These data sets illustrate the opportunities and limitations of thermal history modeling. Examining these synthetic data helps to develop intuition about which thermochronometric data are most sensitive to different thermal events and to what extent user decisions on data handling and model set-up can control the recovery of the true solution. We also use real data to demonstrate the importance of incorporating sensitivity testing into thermal history modeling and suggest several best practices for exploring model sensitivity to factors including, but not limited to, the model design or inversion algorithm, geologic constraints, data trends, the spatial relationship between samples, or the choice of kinetics model. Finally, we provide a detailed and explicit workflow and an applied example for a method of interrogating vague model results or low observation-prediction fits that we call the “Path Structure Approach.” Our explicit examination of thermal history modeling practices is designed to guide modelers to identify the factors controlling model results and demonstrate reproducible approaches for the interpretation of thermal histories.**

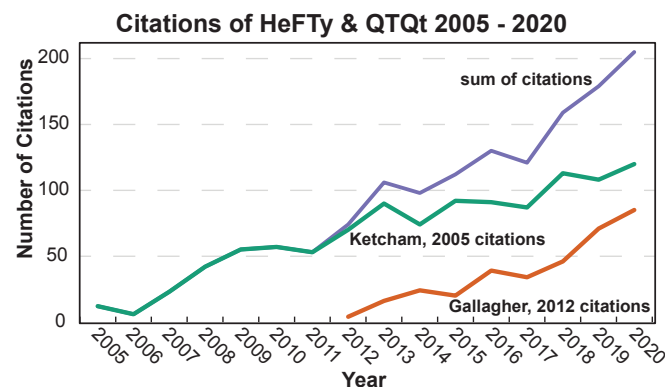
<sup>1</sup>Supplemental Material. Contains tables that detail all model input parameters and decisions, figures that show model predictions from sensitivity tests shown in the main text of the manuscript, and a tutorial accompanied by data and model input files, for becoming familiar with QTQt. Please visit <https://doi.org/10.1130/GEOS.S.21538512> to access the supplemental material, and contact [editing@geosociety.org](mailto:editing@geosociety.org) with any questions.

Alyssa Langford Abbey <https://orcid.org/0000-0002-0617-4859>

## INTRODUCTION

Low-temperature thermochronology is a popular approach for constraining the timing of upper-crustal events and the rates of geological processes. Thermochronologic techniques exploit the temperature-sensitive retention of radiogenic decay products in different minerals. Although an “age” is typically obtained, the geological significance of that age depends on the geologic context and is a product of the length of time the mineral has spent at temperatures where partial retention (or partial loss) of the daughter product occurs. In simple rock-cooling scenarios, the concept of a “closure temperature” is useful for describing the transition between open- and closed-system behavior and the relative temperature sensitivities of thermochronometers (Dodson, 1973). In this concept, daughter products are not retained in the mineral at temperatures hotter than the closure temperature. As a rock cools, daughter products begin to accumulate as the system closes. The rate of daughter retention increases as temperature decreases until it is sufficiently cold enough that daughter products are completely retained. Importantly, a thermochronometer’s closure temperature is not constant, because it depends on the cooling rate (Dodson, 1973), grain composition (e.g., O’Sullivan and Parrish, 1995), and size (e.g., Reiners and Farley, 2001), and even mineral characteristics that are themselves a function of a rock’s thermal history, such as accumulated radiation damage (Shuster et al., 2006; Ault et al., 2019; Flowers et al., 2022a). Thus, forward and inverse thermal history modeling has become an invaluable tool for interpreting cooling ages. Several different numerical modeling approaches have been developed to derive thermal histories from thermochronologic data (e.g., Gallagher, 1995, 2012; Ketcham et al., 2000; Braun, 2003; Zeitler, 2004; Ketcham, 2005; Hager and Stockli, 2009).

Two of the most commonly used thermal history modeling software packages are QTQt and HeFTy. Although based on different optimization and statistical approaches, they typically yield consistent results (see Ketcham, 2005, and Gallagher, 2012, for full details on the methodology of these approaches). The increasing popularity of these programs (Fig. 1) can in part be attributed to their convenient user interfaces and the availability of user support from the developers and peers in the thermochronology community. However, thermal history model solutions are inherently non-unique; so careful planning



**Figure 1.** Graphic reproduced from Murray et al. (2022) showing the trends in citations of HeFTy (Ketcham, 2005, green line) and QTQt (Gallagher, 2012, orange line) thermal history modeling applications from 2005 to 2020. Purple line shows a trend in the sum of citations of both packages. Data sourced from Google Scholar (October 2021).

and decision making while defining model inputs and choosing parameters are needed in order to obtain useful and interpretable results. With increased popularity, there has been increased scrutiny over the use of these programs to make geological interpretations (Vermeesch and Tian, 2014, 2018, 2019; Gallagher and Ketcham, 2018, 2020; Fox and Carter, 2020).

This contribution discusses modeling styles, strategies, and best practices with a goal of guiding users through modeling and interpreting thermochronometric data using QTQt—a discussion that is otherwise relegated to occasional workshops, informal mentorship, or the review process. We do not intend to provide a direct critique or comparison of modeling programs or outline program pitfalls and limitations, which are discussed elsewhere. Instead, we highlight the importance of user decision making when designing and performing thermal history modeling with QTQt (Fig. 2). By presenting sensitivity tests and case studies, we aim to raise awareness of how user decisions affect the model output, outline strategies for novice and experienced users to familiarize themselves with the capabilities of QTQt, and encourage transparent reporting of the decision-making process. The exercises presented here are designed specifically for QTQt. Our companion paper (Murray et al., 2022) aims to achieve a similar goal for the HeFTy modeling software. However, the approaches discussed in both contributions are relevant and important to consider no matter which thermal history modeling program is being used.

Our intention is to help users develop an intuitive understanding of modeling methods and results by providing examples of practical strategies for modeling different types of thermochronologic data sets, and our examples here utilize the apatite fission-track (AFT) and apatite (U-Th-Sm)/He (AHe) thermochronometers. The AFT system relies on the accumulation of damage trails (“fission-tracks”) in the crystal lattice produced during the spontaneous fission

of  $^{238}\text{U}$  and the partial annealing of these tracks on geologic timescales at temperatures of  $60\text{--}110 \pm 10$  °C (partial annealing zone [PAZ]). At temperatures  $>110 \pm 10$  °C, tracks are completely annealed, while at temperatures  $<60$  °C, annealing is assumed to be negligible (Gleadow et al., 1986). Variations in the chemical composition of an apatite grain (primarily due to anion substitution of Cl-F-OH) has a second-order influence on the annealing rate of tracks (Carlson et al., 1999), while anisotropic annealing (e.g., Donelick et al., 1999) and differences in chemical etching protocols (Tamer et al., 2019) may cause variability in fission-track length data across analysts. The AHe system exploits the accumulation of radiogenic  $^4\text{He}$  produced during the alpha decay of U, Th, and Sm. Similar to the AFT PAZ, there exists a temperature range (the partial retention zone [PRZ]) where partial diffusive loss of He occurs by thermally activated diffusion. The temperature range for the AHe PRZ on geologic timescales is  $30\text{--}90$  °C. However, this is dependent on variables such as grain size, rock cooling rate, and amount of radiation damage accumulated in a crystal (approximated by the effective uranium concentration,  $[\text{eU}] = [\text{U}] + 0.235[\text{Th}] + 0.0047[\text{Sm}]$ ) (Shuster et al., 2006; Ault et al., 2019). In routine He thermochronology analyses, multiple single grains are dated from each sample, and in many cases, He ages from single samples are highly variable (also referred to as age “dispersion” or “irreproducibility,” e.g., Brown et al., 2013; McDannell et al., 2018; Ault et al., 2019; Flowers et al., 2022a). Intra-sample He age variability can arise from various factors, which sometimes coexist in real data sets (e.g., Flowers and Kelley, 2011; Murray et al., 2019); see Flowers et al. (2022a) for a recent overview. Here, we quantify a sample’s He age variability as the standard deviation of the single-grain ages divided by the mean of single-grain ages.

First, we illustrate the impact of data input and model design on thermal history model results using six synthetic data sets (see companion paper Murray et al., 2022, for similar analyses in HeFTy). We use forward and inverse modeling to demonstrate the effects of various decisions made during the modeling process. First, we use QTQt to generate synthetic data using a forward thermal history model, and then we use those synthetic data as input data for inverse models, which we run to assess how well we can retrieve the “true” time-temperature (t-T) paths that generated the synthetic data. This approach requires six steps: (1) build a generic input file to define the synthetic data we wish to generate; (2) define t-T paths to forward models; (3) run forward models to predict and generate synthetic data; (4) design inverse models, modifying specific input parameters; (5) run inverse models to obtain thermal histories and data predictions; and (6) interpret inverse model outputs. We provide a tutorial in the Supplemental Material (see footnote 1) to guide readers through these exercises for themselves. By working through these exercises with near “perfect” data (the He data can be effectively “perfect,” whereas the FT data will always have some sampling-related noise) and known “true” histories, new and experienced users alike can build intuition about the importance of different decisions during the modeling process (Fig. 2).

Next, we present examples from published data sets and discuss decision-making strategies for (1) collecting and modeling a vertical profile of data; (2) modeling multiple thermochronometers together; and (3) modeling data

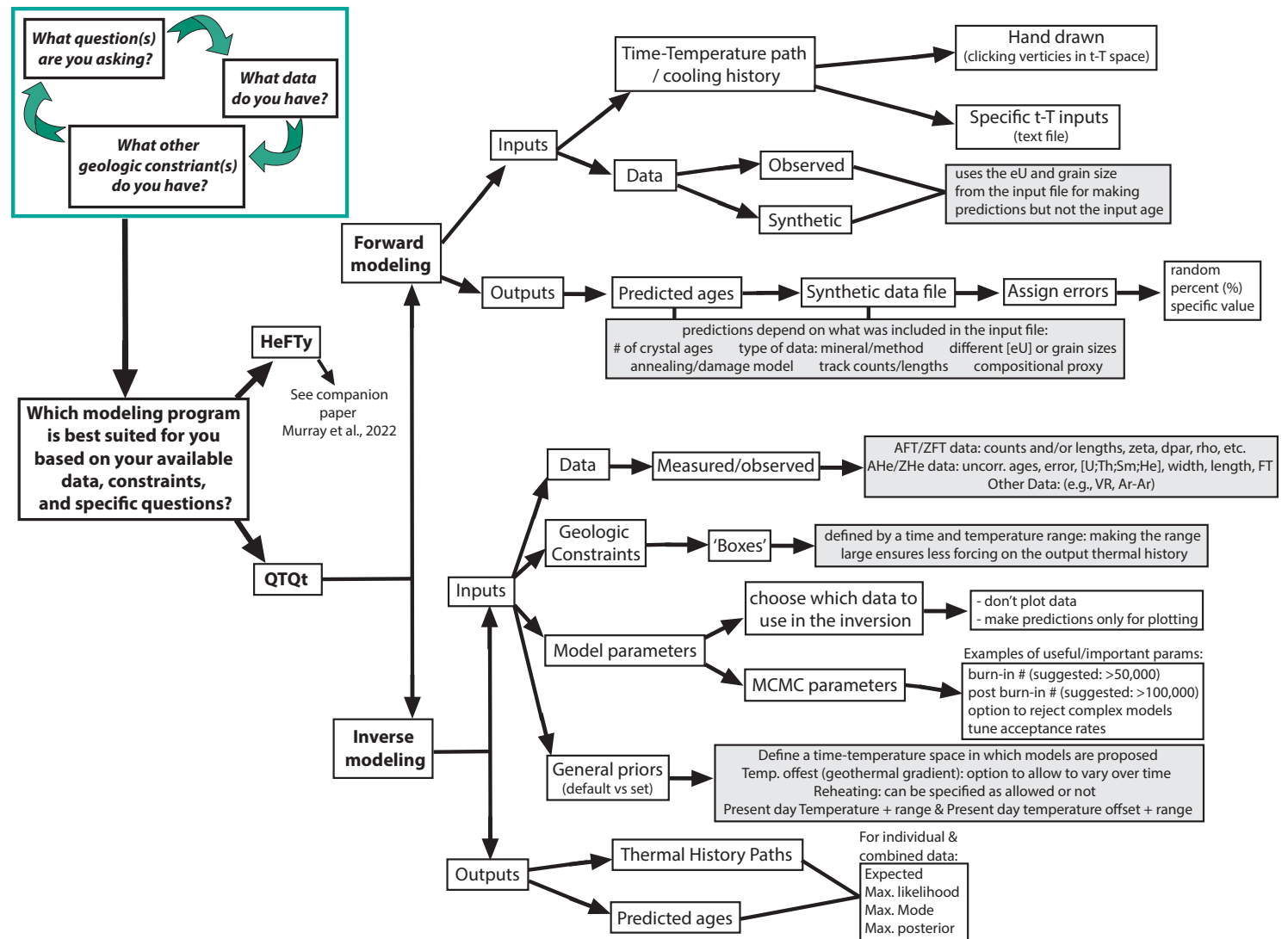


Figure 2. Flow chart showing the numerous decisions users are required to make during the process of modeling thermochronometric data. MCMC—Markov-Chain Monte Carlo algorithm.

sets that include complex AHe and AFT data. We also present an iterative “Path Structure Approach” that elaborates on how to use each of the QTQt outputs to evaluate a variety of possible thermal histories produced by one set of data. These examples are not intended to expand on or revise interpretations made in the initial publications; instead, they illustrate opportunities and limitations of modeling with real data by addressing the practical challenges that face thermochronologists. This explicit discussion of modeling and interpretation strategies offers a resource and framework to help anticipate and evaluate the impact of sampling, data collection, and model design on thermal history modeling.

## ■ BACKGROUND INFORMATION

The following section provides some background information on the methodology, philosophy, and terminology underpinning QTQt thermal history modeling. However, we do not provide a detailed description of the mathematical and statistical underpinnings of the program. Instead, we direct readers to Gallagher et al. (2009) and Gallagher (2012) and references therein for details on the transdimensional Markov-Chain Monte Carlo (MCMC) algorithm and Bayesian approach to thermal history modeling. Furthermore, we refer readers to Vermeesch and Tian (2014, 2018) and Gallagher and Ketcham (2018, 2020) where these concepts (and those applicable to the HeFTy program) are reiterated and thoroughly discussed.

At the heart of QTQt is Bayes theorem:

$$p(m|d) \propto p(d|m) \cdot p(m),$$

which states that the probability of obtaining a set of model parameters given the data ( $p(m|d)$ ) is proportional to the probability of obtaining the data given the underlying model parameters ( $p(d|m)$ ) multiplied by the prior probability of the model parameters ( $p(m)$ ). The term  $p(m|d)$  is referred to as the “posterior” probability and is a product of multiplying the “likelihood” function  $p(d|m)$ , which describes how well the model fits the data, by the “prior,” which is a measure of what we know (or think we know) before we consider the data. The prior, likelihood, and posterior each take the form of probability distribution (Gallagher et al., 2009). In short, our prior knowledge of the model is updated by the likelihood of fitting the data given the model parameters and transformed into the posterior, which is what we have learned about our model from the data (Gallagher, 2012). In QTQt, the inverse modeling function constructs the posterior probability distributions for all model parameters conditional on the data provided (Gallagher, 2012).

In the case of thermal history modeling, the “prior” information comprises all of the knowledge about the samples that is not the measured data itself, including (1) a broad constraint on time and temperature in which the thermal history will be sampled; (2) additional geological and thermal history constraints; (3) if multiple samples are being modeled together, the temperature

difference (i.e., offset) between samples in a profile that arises from their vertical separation, and, if desired, a geothermal gradient that varies over time; (4) allowance for reheating or not; and (5) kinetics parameters related to He diffusion and/or fission or damage annealing (Fig. 2). QTQt employs a reversible jump, transdimensional MCMC algorithm to sample the prior probability distributions of the model parameters and propose a t-T path.

Importantly, the transdimensional MCMC algorithm in QTQt permits the number of model parameters to be treated as an unknown, and therefore the data determine the complexity of the t-T paths that QTQt proposes. The simplest thermal history consists of two points in time-temperature space (t-T points) with a constant cooling rate between them. Increasingly complex thermal histories have more and more t-T points, which mark a potential change in the cooling/heating rate (i.e., an inflection point in the t-T path). Given a proposed t-T path, QTQt calculates a likelihood probability (how well the thermal history fits the data) and a posterior probability (a product of the likelihood and the prior). The MCMC algorithm then continues on a random walk of the model space by proposing a new t-T path that is conditional on the previous one (i.e., a “learning” search algorithm) by: moving a point in time or temperature, changing the value of temperature offset for profile samples, adding (“birth”) or removing (“death”) a time-temperature point, sampling a kinetic parameter for an individual sample, or resampling the observed data value (Gallagher, 2012). Proposed t-T paths are either accepted or rejected using an acceptance criterion outlined in Gallagher (2012). A natural consequence of the Bayesian transdimensional MCMC approach used by QTQt is that simpler t-T paths (i.e., fewer t-T points) are preferred to complex t-T paths (i.e., more t-T points) if both models fit the data equally well (Gallagher and Ketcham, 2018). QTQt’s inherent preference for the simplest t-T paths that fit the data is perhaps the most important difference between QTQt and HeFTy model results. QTQt’s MCMC algorithm is run for several thousands to millions of iterations to sample the model space and produce a collection of acceptable thermal history models (Gallagher, 2012).

## ■ GENERAL RULES FOR ENGAGEMENT WITH QTQt

After completing the hard work of data acquisition and reduction, the first question a thermal history modeler may ask is: Which thermal history modeling software should I use? There are multiple numerical approaches adopted by different research groups, but as noted above, for many HeFTy and QTQt are the most accessible. The user should explore the functionality of QTQt and HeFTy and determine whether one or the other best suits their needs. QTQt has many strengths, including, but not limited to, the ability to model samples in a vertical and/or borehole profile (as a forward or inverse model) by defining a thermal offset between samples, to incorporate a large number of single grain U-Th/He ages with control over diffusion parameters, and to resample observed data and uncertainties such that outlier or enigmatic data do not have a large influence over the solution.

Ongoing discussions comparing the QTQt and HeFTy programs may lead some to conclude that there is a rigid dichotomy between these two tools, but this is not the case. For example, Vermeesch and Tian (2014, 2018) concluded that for a complex data set, HeFTy may “break” and fail to yield any acceptable solutions, whereas QTQt will always yield a solution—one that can appear to be well resolved but actually not fit the data well. In response to these assertions, Gallagher and Ketcham (2018, 2020) highlighted that despite substantial differences in their theoretical approaches, both modeling tools, when used with careful handling of input data and robust understanding of software functionality, will yield comparable results. When a user is faced with a HeFTy model that fails to yield any good or acceptable fits to the observed data or a QTQt model that yields a well-resolved thermal history that does not reproduce the data, in both cases this is an indication that the model design, assumptions, and input data should be re-assessed. Importantly, in such cases, the user has still learned something useful through the modeling process. Thus, it is crucial for users to always compare the model predictions with the measured data. Users are able to extract model predictions and employ statistical techniques to quantify the degree of data fit to the observations; however, QTQt provides several graphical outputs to provide an accessible representation of how well the observed data are predicted by the model (Gallagher, 2016). This manuscript is designed to aid modelers in developing a path forward in these situations in QTQt. In concert with Murray et al. (2022), here we reiterate the conclusions of Gallagher and Ketcham (2018, 2020) and describe specific functionalities and modeling strategies for QTQt, so that modelers are equipped to choose the best approach for their own data sets.

The next decision in the modeling process is whether to use a “forward” or “inverse” modeling approach, or an iterative combination of the two (for example, our “Path Structure Approach”). Forward thermal history modeling generates predicted thermochronometric data for manually defined t-T paths. Predictions from forward models can be compared to observed data, used to support or reject a particular hypothesis, or be used to develop a future sampling strategy. For example, forward modeling may direct a researcher to use a thermochronometric system that is sensitive to a particular magnitude and rate of cooling or demonstrate that a vertical sampling campaign will be effective for resolving a hypothetical thermal history. Forward modeling exercises will produce identical results regardless of the program used; however, we provide examples of the forward modeling workflow specific to QTQt in the next section on evaluating thermochronometric behaviors using forward models and synthetic data and in the tutorial in the Supplemental Material (see footnote 1) (Fig. 2).

Here we outline the decision-making process one may be faced with when embarking on a geological investigation involving thermal history modeling. The first step is an assessment of the following questions: (1) What geologic question(s) are you asking?; (2) what data do you have?; and (3) what other geologic constraint(s) do you have? (Fig. 2). This may be clear from the outset; for example: (1) When was a sedimentary basin formed, and what was its post-burial exhumation history; (2) high-quality, state-of-the-art AFT and AHe

data from a borehole; (3) tight stratigraphic constraints for the sedimentary units sampled. However, it may be necessary to expand your initial inquiry to address complex related questions such as: Are there existing competing hypotheses to be tested? What other data might you need to address the scientific question (e.g., additional thermochronometers)? Are there quality control concerns on your data (e.g., legacy AFT data lacking compositional proxies or standardized etching techniques, “over-dispersed” AHe ages)? What is your confidence in each geologic constraint? All of these considerations inform the prior knowledge as we’ve described, and, in different ways, can be built into the modeling process as we will illustrate.

## ■ EVALUATING THERMOCHRONOMETRIC BEHAVIORS USING FORWARD MODELS AND SYNTHETIC DATA

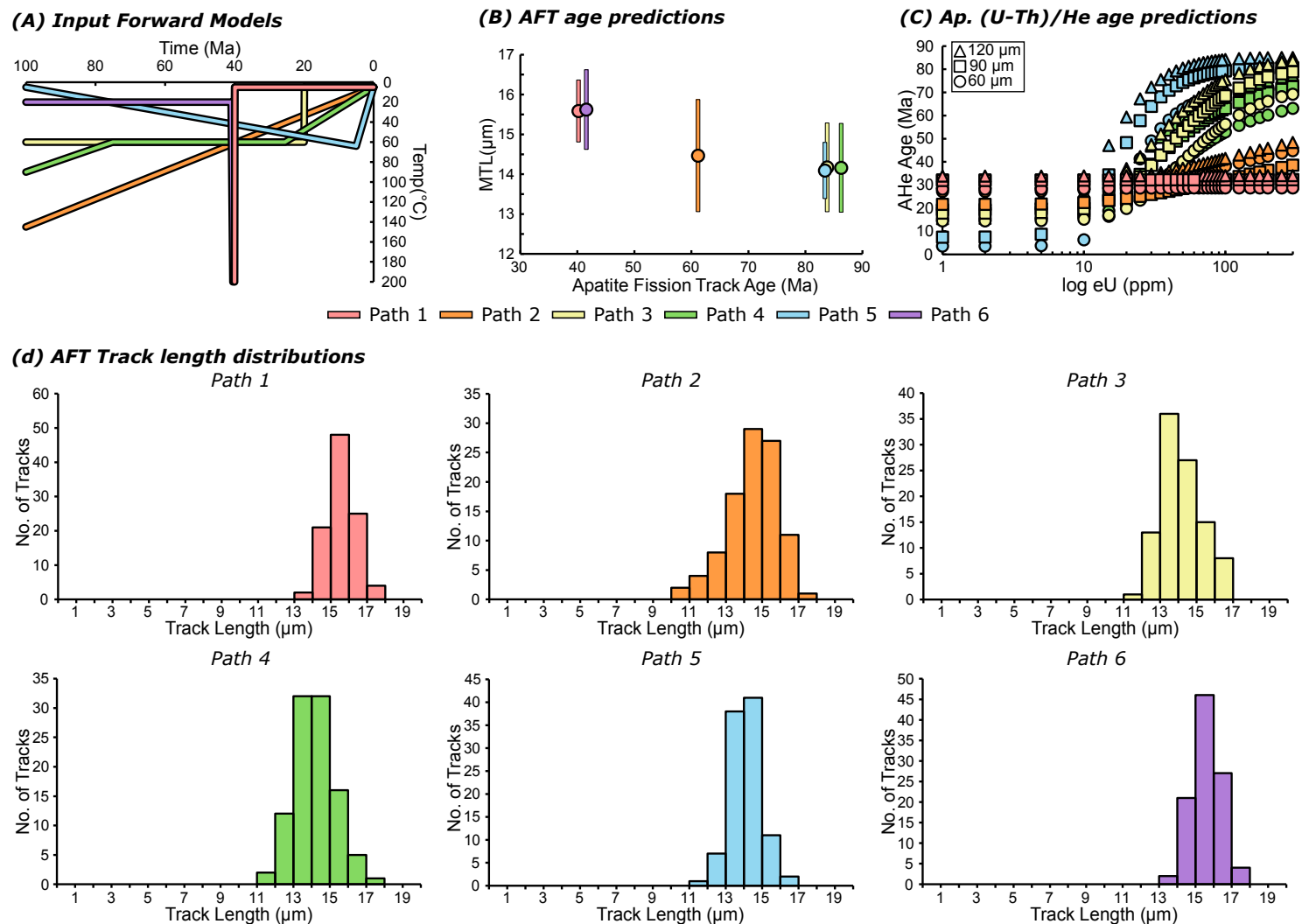
We illustrate the utility of forward thermal history modeling by using six t-T paths, representative of common geologic settings, to create synthetic AFT and AHe data sets. These thermal histories are modified from histories first used by Wolf et al. (1998) to highlight the non-unique nature of AHe ages. These paths are specifically designed to explore the temperature ranges over which the AHe system is most sensitive (30–90 °C) and investigate the consequences of partial-retention behavior in that system. However, we also interrogate predicted AFT behavior and recognize that alternative thermal histories could easily be designed to illustrate the non-unique nature of different thermochronologic data sets.

The original demonstration, from Wolf et al. (1998), presents five t-T paths that all predict a 40 Ma AHe age, assuming a constant grain size (60  $\mu\text{m}$ ) and the diffusion kinetics of Wolf et al. (1996). Here, we use those same t-T paths from Wolf et al. (1998), plus one additional Path 6, but apply the Flowers et al. (2009) radiation damage accumulation and annealing model (RDAAM) during forward and inverse modeling, and as such, we do not obtain a 40 Ma age for a 60  $\mu\text{m}$  grain for all six thermal histories (as the AHe age will be different depending on the eU concentration of the grain). Murray et al. (2022) present minor adjustments to the t-T paths such that they each produce a 40 Ma age for a Durango-like 60 ppm [eU] apatite with a grain size of 60  $\mu\text{m}$  using the RDAAM (see table 1 in Murray et al., 2022).

### Forward Modeling Data Predictions

The AFT and AHe data predicted by our six thermal histories are shown in Figure 3 and summarized in Table S1 (footnote 1). Because each history starts at 100 Ma, the apatite grains effectively “crystallized” at that time and have no prior history. Predictions of AFT data are made assuming a Dpar of 2.05  $\mu\text{m}$ . We predicted AHe ages for a range of eU concentrations (1–300 ppm) for grain radii of 60, 90, and 120  $\mu\text{m}$  and present them without an alpha-ejection correction because uncorrected (measured) AHe ages are used as the input





**Figure 3.** (A) Six thermal history paths used to generate synthetic apatite fission-track and apatite (U-Th-Sm)/He (AFT and AHe) data. Input thermal history paths are constructed by specifying and connecting a number of points in time-temperature space. (B) Predicted AFT ages plotted against predicted mean track length for each model. (C) Predicted (uncorrected) AHe ages against [eU] over the range of 1–300 ppm. Different symbols are used to indicate different apatite grain sizes (circle—60 μm; square—90 μm; and triangle—120 μm) for each model. (D) AFT track-length distributions for each model. Specific details on these inputs and outputs are described in Table S1 (see text footnote 1).

ages in QTQt (see Table S1 for corrected ages) and the alpha ejection calculation is performed during the thermal history run.

We observe that for Path 1 and Path 6, the AFT age is ca. 40 Ma, which corresponds to the timing of rapid cooling through the PAZ. Paths 2, 3, 4, and 5 all predict older AFT ages (59–86 Ma), which do not obviously relate to a discrete thermal “event” (e.g., change in rate of heating or cooling) in the corresponding history. This is because these samples have spent a significant portion of their thermal history in the PAZ, where variable amounts of track annealing have occurred (Fig. 3). The predictions for Paths 3, 4, and 5 highlight the challenges with interpreting AFT data when events in the thermal history occur at temperatures close to the AFT thermochronometer’s lower limit of temperature sensitivity (i.e., 50–60 °C). We observe that the AFT age and MTL for the Path 3, 4, and 5 models are almost indistinguishable even without the additional uncertainty typical of “real-world” data. Although there are some subtle differences in the track length distribution (TLD) (e.g., slightly left-skewed TLD for Path 3, and a slightly narrower TLD for Path 5), resolving these features in a real data set would be challenging.

In contrast, the predicted AHe ages produced by these six thermal histories are more variable (Fig. 3C), as expected given the temperatures experienced by these grains and the radiation damage kinetics model we chose. The AHe ages for Paths 1 and 6 are approximately equal to the timing of rapid cooling, regardless of the grain size or eU, due to the limited time spent in the PRZ. Slowing the cooling rate and increasing the amount of time the grains spend in the PRZ, as shown by Path 2, creates more variability among the single-grain ages, primarily controlled by eU, but still yields a comparable mean AHe age to Path 1. Paths 3, 4 and 5 spend even more time at PRZ conditions and predict older mean AHe ages and, importantly, significantly greater age variability among the single-grain ages than Paths 1, 2, and 6 (Fig. 3C). The age-eU trends predicted from Paths 3 and 4 are very similar, with subtle differences due to differences in cooling rate at 100–80 Ma and 20–0 Ma. The AHe data predicted by Path 5 are also highly variable. The moderate and low eU grains are young because the radiation damage model predicts that they lose He during the reheating to 60 °C. In contrast, the high-eU grains (>100 ppm) are more damaged than the low-eU grains that experienced this same history, and thus retain more He and have older ages. These older grains record the initial rapid cooling through the PRZ at 100 Ma and are not sensitive to the later reheating. At low to moderate eU values (10–50 ppm), the AHe age variability controlled by grain size is also more significant for Path 5 than for Paths 3 and 4.

### Summary: Effect of Thermal History, Radiation Damage, and Grain Size on Predicted Apatite Thermochronometry Data

The forward model exercises detailed above highlight the well-established effects that radiation damage and grain size have on AHe ages and show that thermal events occurring at ~60 °C will result in only minor differences in the predicted AFT data. Simple cooling histories, such as fast cooling through the

PRZ and PAZ (Paths 1 and 6), produce nearly identical AHe age trends regardless of grain size or eU concentration, and the AFT TLD is symmetric with long MTLs (>15 μm). The lack of AHe age variability over a large range of grain size and eU as well as the AFT track-length data in this case are useful qualitative indicators of the relatively simple scenario of rapid cooling. More complex cooling histories, such as those with protracted cooling (Path 2), moderate reheating (Path 5), or extended time spent at temperatures where there is partial loss of the daughter product (Paths 3, 4, and 5) produce more variable AHe ages, skewed or broad TLDs, and shorter MTLs. Our examples reiterate how the degree to which AHe age variability is influenced by the eU and grain size (e.g., Reiners and Farley, 2001; Flowers et al., 2009; Gautheron et al., 2009). They demonstrate that eU is typically the dominant control of AHe age variability, and its effects are most pronounced for samples that experienced long durations in the PRZ. However, at very high and very low eU, He ages are more consistent for a particular grain size, and grain size variations are a bigger control of age variability. In “real-world” cases, additional age variability can be introduced from other sources such as intracrystalline zonation, He implantation, inclusions, and grain fragmentation (see Wildman et al., 2016, and Flowers et al., 2022a, for summaries). In our idealized case, assuming the Flowers et al. (2009) RDAAM model is a suitable numerical model to simulate diffusion in radiation-damaged apatite, strong correlations exist between the AHe age and the damage for which eU is a proxy, making interpretations based on age-eU trends and extracting thermal history information feasible (Flowers and Kelley, 2011). However, radiation damage in apatite and models simulating the accumulation and annealing of defects, and the impact these defects have on He diffusion, remain under investigation (e.g., Gautheron et al., 2009, 2013; Djimbi et al., 2015; Gerin et al., 2017; Willett et al., 2017).

### ■ USING SYNTHETIC DATA TO EXEMPLIFY SENSITIVITY TESTING WITH INVERSE MODELS

Our examples above show that a single t-T path proposed as a forward model will produce an AFT and AHe data set specific to that path. However, the reverse is not true; it is not possible to recover one unique t-T path from an AFT and/or AHe data set. Multiple histories may produce the same data, especially “real-world” data sets with uncertainties. Inverse modeling offers an approach for addressing the significance of this non-uniqueness, because it is a process where thousands to millions of t-T paths are sampled, and their fit to the data assessed with the aim of finding those paths that best fit, or at least are consistent with, the observed data.

The inverse modeling approach in QTQt utilizes the transdimensional MCMC method to search the model space and infer acceptable values for the model parameters from the data. In practice, QTQt does this by proposing a large number of thermal history paths, assessing their fit to the observed data, and creating an ensemble of acceptable paths. The user creates a data file input containing the observed data and a specification of the kinetics models

used to simulate annealing and/or diffusion (Fig. 2). The user then defines their prior information: a broad time and temperature box, the present-day temperature, any additional geological information in the form of constraint boxes, and the temperature offset between samples if modeling a vertical profile (Fig. 2). Note that QTQt, unlike HeFTy, can only accommodate up to five geologic constraint boxes, which reflects the substantial philosophical difference between these programs (Ketcham, 2005; Gallagher and Ketcham, 2018, 2020). In QTQt, the thermochronologic data, as input by the user, define the complexity of the thermal history model result, and therefore the user is encouraged to only use a few well-defined geologic constraints.

After designing an inverse model, users are advised to perform several short modeling runs (tens of thousands of iterations) to assess the performance of the MCMC algorithm and to monitor whether it has converged. One important evaluation is of the acceptance rates for the specified proposal time and temperature moves, which is the scale of the distribution for each model parameter from which a point can be sampled in each iteration. The proposal moves should be set such that the acceptance rate for that parameter is ~20%–60%, and the rate of new t-T points being added (“birth”) is similar to the number being removed (“death”). Increasing or decreasing the value of the proposal move will tend to have an inverse effect on the acceptance rate. Additionally, the user should assess the stability of the log likelihood (i.e., data fit) function, posterior probability log, and the number of time-temperature points as a function of iteration (e.g., likelihood chains and acceptance rates; see tutorial in Supplemental Material for examples). There should be no trends in the chain with values changing for most iterations (Gallagher, 2012; see tutorial, footnote 1). Once the chain is stable, longer runs can then be completed to obtain the ensemble of accepted thermal histories that construct the posterior distribution. We encourage users to report in publications the proposal moves and acceptance rates for the MCMC parameters and that the MCMC has converged (Tables S2–S7). We demonstrate how users can practically engage with the inverse modeling set-up and this MCMC tuning, and how engagement with the inverse modeling set-up can influence the output models, in several examples below and in the tutorial.

Once all of the above decisions are made for the model inputs, the inversion can be run. Outputs from an inversion include predicted data, thermal history paths, and the MCMC parameters including time and temperature acceptance rates and proposal “birth” and “death” rates. The predicted data include uncorrected He ages, which can be displayed in a plot against the uncorrected input measured ages such that a perfect fit of the data would fall on a 1:1 line. Predictions can also be plotted in elevation space, which is useful when modeling samples collected in a vertical profile. Predicted FT lengths and track-length distributions are plotted as a histogram.

After completing an inverse model, QTQt generates a series of outputs that convey the performance of the MCMC sampling of model parameters (see tutorial, footnote 1), thermal history models, and data predictions. Specific thermal histories that can be examined and are discussed here include the maximum likelihood, maximum posterior, and expected thermal history. The

maximum likelihood (MaxLike) model is the t-T path that fits the data the best. However, the MaxLike model often fits the data using high degrees of complexity (i.e., large number of t-T points), which may be geologically unrealistic. In contrast, the maximum posterior (MaxPost) model is essentially the simplest t-T history that fits the data. It has the maximum posterior probability, where the posterior probability is proportional to the likelihood multiplied by the prior. As a result, even if data were truly produced by a complex thermal history, if the data are not suitably informative, or the underlying thermochronometric kinetics models are not correct, the MaxPost model may not resolve the complexity of the true thermal history. The expected thermal history (ExTH) path is a weighted-mean model where the weighting is provided by the posterior probability for each model. This ensemble reconstructs the posterior probability distribution of the accepted thermal histories and is presented along with 95% credible intervals (Gallagher, 2012). Finally, a color map showing the “marginal posterior distribution” can also be visualized that shows the relative probability of a sample cooling through a box of size 1 °C and 1 m.y., or a time interval defined by dividing the model time range by 100; the marginal posterior distribution is conditional on the temperature of the thermal histories at all other times (Gallagher and Ketcham, 2018).

If the MaxLike, MaxPost, and ExTH outputs from a model run are similar, and all reproduce the data well, then the thermal history can be considered well constrained. The ExTH should reflect the true solution where it is well resolved with tight credible intervals. However, because the ExTH is a weighted-mean model, parts of the ExTH path could be the product of individual paths with very different behaviors (see our “Path Structure Approach” and the “Path Family Approach” from Murray et al., 2022). For example, there may be thermal histories that predict reheating over a given interval and others that predict cooling over the same interval, and therefore the ExTH suggests a scenario in between. In this case, the MaxLike and MaxPost models may be very different but still predict the data very well, while the averaging of the ExTH produces another different thermal history that does not fit the data as well. For all thermal histories, it is important to present and assess the data fit, either visually or with appropriate statistical tests and then evaluate the significance of the fit or lack thereof.

In many studies, there is a desire to arrive at a single thermal history solution, and the ExTH is commonly the appropriate choice because it reflects an objective sampling of all model parameters, penalizes complexity that is not resolved by the data, and can provide information on the uncertainty of the model in the form of 95% credible intervals. However, given the nature of the ExTH, the other model outputs need to be examined closely, and, in many cases, additional inverse or forward modeling is required to identify and understand the “preferred” solution. We demonstrate a strategy for evaluating model outputs and selecting a “preferred” solution, the “Path Structure Approach,” through a worked example at the end of this contribution.

In this work, most of our interpretations rely on the ExTH output, although we sometimes refer to the MaxLike or MaxPost to investigate what might be controlling the structure of the ExTH. With all of the user decisions made



when interacting and engaging in the modeling process, it is critical to record and report the decisions made and the reasoning behind those decisions. We encourage users to refer to the previously published templates and the versions we provide here for clear reporting (Flowers et al., 2015; Wildman et al., 2019b; Tables S3–S7 of this contribution, footnote 1).

One approach for assessing how model design choices impact an inverse model result is to generate synthetic data using a forward model and then invert that data to investigate how effectively the original parameters can be recovered. Next, we illustrate how both the non-unique nature of cooling ages and the choices made when designing inverse models control our ability to recover the six “true” t-T histories modeled in a forward sense in the previous section.

### Illustrating the Non-Unique Nature of Single-Grain He Data

An inversion of a single-apatite He age illustrates how QTQt handles the non-unique nature of thermochronometric ages. We use a single apatite grain with a 60  $\mu\text{m}$  radius, an eU concentration of 60 ppm, and a corrected He age of 40 Ma, which Wolf et al. (1998) demonstrated is produced by a wide range of thermal histories, including Paths 1–6 here (Fig. 4). The resultant color map (Fig. 4A) shows the possibility that any of the six cooling paths, and any number of other cooling paths, could produce the measured age for that specific grain. However, given the single data input, QTQt converged on the simplest, Path-2-like solution (Fig. 4). Indeed, a casual assessment of the color map (Fig. 4A) might lead one to think that a Path-2-like thermal history offers the best fit to a single 40 Ma age, and thus QTQt has constrained a specific thermal history from this single AHe age and excluded the possibility of more complex solutions fitting the data. However, this is not the case. A careful look at the percent likelihood of individual cooling paths shows that more complex cooling histories fit the single 40 Ma age equally well (Figs. 4C and 4D), although such histories were not extensively explored because QTQt preferentially converged on the simplest solution. For example, several of the MaxLike models show a percent likelihood >80% for thermal paths that resemble the true Paths 1–6 (Fig. 4C). This is why, to better constrain thermal histories, data sets should typically include more information than a single thermochronometric age, such as ages from multiple independent chronometers or additional thermal history proxies such as fission-track lengths or the variability in multi-aliquot AHe or zircon (U-Th-Sm)/He (ZHe) ages caused by grain size or radiation damage.

Comparing these results with those produced by an inversion of a single 40 Ma AHe age in HeFTy (Fig. 4E; Murray et al., 2022) illustrates several fundamental differences between the way these programs search t-T space and present inversion results to the user. In contrast to QTQt’s MCMC algorithm, HeFTy employs a non-learning random Monte Carlo search of t-T space; it presents results as either acceptable- or good-fit paths with no colored weighting such as the relative probability color maps produced by QTQt. As a result, the

wide range of acceptable- and good-fit paths are more apparent. Upon close inspection of QTQt’s color map (Fig. 4A) and MaxLike plots (Figs. 4C and 4D), we see that the results from the two programs are, in fact, quite similar. The QTQt ExTH color map covers the same t-T ranges as the HeFTy acceptable field, and the QTQt paths with higher-percent likelihood represent the same possibilities seen in the HeFTy good-fit paths (Fig. 4).

The difference between the QTQt inversion results (Figs. 4A–4D) and HeFTy inversion results (Fig. 4E) directly reflects the philosophical differences in how these programs handle the non-unique nature of cooling ages. In HeFTy, a single cooling age that is fit equally well by a wide range of simple and complex thermal histories will produce an inverse model that easily finds a wide range of good- and acceptable-fit t-T paths (see also other examples in Murray et al., 2022). In contrast, QTQt penalizes complexity when the data inputs are simple. As a result, it converges upon the simplest solution that fits the data to the apparent exclusion of more complex histories that fit the data equally well. One approach is not generally better than the other; they are simply different. Importantly, these differences prompt the user to conceive of modeling questions, model design, and how to make appropriate geologic interpretations of inversion results in different ways. For more comparisons between these programs, see Fox and Carter (2020), Flowers et al. (2022b), and the commentaries related to Flowers et al. (2015): Vermeesch and Tian (2014); Flowers et al. (2016); Gallagher et al. (2016); Gallagher and Ketchum (2018, 2020); and Vermeesch and Tian (2018, 2019).

### Model Set-Up Decision Making: Inputs and Outputs

Using the synthetic data generated from the forward models of thermal history Paths 1–6 as “perfect” input data for inverse models, we explore how decision making during the model design process controls the inverse thermal history modeling outputs. First, we jointly invert the AFT and AHe data and run all inverse thermal histories using QTQt’s default t-T prior information (time = oldest observed age  $\pm$  oldest observed age; temperature = 70  $\pm$  70  $^{\circ}\text{C}$ ) and an initial constraint that ensures that each model starts under the same conditions and that all thermochronometers have a zero age at the beginning of the model ( $t = 200 \pm 5$  Ma,  $T = 200 \pm 5$   $^{\circ}\text{C}$ ). This initial constraint ensures that the thermochronometers are initially completely reset for each model and has no geological significance; the effect of the initial constraint is discussed more in the Effect of an Initial Constraint section below. We add no other constraint boxes to guide QTQt to attempt paths that start at 100 Ma. This exercise gives us (1) a baseline for how well QTQt inverse models can reproduce known cooling paths given “perfect” data, an important reference point for the sensitivity tests we outline below, and (2) an opportunity to discuss the differences in the various outputs that QTQt produces and how they may be useful for interpretation of different features of the thermal histories.

In the following subsections, we iterate between designing, running, and interpreting different inverse thermal history models to explicitly test the

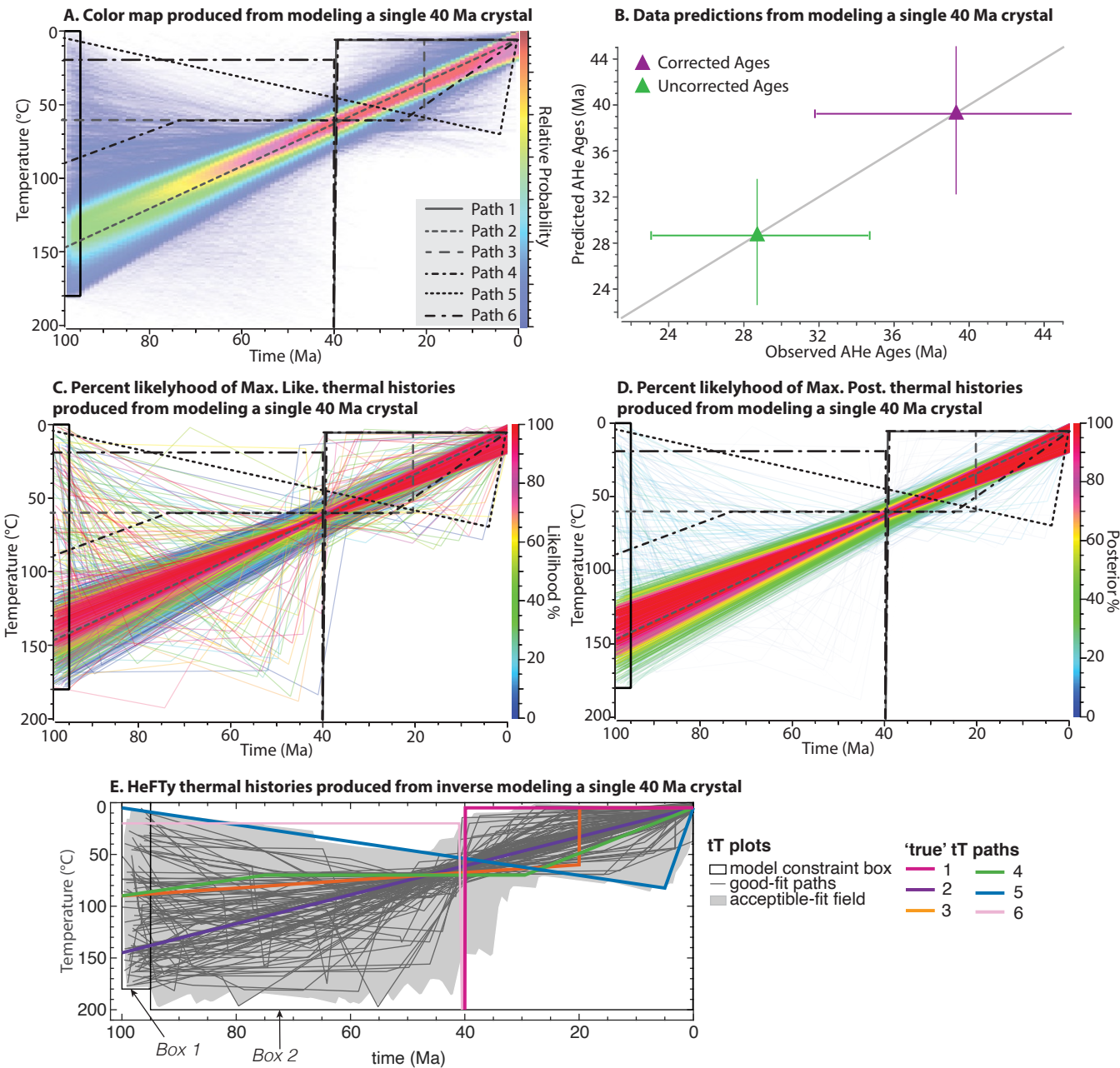


Figure 4. (A) Expected t-T produced from inverting a single apatite grain with a 60  $\mu\text{m}$  radius, an eU concentration of 60 ppm, and an age of 40 Ma. Six paths used for forward modeling and synthetic data production overlaid on a relative probability color map. (B) Data predictions for the expected thermal history compared with input data (observation); the diagonal line is 1:1 correlation between predicted and observed, and the error bars are  $2\sigma$ . Note, we use the Flowers et al. (2009) radiation damage accumulation and annealing model (RDAAM), and the ages reported here show both the uncorrected and corrected ages. (C) Representative population of accepted thermal history paths, colored by their likelihood probability in percent. (D) Representative population of accepted thermal history paths, colored by their posterior probability in percent. (E) Results for a single 40 Ma AHe age modeled in HeFTy (after Murray et al., 2022) demonstrate key differences in the modeling philosophy and model result visualization between these two commonly used programs.

impact of the following choices: using one or multiple chronometers as data, changing the size of the time-temperature range prior, changing the time-temperature position of an initial constraint box, and changing the AHe age uncertainty.

In our examples, we use the Ketcham et al. (2007) annealing model to simulate fission track production and annealing. We use  $D_{par}$  as the compositional proxy and assign a typical Durango value of  $2.05 \pm 0.20 \mu\text{m}$  (Sobel and Seward, 2010) and determine the initial track length for each proposed model based on a composition sampled from this  $D_{par}$  range. We use the synthetic Ns and Ni (number of spontaneous and induced tracks, respectively), data for 20 single grains, and 100 c-axis projected track lengths generated during the forward modeling phase as the raw input data (Table S2, see footnote 1).

To simulate He diffusion from apatite crystals, we use the Flowers et al. (2009) RDAAM. Although in our forward modeling exercise, we generated many AHe ages across an eU range of 1–300 ppm for three different grain sizes, in this exercise we only use six synthetic single-apatite grains that all have the same spherical equivalent radius of  $60 \mu\text{m}$  but different eU values (eU values = 10, 25, 50, 75, 150, and 300 ppm). For a discussion on how to model more, or fewer, grains with different degrees of age variability linked to grain geometry and radiation damage, see Brown et al. (2013). For each AHe age, we assign an uncertainty of 10% to reflect analytical uncertainty, typical reproducibility of standards, and uncertainty due to our incomplete understanding of the causes of AHe age variability. The specified model inputs described here are used in the following exercise to resolve the “true” thermal histories (Fig. 5) and reported in our template found in Table S2 (see footnote 1).

### Resolving “True” Thermal Histories

Visually, the inverse model outputs show good recovery for some of the “true” cooling paths (e.g., Paths 1, 2, 5, and 6) whereas others (Paths 3 and 4) less clearly recover the true path (Fig. 5). For all paths, the parts of the model results at temperatures above  $120 \text{ }^\circ\text{C}$  have no real significance because the apatite chronometers are only sensitive to temperatures between  $30\text{--}120 \text{ }^\circ\text{C}$ ; so we focus our discussion on the parts of the paths  $<120 \text{ }^\circ\text{C}$ . The inverse modeling results for Paths 1, 2, and 6 yield ExTH, MaxLike, and MaxPost models that all reproduce the “true” paths well (Fig. 5). The 95% credible intervals for these three cooling scenarios are narrow along the entire t-T space. For each of these three paths, the three output models predict ages and track lengths that are nearly indistinguishable from one another. Additionally, in Paths 1 and 6, we observe that some thermal histories imply cooling and reheating prior to the rapid cooling event at ca. 40 Ma. However, although these paths may result in a reasonable data fit, the number of paths that do this is very low and as such can be considered to have a very low relative posterior probability (as indicated by their blue color on the color map), and they lie outside the 95% credible intervals of the expected thermal history (Figs. 5A and 5F). The

monotonic cooling inferred in Path 2 is extremely tightly constrained and is inferred from almost all proposed models (Fig. 5B). This can be attributed to the fact that the monotonic cooling is a very simple t-T path, which in this case, fits the data extremely well, and so there is no need to introduce any additional complexity.

Paths 3, 4, and 5 show notable differences between the MaxLike, MaxPost, and ExTH t-T paths. Additionally, these paths show broader 95% confidence intervals where acceptable predictions vary greatly and incorporate the most structure (i.e., inflection points) in output paths. The MaxLike thermal history captures the true shape of Paths 3 and 4 the best, both visually and in the predicted AHe and AFT ages (Figs. 5C and 5D). The MaxPost paths for Paths 3 and 4 are simpler (have fewer t-T points) than the MaxLike model but still have structure (i.e., inflection points in the path) inconsistent with the true path. The ExTH model incorporates the structure inferred by each individual model but is smoothed due to the weighting toward the posterior (i.e., simpler) models. In contrast, for Path 5, the MaxPost model visually fits the true history better, but the age predictions are virtually indistinguishable from the more structurally complex MaxLike model predictions (Fig. 5E). The ExTH model for Path 5 smooths the structure of the cooling paths so that t-T changes are less abrupt.

A major challenge when assessing inverse thermal history models is determining which features of the model are justifiable for geological interpretations and how these features should be interpreted. In this example, we are able to compare the model results with a known “true” thermal history. In a real data set, we may compare and assess results with geological hypotheses. In Paths 3 and 4, we see that the 95% credible intervals, between 100 and 30 Ma, are broad, and the ExTH path shows some minor cooling and heating. Given the uncertainty on the model over this timeframe, and the ambiguity between the MaxLike and MaxPost models, it is not advisable to interpret these minor temperature deviations with any geological significance unless justified by some other data, observation, or hypothesis. The 95% credible intervals become narrow and tightly constrain linear cooling from ca. 30–25 Ma until 0 Ma through temperatures of  $10\text{--}60 \text{ }^\circ\text{C}$ . For Path 4, this inference matches the known t-T path, but for Path 3 the ExTH infers that cooling starts earlier than the true history and does not capture the rapid cooling. If the true path was not known, there would be no way of identifying that the Path 4 ExTH model was consistent with the true history while the Path 3 ExTH model was not without additional, independent geologic information. In fact, the true timing for the onset of the final cooling event in the Path 3 model does fall within the 95% credible intervals and is recovered well by the MaxLike model. This suggests that the thermochronologic data set is consistent with cooling at this time; however, this solution is non-unique given the input data, the uncertainties, and the preference for the Bayesian approach of QTQt to find simpler solutions (Gallagher, 2012). Geologic interpretations should reflect these uncertainties.

These simple examples demonstrate how resolving near-surface, low-temperature, thermal events such as those in Paths 3, 4, and 5 with apatite chronometers from a single outcrop sample is challenging and may not be possible. In such a circumstance, it is therefore up to the user to decide which





model (e.g., ExTH, MaxLike, MaxPost) they prefer to interpret, to justify this choice using the observed data and coherence with available independent geological information, and to discuss the limitations of the inversion result and how additional observations or data could improve the result. In the Supplemental Material (footnote 1), we reproduce Path 3 very well with more information that is rarely attainable in a real-life scenario, such as a very high number of AHe ages covering a larger range of grain sizes and eU (and therefore temperature sensitivities), perfect predictive models for He diffusion and fission-track annealing, and adding the requirement that the sample experiences no reheating during its geological history (Fig. S1, footnote 1).

### Effect of Using One Thermochronometer versus Multiple Chronometers

The commonly used thermochronometric systems are sensitive over different and sometimes overlapping temperature ranges. Therefore, applying multiple dating techniques on a single sample can provide more information on the sample's thermal history than using any one system in isolation. However, in practice, research questions and budget considerations may limit the researcher to analyzing only one or two systems. Therefore, it is important to appreciate and acknowledge the temperature sensitivity range and thermal history resolution that can be obtained with the thermochronometers used and to identify where the thermal history will be poorly resolved. Next, for Paths 1–6, we invert synthetic data from the AFT and AHe systems separately to illustrate their separate contributions to resolving the true paths.

Paths 1 and 2 are equally well resolved if AFT or AHe data are used in isolation, or if they are combined and jointly inverted together (Figs. 5 and 6). For Path 1, this is because a thermal history with very rapid cooling through partial-annealing and retention temperatures produces a distinct form of AFT track-length data (long, unimodal, narrow distribution) and AHe single-grain age data (flat age-eU and age-grain-size relationships), respectively, and thus there is a narrow range of thermal histories that reproduce these data (Fig. S2, footnote 1). For both paths, any additional complexity in the attempted t-T histories produces a poorer data fit and is rejected. The inverse thermal history deviates from the true histories at temperatures of 100–140 °C; however, this is to be expected because such temperatures are too hot for the AFT and AHe data to record thermal history information, and the thermal history path is being pulled in the direction of the initial constraint added to the model (at 200 Ma, 200 °C).

For the more complex history of Path 4, the ExTH model obtained using either AHe or AFT data alone shows that cooling through 90 °C at 100 Ma is well constrained. At temperatures <70 °C the AFT data do not tightly constrain detail in the thermal history, as shown by the broad 95% credible intervals. However, the ExTH path does approximate isothermal holding, albeit with a subtle oscillation and at a lower temperature than the true Path 4, before the initiation of cooling at the correct time of ca. 25 Ma. The subtle oscillation of the ExTH within the 95% credible intervals is likely a consequence of individual t-T

paths, from which the ExTH is derived, having more extreme oscillations (i.e., heating and cooling events around the temperature where the true history has isothermal holding). With only AHe data, the model tightly constrains the early part of the history (until 80 Ma) and the final episode of cooling from 25 Ma to present. In the period between 80 and 30 Ma, the ExTH path shows minor cooling and reheating within broad 95% credible intervals (Fig. 6). This is a consequence of individual thermal histories of varying complexity (spanning the credible interval range) being averaged out.

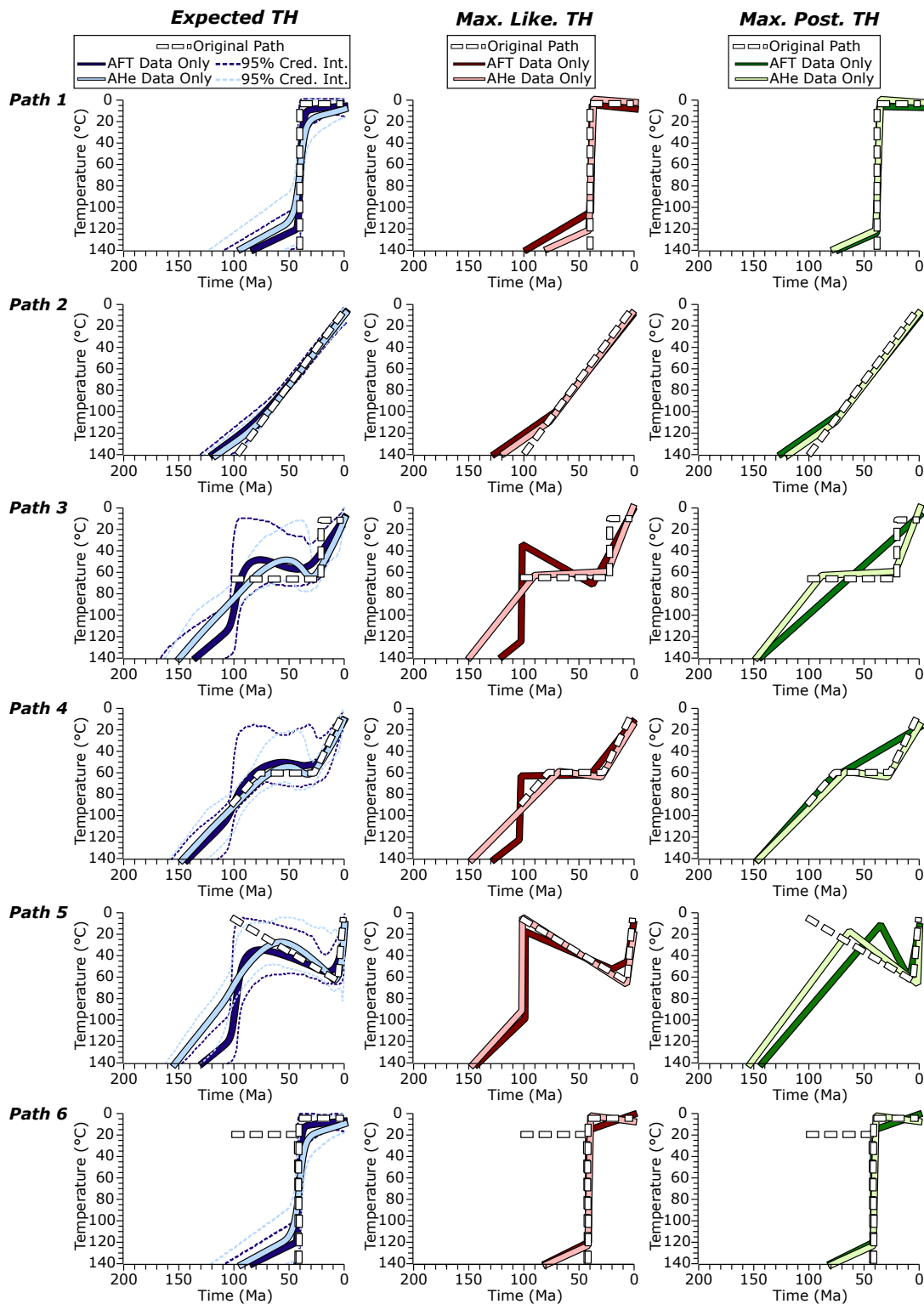
In the Path 4 example, the MaxLike and MaxPost models, when using AHe data only (MaxLike<sub>AHe</sub>, MaxPost<sub>AHe</sub>), are a very good reflection of the true thermal path (Fig. 6). The AFT data resolve only certain parts of the history, such as the isothermal holding at 60 °C in the MaxLike<sub>AFT</sub> model, or only the slow cooling at the beginning of the thermal history in the MaxPost<sub>AFT</sub> model. This suggests that if the RDAAM model is accurate and well calibrated and data spans a sufficient range of eU values, then an inversion retrieves the true thermal history. It is also worth noting that the main feature that distinguishes Path 3 from Path 4, rapid versus slow cooling for the final event starting at 20 Ma or 25 Ma, respectively, cannot be resolved in any of the models; the more slowly cooled scenario is preferred across all of them. As discussed in the Effect of Using One Thermochronometer versus Multiple Chronometers section above, when we jointly invert the AFT and AHe data, we produce an ExTH model that is mostly similar to the AHe-only ExTH model, with the same oscillating cooling and heating structure over the time period where there should be isothermal holding (Figs. 5 and 6). Together, these results suggest that the AHe data exert more influence on the thermal history model results presented in this work than the AFT data. Tightly constrained parts of the thermal history include cooling through 100 °C at 100 Ma and the onset of cooling at 25 Ma. The erroneous cooling and reheating episode is evident in the MaxPost model, while the MaxLike model is a good approximation of the true solution.

### Effect of the Prior Time-Temperature Ranges

Next, we test the effect of changing the “Ranges for General Prior,” which sets the prior information on the time and temperature parameters (i.e., defines an initial probability distribution for time and temperature before any other evidence is introduced) and defines the area in t-T space where t-T paths can be proposed. Due to the nature of transdimensional MCMC sampling and the criterion to accept proposed models, a broader range of the t-T priors will often produce simpler thermal history solutions than a more tightly constrained prior. A broad prior can be thought of as a relatively vague state of information that is rapidly improved upon when data are added, even with a simple thermal history. Therefore, it is important that the ranges for the general prior are both justified and sufficiently broad such that the complexity (or simplicity) of the model is not overly forced, but not too broad such that model complexity is overly penalized.

For example, setting a temperature prior that covers a temperature range from the minimum estimate of the present-day surface temperature to a little





**Figure 6.** Model results from inverting (1) the apatite fission-track (AFT) data only and (2) the six single-grain apatite (U-Th-Sm)/He (AHe) ages only (see text for details) generated from each of the six cooling paths. Left: ExTH model. Middle: MaxLike model. Right: MaxPost model. In all plots, the white dashed line is the true path. Darker line colors are the model path for (1) and lighter line colors are the model paths for (2). Thin dashed colored lines in left-hand plots are the 95% credible intervals (Cred. Int.) on the ExTH models. The data are reproduced in a manner comparable to that observed in Figure 5, or in many cases, the data are better reproduced because the models in these exercises are only fitting a subset of the data (i.e., the AFT data only or the AHe data only). See Figure S2 (text footnote 1) for data predictions.

more than the maximum temperatures a thermochronometer is sensitive to on geologic timescales is reasonable. The default prior temperature range when working with apatite thermochronometry data is automatically set to  $70 \pm 70$  °C in QTQt. This covers temperatures from 0 to 140 °C, which encompasses the typical AFT-PAZ and the AHe-PRZ and sufficiently focuses model sampling in the area over which such data are informative.

The “time” prior is harder to define. In our initial examples, the time prior is set to the oldest AFT or AHe single grain age  $\pm$  oldest AFT or AHe single grain age (e.g., if the oldest age in the model is 50 Ma, the time space is  $50 \pm 50$  Ma or 100–0 Ma); this is the default option in QTQt v. 5.7.1, used in this contribution. However, a broader prior may be more appropriate if the model should be investigating a period of geologic time longer than the total time period defined by twice the oldest age. Alternatively, when modeling multiple samples separately, it may be more appropriate to have a prior that is the same across all models, rather than a prior that changes depending on the oldest AFT or AHe age yielded by a particular sample. Here, we demonstrate how broadening the time component of the prior influences how well we can reproduce a simple thermal history and a complex thermal history. We also show to what extent this effect is amplified when we only use the AFT data set. Unless specifically mentioned, the data fit for all models is good and indistinguishable between models (see Fig. S4, footnote 1, for data predictions).

We show the inverse modeling results when using Paths 1 and 4 AFT and AHe synthetic data, respectively (Fig. 7). For both models, the default prior and an initial constraint of  $1000 \pm 5$  Ma,  $200 \pm 5$  °C have been used. The expected model reproduces Path 1 almost exactly. For Path 4, the model reproduces the timing of cooling, and paleo-temperature at the onset of cooling reasonably well. However, the isothermal holding is not well reproduced due to the averaging effect of the range of acceptable models, although it is within the 95% credible intervals.

Next, we use the same input data but change the time component of the general range for prior: (1)  $150 \pm 150$  Ma, (2)  $250 \pm 250$  Ma, and (3)  $500 \pm 500$  Ma (Figs. 7 and S4, footnote 1). For Path 1, the inverse models show that altering the time prior has no effect on our ability to retrieve the true model, at least in this idealized case with “perfect” input data and models of annealing and diffusion. The nature of the data produced from the Path 1 thermal history is so tightly constrained that it easily overcomes the prior effect. Path 4 shows that the difference between the ExTH model for the default prior and for the largest prior is negligible, with only a minor difference in the ExTH path where there is cooling and reheating between 100 and 50 Ma, and comparable 95% credible intervals.

We then repeat the test using only AFT data to investigate to what extent the prior effect is amplified when using a smaller data set. Again, for Path 1, changing the prior causes no impact on our ability to retrieve the true model. The AFT data are simply too well constrained for any models other than rapid cooling to be acceptable and broadening the prior has no influence over this.

For Path 4, the effect of broadening the prior can have a significant impact on our ability to find the true model. For the default prior, the ExTH path initially

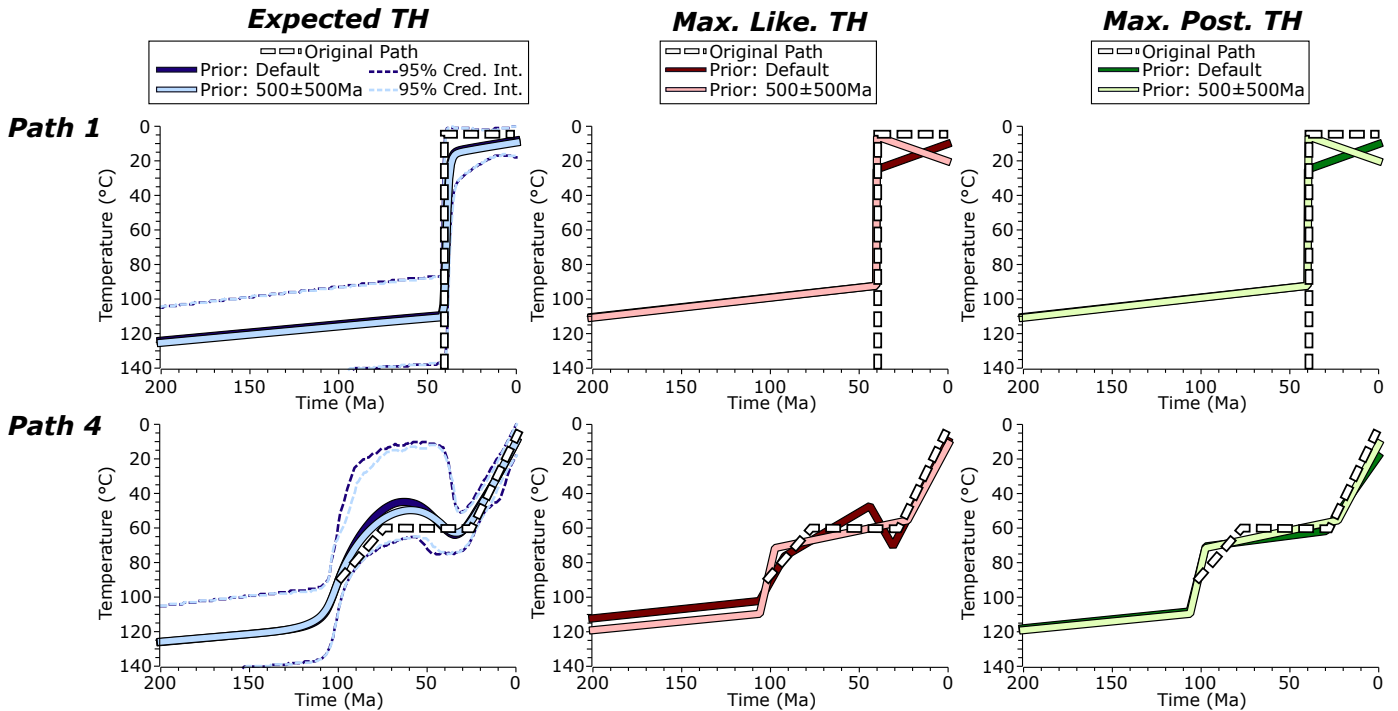
shows cooling followed by holding at 55 °C between 70 and 45 Ma before cooling again to present-day surface temperatures (Fig. 7). As we increase the prior, we observe that the inflections in the ExTH path over this time become lessened and smoothed out (Fig. S4, footnote 1) until the ExTH path is reduced to a simple continually cooling path (Fig. 7). In addition to smoothing out of the path, the 95% credible intervals also become narrower and tighter to the ExTH path. With small to moderate increases in the prior the uncertainties are large enough such that the models are still comparable, and any misinterpretations caused by the prior effect will be relatively minor. However, with the largest prior, the 95% credible intervals are narrow and very tight to the ExTH path suggesting this thermal history inferring slow monotonic cooling, is particularly well constrained (Fig. 7). Clearly, the model produced using the default prior (oldest age  $\pm$  oldest age) and that produced with the largest prior ( $500 \text{ Ma} \pm 500 \text{ Ma}$ ) would have entirely different geological implications, and so broadening the time prior by large amounts should be done with caution, particularly if modeling a small data set with limited thermal history information in the data.

In summary, broad priors tend to produce simpler solutions. However, large or independently constrained data sets will counteract this effect; data sets with multiple thermochronometers have more information that is able to overcome the weighting exerted by a wide prior. At the same time, a too tightly defined prior can lead to misleading well-resolved results. Therefore, the modeler should purposefully tailor their choice of a prior to their data set and the questions motivating their modeling and explain this choice as a part of documenting their modeling process. When working with large data sets, testing many different prior ranges for each sample is likely to be inefficient; so a logical starting point would be to define a time prior that is sufficiently broad such that it encapsulates all of the observed ages and an estimate of the geological time period where the thermal history may be resolved (e.g., the oldest observed age  $\pm$  oldest observed age). In contrast, if the modeler is considering using a prior that is excessively large relative to their observed ages and temperature sensitivity of the thermochronometers used, then they should justify this choice and demonstrate what effect it may have on the thermal history, particularly if the data set contains limited information (e.g., a single thermochronometer).

## Effect of an Initial Constraint

Next, we explore whether explicitly stipulating an initial constraint (i.e., a specific model start time) impacts the output thermal history. In QTQt, a starting time and temperature for running inversions are not required. If no initial constraint box is defined, QTQt will commonly infer rapid cooling immediately before the first point proposed by the model that is constrained by the data. This implied period of rapid cooling will depend on the nature of the data, which could require that the initial cooling is through both the PAZ and PRZ (e.g., Path 1) or is only to temperatures somewhere *within* the PAZ

**AFT and AHe Data**



**AFT Data only**

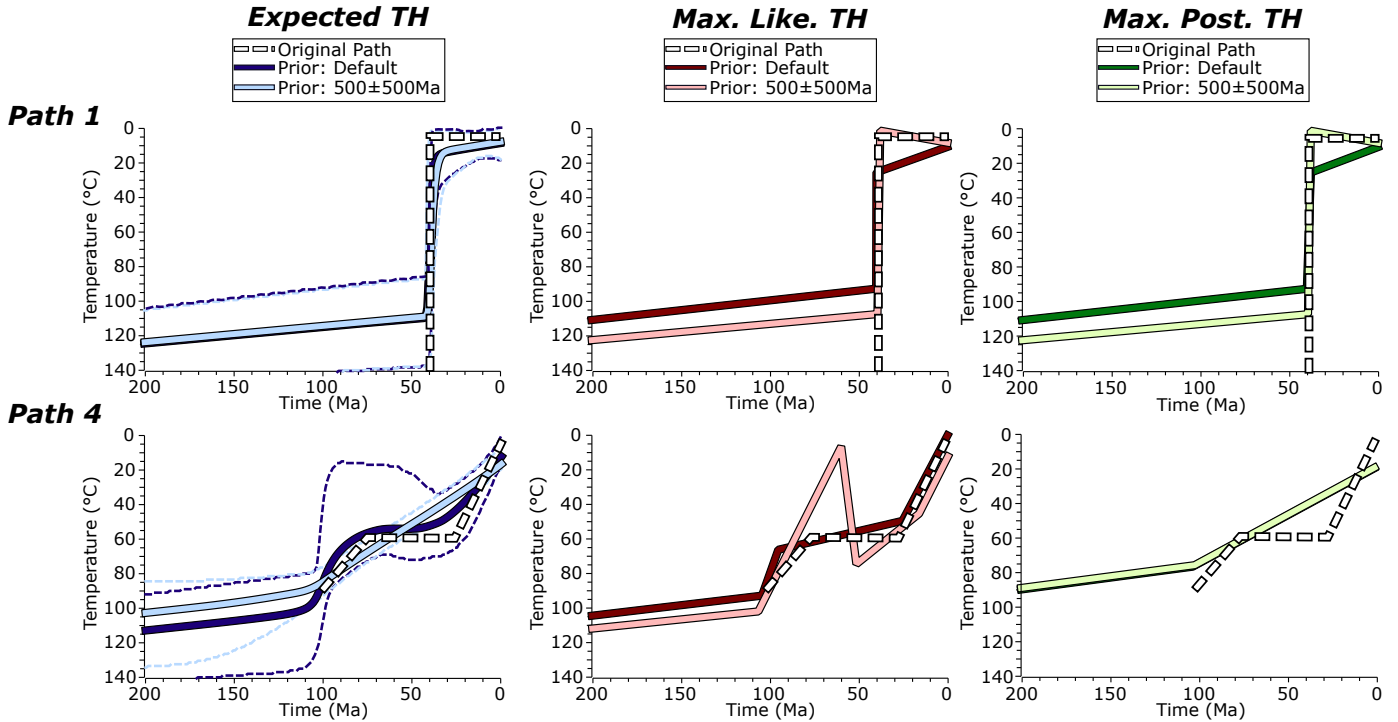


Figure 7. Model results from inverting the synthetic data from Paths 1 and 4 with different inputs for the “General range for prior.” Additional plots for a prior of  $150 \pm 150$  Ma and  $250 \pm 250$  Ma are provided in Figure S3 (text footnote 1). Top plots show the results from inverting apatite fission track (AFT) and apatite (U-Th-Sm)/He (AHe) together. Bottom plots show the results of inverting only AFT data. Left: ExTH models. Middle: MaxLike models. Right: MaxPost models. In all plots, the white dashed line is the true path. Darker line colors are the model path when using the default “General range for prior,” and lighter line colors are the model paths when setting a much larger time span for the “General range for prior.” In both scenarios, the prior temperature remains the same at  $70 \pm 70$  °C. Thin dashed colored lines in left-hand plots are the 95% intervals on the ExTH models.

or PRZ (e.g., Path 4) (Figs. 8A and 8B). Here we investigate whether including an initial constraint is useful and, if so, where in t-T space is an appropriate or inappropriate place to put that initial constraint. We address these points using Paths 1 and 4 for examples and use the default prior (explored above). We compare an inverse model with no initial constraint box to inverse models testing an initial constraint with a small temperature window ( $200 \pm 5$  °C) and a large temperature window ( $100 \pm 100$  °C). Each of those constraint boxes is placed at three different times before the timeframe encompassed by the “true” path ( $200 \pm 5$  Ma,  $500 \pm 5$  Ma, and  $1000 \pm 5$  Ma).

Without an initial constraint box, the temperature at the beginning of the Path 4 ExTH model is  $\sim 75$  °C at 100 Ma and  $\sim 30$  °C at 40 Ma in the Path 1 ExTH model (Figs. 8A and 8B). In both cases, these times and temperatures correspond to the conditions when daughter products in both the AHe and AFT systems were starting to be retained in the true paths. The inverse model results suggest rapid cooling to these temperatures immediately before this time. As such, Path 1 is reproduced very well. However, without an initial constraint the Path 4 ExTH fails to capture the higher temperature cooling from 100 to 80 Ma, although the 95% credible intervals do encompass that part of the true path (Figs. 8A and 8B).

Adding a constraint box at any time before the initial time point of the true t-T paths (i.e., before 40 Ma for Path 1 or before 100 Ma for Path 4) helps to retrieve the full structure of the true cooling paths (Figs. 8C–8F). We observe an apparent improvement in the 100–80 Ma cooling in the Path 4 inversion as the ExTH is drawn toward the imposed constraint, although the difference is small, and both fall within the 95% credible intervals. There is essentially no difference in cooling history output if the box is placed at different times before the model start time defined by the original input. Similarly, there is little difference in model output if the constraint box is tight, spanning a narrow temperature range, or expanded, spanning a broad temperature range (Figs. 8C–8F).

We conclude that if the initial constraint box is placed at a time before the onset of daughter product retention in the highest-temperature thermochronometer, then it does not matter what range of temperatures and time it covers. We recommend that the temperature range includes the possibility of exploring temperatures hotter than the PRZ/PAZ of the highest temperature chronometers being modeled. However, users should not make interpretations (i.e., heating or burial rates) based on the cooling path between the initial constraint to the point where the path passes through the maximum temperature limit of the modeled thermochronometric systems (e.g.,  $\sim 110$ – $120$  °C in the case of the systems modeled in this example). In the absence of an initial constraint, the thermal history may start at a point colder than the thermochronometer PRZ/PAZ, with the implication of rapid cooling immediately before this time. This may be aesthetically unappealing, but as we have shown, imposing an initial constraint will create an additional segment of the thermal history path between the constraint box and the point where the thermal history passes through the maximum closure temperature of the methods being used that is entirely the product of the constraint box, not the data. An initial constraint box

can be useful beyond aesthetics if we know that at some point in a sample's history it existed at higher temperatures (e.g., during igneous rock formation).

We recommend that the data are first inverted with no initial constraint box. Then, if the user wishes, an initial constraint box that is hypothetical or supported by independent geological or geochronological information can be added. Clearly, if there is a known, well-dated geological constraint, then this should guide the time and temperature range of the initial constraint.

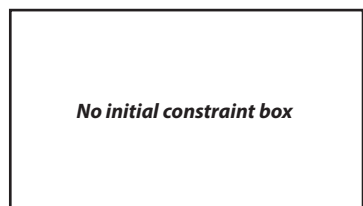
## Effect of Uncertainties

Lastly, using Paths 1 and 4 as examples, we explore the effects of data uncertainty by investigating the differences in model outputs depending on data uncertainty assignment choice. Presently, most ages are reported with analytical uncertainty, which can be quite small and generally does not capture all the factors that can influence the uncertainty on a thermochronologic age. Confidently identifying and quantifying all possible uncertainty on cooling ages is a challenge being addressed by ongoing research across the thermochronology community (e.g., Flowers et al., 2022b; Ketcham et al., 2022). In acknowledgment of this challenge, QTQt offers users the ability to choose the option of changing uncertainties on data when modeling. Here we test three types of uncertainty: (1) a random error assignment made by QTQt when producing the synthetic AHe data (these errors will change with every forward model run). This assignment skews to lower errors that are  $<1\%$ – $10\%$  of the predicted age and would reflect analytical errors that may be low but variable. (2) This type of uncertainty encapsulates analytical and methodological uncertainty and/or reproducibility of standards. This can be subjective and result in uncertainties that have absolute values that scale with the observed age such that older ages are less precise. (3) A 2 Ma uncertainty assignment to all ages produces 3%–13% error on the various input ages in our examples presented in this work. This would cause older ages to be treated as being more precise than younger ages.

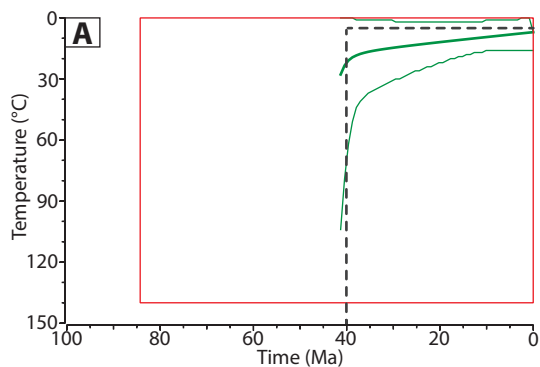
For Path 1, the model output is similar regardless of the choice of uncertainty assignment. The predicted AHe ages are almost identical, and thus changing the uncertainty does not have any influence on the model output. On the other hand, Path 4 shows marked differences between the cooling path outputs. The main difference comes from the model output generated from the data with the random error assignment (Fig. 9).

This random error assignment can be considered as an analogue for the analytical error attributed to analyzed “unknown” samples. In this case, the random error assignment put a relatively small uncertainty on all of the AHe ages ( $<10\%$ ), and a few of the ages were given very small uncertainties (e.g., the youngest and third youngest ages; Fig. 9). These small uncertainties imply that we are confident that the data from those particular samples are highly precise and therefore will have a greater influence over the inferred cooling history. In the case of Path 4, the tightly constrained samples add much more detailed structure to the output cooling history. That detailed structure shows

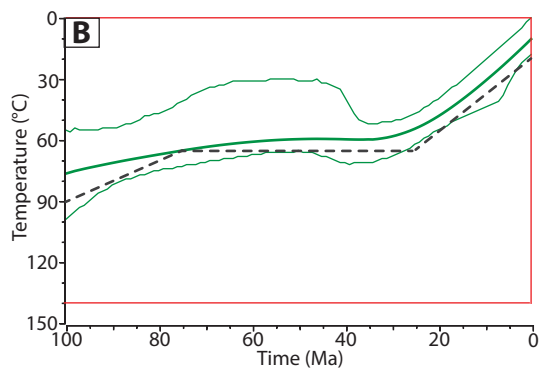
Model set-up



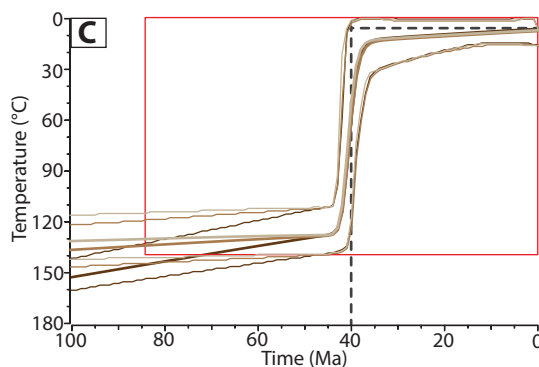
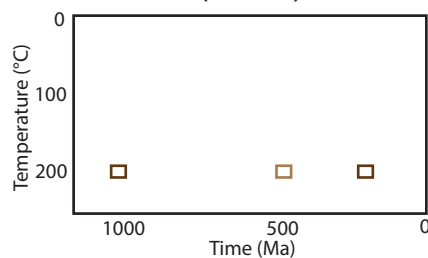
Path 1



Path 4

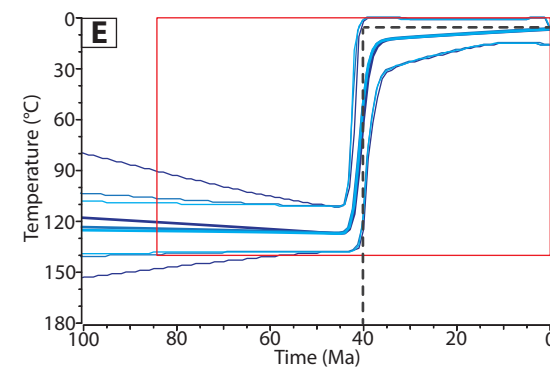
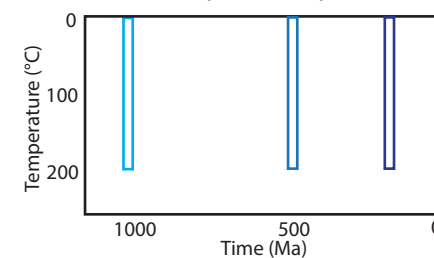


Tight temperature constraint  
( $200 \pm 5^\circ\text{C}$ )



- Constraint time:  $200 \pm 5$  Ma
- Constraint time:  $500 \pm 5$  Ma
- Constraint time:  $1000 \pm 5$  Ma

Expanded temperature constraint  
( $100 \pm 100^\circ\text{C}$ )



- Constraint time:  $200 \pm 5$  Ma
- Constraint time:  $500 \pm 5$  Ma
- Constraint time:  $1000 \pm 5$  Ma

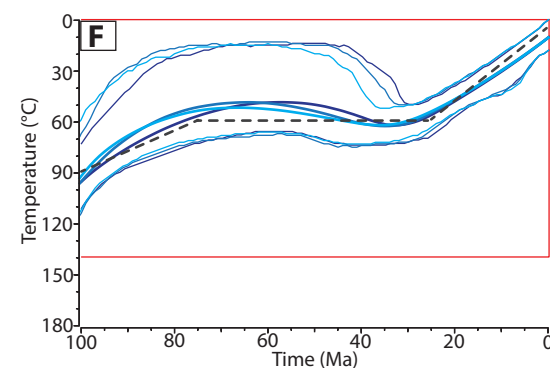
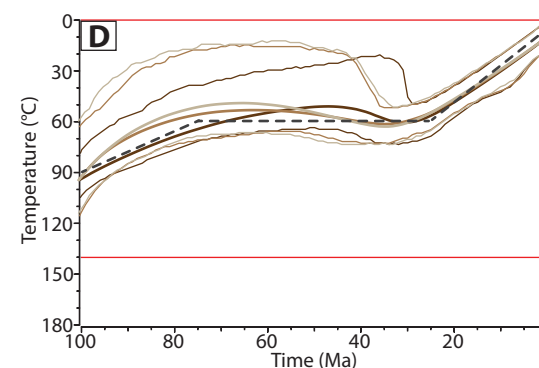


Figure 8. Model results from initial constraint box testing. Top: Model set-up: No initial constraint box (left); tight temperature range initial constraint box placed at three different points in time previous to the start of the part of the model informed by the data (middle); broad temperature range initial constraint box placed at three different points in time previous to the start of the part of the model informed by the thermal history data (right). (A, C, E): Path 1 ExTH (data predictions in Fig. S5, see text footnote 1). (B, D, F): Path 4 ExTH (data predictions in Fig. S6).



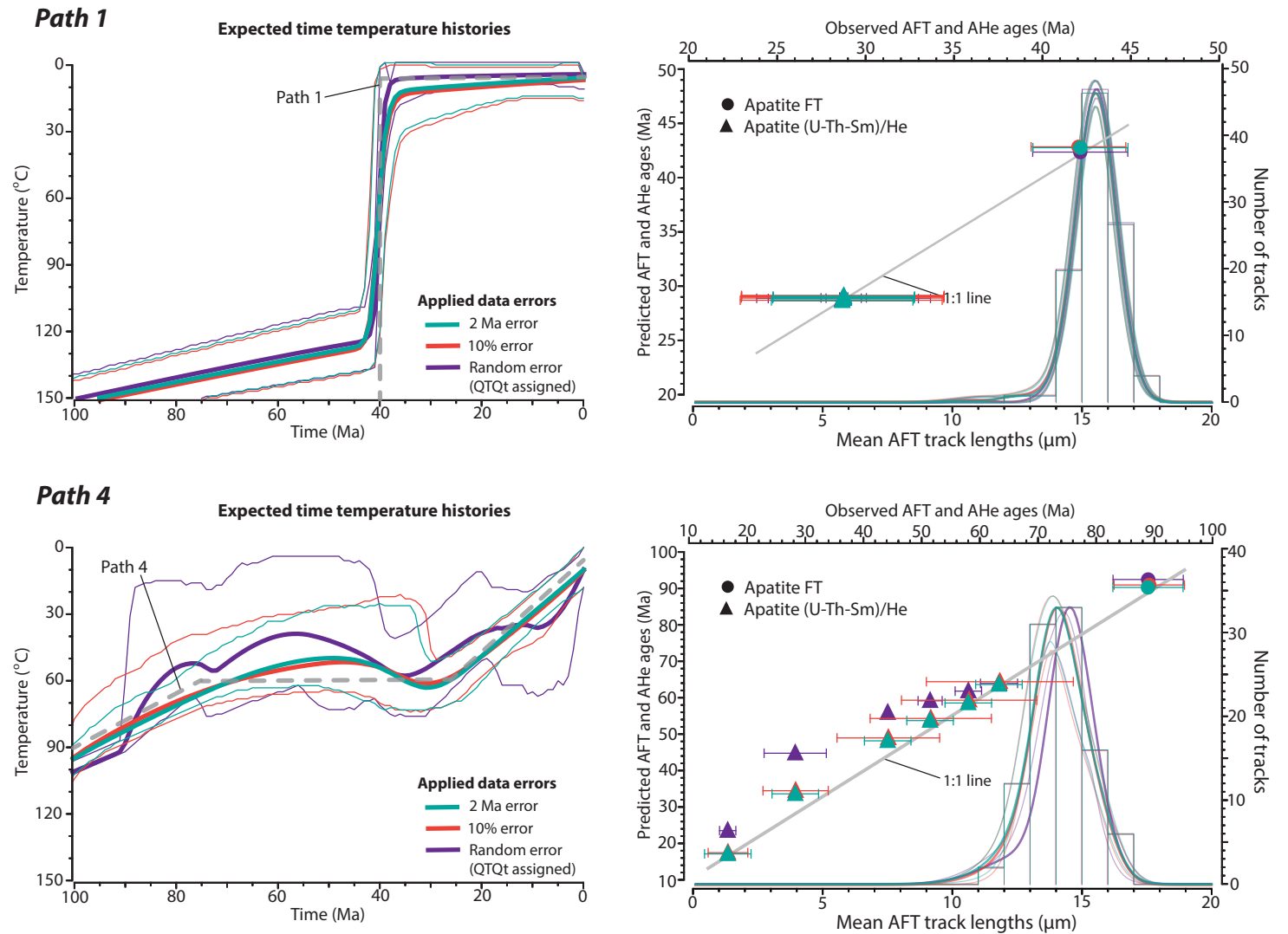


Figure 9. Comparisons between the ExTHs produced when testing data error assignment. Top: Path 1 thermal history and predictions. Bottom: Path 4 thermal history and predictions. Apatite fission-track (AFT) ages are central AFT ages, and apatite (U-Th-Sm)/He (AHe) ages are uncorrected for alpha-ejection.

the appearance of multiple heating and cooling events; however, those events are small—changes of 5–20 °C over time spans of 10–20 m.y.—and all occur within the AHe PRZ. In this case, the assignment of incredibly small errors on some of the data is giving a false signal of great detail in the cooling history, and the 95% credible intervals show how varied and non-unique those slight heating and cooling episodes are. This is an example of why we generally caution against interpretation of small (<15 °C) oscillations in a thermal path resolved by a QTQt inversion, especially in a PRZ or PAZ, without detailed sensitivity testing to determine what is controlling those oscillations.

We are better able to recover the “true” cooling history of Path 4 once we expand the data uncertainty and give more of an equal weighting to the individual AHe ages. In fact, there is little difference in the inverse model output using an error assignment of 10% versus a uniform 2 Ma (Fig. 9). However, using more expanded or uniform error assignment means that we chose to forgo the ability of QTQt to incorporate our confidence in the quality of data used within a model. Often, with real data, there is a reason to expect and retain a small error value on certain data that are deemed to be “high-quality.” In such a case, we do not want to lose that quality assessment, because that information is an input used by QTQt to find acceptable cooling paths. However, there is also the potential that an age is highly precise, but inaccurate, which would lead to confidence in an erroneous model solution. Additional user options in QTQt (discussed below) allow users to resample AHe ages or errors to assess the impact of different individual age measurements.

QTQt also has the ability to scale AHe errors such that they are treated as unknown parameters that can be sampled during the model run. The effect of this error scaling is typically to increase the error on, and thus reduce the influence of, single AHe ages that may be outliers because QTQt finds that their measured ages are not a function of known sources of He age variability that are modeled, such as grain size or radiation damage. The error rescaling factor can be defined by the user from 0.001 to 1000. This error handling approach is a way of recognizing that we know the observed age, but we have a poor understanding of the uncertainty on it. It offers a way of handling dispersed AHe data sets without manually excluding individual ages from thermal history analysis.

Another option QTQt offers is resampling the observed age from an assigned uncertainty. In this context, we recognize that our observed age is only an estimate of the true age. Intuitively, we can argue that it is better to try and fit the true, but unknown, age than the noisy measured age. This true age lies somewhere within the age uncertainty range. Then instead of comparing the predicted model age to the measured observed age, the predicted age is compared to a “sampled observed” age that is randomly sampled from the uncertainty range during the inversion. Typically, these ages are sampled from a normal distribution with a mean equal to the observed age and standard deviation equal to the input error. So, the resampled ages will tend to be close to the observed age, but it is possible to sample younger or older values. This approach is really only recommended when we have multiple AHe ages, because a single age resampling will tend to trade off with the

thermal history sampling. Note also, it does not make sense to combine the two resampling strategies mentioned above (i.e., to resample both the ages and the errors at the same time).

We do not exhaustively test the resampling feature here but show an example of it in practice in the section Handling Complex Apatite (U-Th)/He Data in QTQt. Caution should be used when changing data uncertainties, especially when the change is to make them smaller (error scaling factor <1). In our example, we show that having narrow uncertainties on select data with no apparent reason to do so creates false structure in the model output and can create a much more complex cooling history that may be tempting to interpret but is, in fact, a consequence of the model attempting to fit overly precise data. QTQt lets the user define the range of the scaling; so the control of very precise data on the model can be readily assessed as needed.

## Summary

In the examples shown here, we use “perfect” AHe and AFT data to explore effects of model design on inversion results. Even with these “perfect” data, we are unable to retrieve all features of the “true” thermal histories. Some key t-T features that are less well resolved are rapid cooling through low temperatures (e.g., Path 3) and isothermal holding in the AHe PRZ (e.g., at 60 °C in Paths 3 and 4). In the second case, the inverse model results have more complexity (i.e., heating and cooling) than the true paths. These results reflect the non-uniqueness of thermochronometric data, even for large and multi-chronometer data sets. User decisions during the model set-up, such as imposing an initial constraint and setting the general ranges for prior, can influence the final form of the thermal history. Excessively broadening the prior will generally lead to simpler model solutions; however, the more robust the data set is (i.e., statistically significant AFT data and abundant single-grain AHe ages that can be explained by current radiation damage models) the less impact the prior has, because the data are able to better constrain the structure of the t-T path. The idea that the prior should not be too important, and therefore the data should dominate the solutions, is fundamental in the Bayesian approach. However, in all cases, it is important to clearly define the prior and initial constraints. Setting an initial constraint exerts a largely superficial influence on the model result so long as the constraint is not set to a time and temperature that the data are informing, and the user takes care to not over-interpret the part of the inversion result that is not informed by data or geologic information. In other words, so long as the initial constraint is old enough and/or hot enough, the data will still drive the solution.

An important take-away is that over-interpreting the expected thermal history (ExTH) path is perilous, but the 95% credible intervals provide useful guidance as to what features may or may not be resolvable from the data. The ExTH is an average of all accepted models, weighted by the posterior probability of the models. Therefore, it can be based on a variety of simple models (for example, protracted monotonic cooling) and also complex models, such

as those involving several episodes of heating and cooling. When averaged together as the ExTH model, this output may then show minor heating and cooling because the simple models smooth out the complexity. However, these portions of the thermal history typically have broad 95% credible intervals, which indicate that the path is not well constrained and should not be interpreted to have geological meaning. Also, estimates of maximum temperature during reheating in the ExTH tend to be skewed to lower values relative to the MaxLike or even MaxPost models. This is often manifested as predicted ages being somewhat older for the expected model (Fig. 9). Thus, the ExTH results should always be interrogated further by looking at the MaxLike and MaxPost models and their predictions in comparison to the observations, either via a graphical representation or through a statistical test conducted by the user (e.g., Gallagher, 2016), in order to determine how much complexity and what thermal path features or geologic events are necessary to fit the data. To further investigate structure in the ExTH, the model can be constrained using relevant geologic information and geochronologic data sets or by using forward modeling to test whether the data are able to resolve different hypotheses.

Overall, we encourage users to critically think about decisions regarding data, uncertainties, and any prior assumptions before initiating any modeling. Use QTQt to explore how each input will control the model design or model output by designing a set of sensitivity tests. If certain choices and assumptions do control the output, then the user should be clear and explicit about why certain decisions were made.

## ■ STRATEGIES FOR AND EXAMPLES OF INVERSE THERMAL HISTORY MODELING USING QTQt

The above examples of forward and inverse modeling illustrate challenges of modeling thermochronometric data sets using QTQt or any modeling program. Because thermochronologic data inherently yield non-unique thermal history solutions, a representative range of all possible t-T histories must be presented as either forward and/or inverse model results. Once the possible t-T histories have been identified, these results must be evaluated and interpreted in a geologic context. These challenges require that (1) a modeler understands what is controlling model behavior (e.g., input data, model design, and selected kinetics), and (2) the modeler determines which aspects of a thermal history are geologically possible and geologically meaningful.

It is critical to document the numerous decisions the modeler makes in the process of data collection, model design, model inputs, and model interpretation, in order to complete model sensitivity testing approaches (this study and Murray et al., 2022) and provide a testable model for future evaluation. This sensitivity testing and documenting requires an iterative approach that may combine both inverse and forward modeling strategies. In the following sections, we present examples of sensitivity testing and interpretation strategies for published thermochronologic data sets. These exercises are useful to develop an intuitive understanding for strategies that interrogate the relative

impact that model design and thermochronologic data have on inversion results—a critical consideration for all geologic interpretations based on model results. These scenarios include examples from vertical profiles, samples with data from multiple chronometric systems, and cases with complex AHe data that may be challenging to interpret along with the corresponding AFT data. Although the details of these examples are tailored to the algorithms and model outputs of QTQt, the strategies are transferable to studies that use other thermal history modeling programs. Additionally, we emphasize that there are multiple appropriate approaches to handling thermal history modeling results, and that the chosen method will depend on the data available as well as the geologic problem of interest.

## Sensitivity Testing: A Vertical Profile Sampling Strategy

One of the powerful attributes of QTQt is the ability to jointly model multiple samples with a known vertical relationship, a spatial context that can be leveraged to resolve more complete or complex thermal histories (Gallagher, 2012; cf. Stephenson et al., 2006). For example, samples taken in a long vertical transect from an exhuming fault block may be capable of resolving a larger magnitude of cooling compared to a single sample or samples spanning a smaller spatial (particularly vertical) range. The key assumption here is that a deeper (or lower elevation) sample was always hotter, or at least as hot as a sample that is shallower (higher elevation). In this case, having samples from different present or paleo-elevations or locations could represent a useful paleo-depth spectrum within the fault block (e.g., Gallagher, 2012; Abbey and Niemi, 2018). By modeling all samples together in a vertical profile defined by their estimated paleo-depths, there is a greater chance of recovering a better resolved thermal history, including the paleo-temperature gradient or even determining a rate of heating or cooling activity. QTQt uses the highest sample in a profile as a reference and a linear interpolation to determine thermal histories for the lower samples in the profile based on provided elevation or depth information and assigned prior temperature offset, which can be allowed to vary over time (Gallagher, 2012; Fig. 2; see screenshots of the user interface in the tutorial provided in the Supplemental Material, footnote 1). Here, we consider vertical profile modeling and the magnitude of cooling captured by different systems, different elevation relationships, and a combination of both. The question(s) that arise when building vertical profile models in QTQt include: How many samples are needed in a vertical profile? Which samples are the most important for obtaining different information about the thermal history? What might be missing information if a vertical profile cannot be collected? We illustrate a sensitivity-testing approach for evaluating these questions through thermal history modeling of vertical profiles with an example from the northern Rio Grande rift in Colorado (Abbey and Niemi, 2018). The data set we use is from a vertical profile collected on Mount Princeton in the upper Arkansas River valley. This profile includes samples with AHe, AFT, and ZHe data providing a wide temperature range to capture thermal events.

The original Mount Princeton vertical profile was made up of nine different samples collected by three different research teams (Kelley et al., 1992; Ricketts et al., 2016; and Abbey and Niemi, 2018). The collected samples were projected onto the exhuming fault plane, to obtain paleo-distance relationships (Abbey and Niemi, 2018). Of the nine samples collected, seven were analyzed for AHe (Ricketts et al., 2016; Abbey and Niemi, 2018). Although this profile also has ZHe and AFT data, here we focus solely on the questions related to how many and which samples are the most useful in a vertical profile, and therefore for simplicity, we model only the AHe data in this section (Fig. 10).

Modeling all seven AHe samples together, with no other constraints, and using the same model design as was used by the original study (Table S3, see footnote 1; Abbey and Niemi, 2018) produces an ExTH model for all samples in the profile. Because no initial constraint box was used, the ExTH begins where the data begin to provide information for the model and implies rapid cooling prior to this time (Table 1). The thermal history begins at temperatures hotter than the AHe PRZ and shows relatively slow cooling until ca. 4 Ma, at which point, the cooling rate increases (Fig. 10A; Table 1). QTQt also predicts an AHe closure time with these models. This is an estimate of the last time a sample passed through the base of the PRZ and was not subsequently reset. For the run with all the samples, the predicted closure times range from ca. 14 Ma for the higher samples to ca. 4 Ma for the lower samples in the profile (Fig. 10A; Table 1). Comparing the model predicted ages to the measured ages shows that the spread in ages observed in the higher samples of the profile is not predicted and that observed ages skew slightly higher or lower than the predictions for the samples higher in the profile (Fig. 10A).

We leverage this sample-rich vertical profile to interrogate the potential effects of sampling on the timing and precision of cooling identified in the inversion. Specifically, we use sensitivity testing that omits key samples in this transect to determine if specific samples or number of samples are controlling the model output (Table S3, see footnote 1). It is also worth noting that QTQt offers the option to include all samples but turn some off as desired. In this case, QTQt provides the predicted ages for all the samples, but the turned-off samples are not used in the inversion (i.e., they do not influence the thermal histories).

First, we test the removal of the lowest sample in the profile. Removing the lowest sample shifts the model start time to begin earlier and at cooler temperatures (Fig. 10B; Table 1). Thus, all the samples are beginning in and spanning the entire range of the PRZ for AHe. Without the lowest sample, the output shows no slower cooling prior to the 4 Ma cooling event, and instead shows the samples all undergoing isothermal holding from model start to final cooling at 4 Ma. The final cooling phase, beginning at 4 Ma, occurs at the same rate as the original vertical profile output containing all of the AHe samples. The predicted closure time shifts and the predicted ages for the higher samples cover a narrower range and do not capture the spread of ages in the highest three samples (Fig. 10B; Table 1).

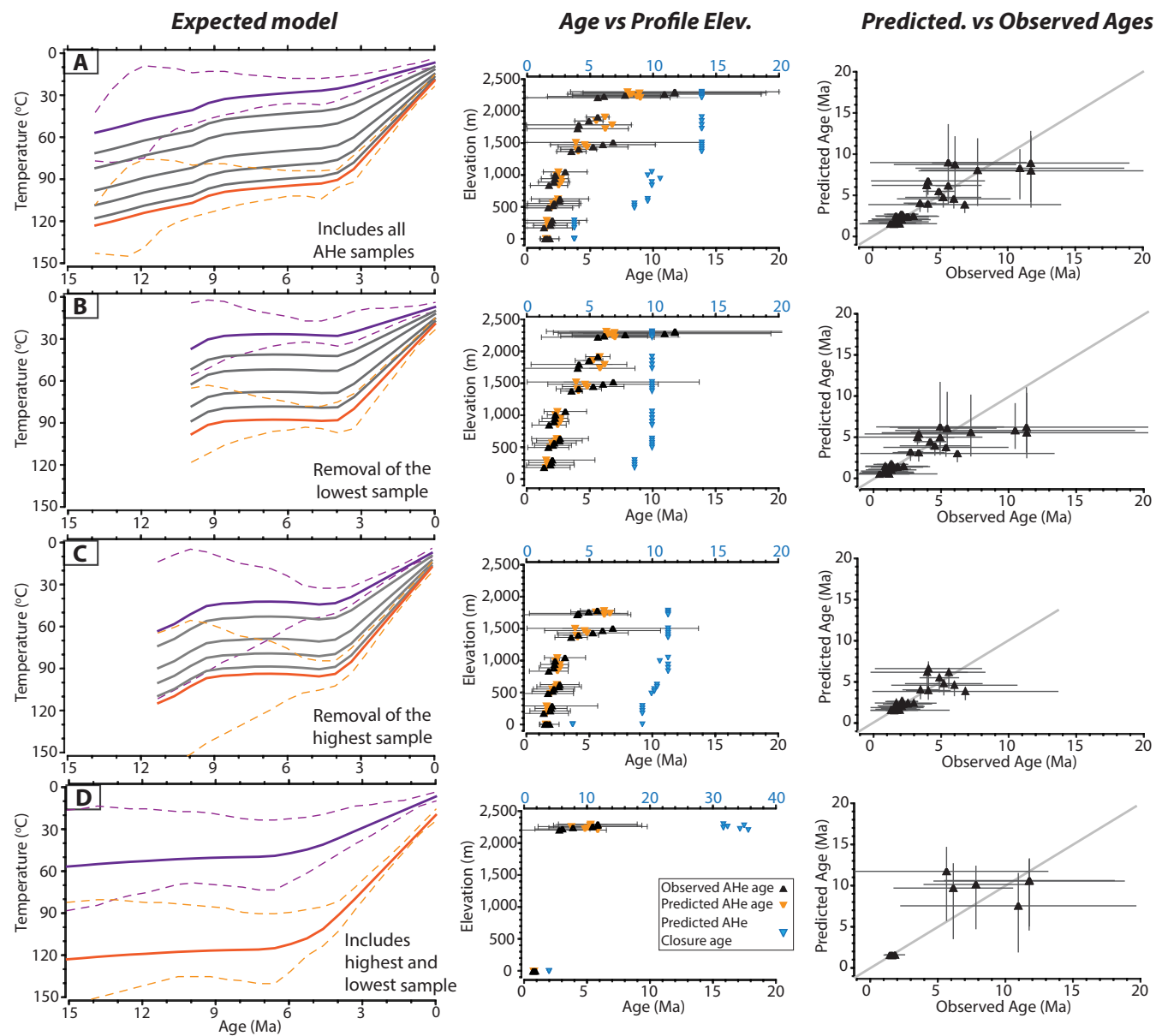
Removing the highest sample produces an ExTH output that begins at a similar time to both the original profile and the test without the lowest sample

(Fig. 10C; Table 1). The initiation temperature range is the same as the original profile (Fig. 10A; Table 1). From the model start until ca. 9 Ma, there may be a possible cooling event at a rate of  $\sim 10$  °C/m.y., although the 95% confidence envelope is wide over this timeframe (Fig. 10C; Table 1). The model then shows the samples are isothermal until 4 Ma, when the final cooling begins. The predicted closure times range from ca. 11 Ma to ca. 4 Ma for the higher and lower samples, respectively (Table 1). The predicted ages are similar to the observed ages with the two higher samples having a slightly wider range in predictions (more like the observations) compared to the predictions when the top sample was included (Fig. 10A).

What if only the top and bottom samples were collected and analyzed? In other words, can one get these results from only collecting a few samples? That would reduce the need to collect the samples at the mid-elevations. In this case, the output model initiates much further back in time than the models that include the mid-profile samples (Figs. 10A,D; Table 1). Both the samples begin at temperatures greater than the PRZ for AHe with relatively slow cooling from initiation to ca. 5 Ma (Fig. 10D; Table 1). However, the resolution on the thermal history is not very precise from model initiation to ca. 10 Ma, when we see narrowing of the 95% confidence interval (Fig. 10D). The final cooling event, which appears to be slightly earlier, compared to the models from the other tests, takes the samples to surface temperatures at a rate of 25 °C/m.y. from 5 Ma to the present. The predicted closure ages shift to much older for the highest sample and younger for the lowest sample compared to the model runs that included the mid-profile samples (Figs. 10A,B,C; Table 1). The predicted ages are much more in line with the observed ages, with more age variability in the top sample and very reproducible ages in the bottom sample (Fig. 10D).

## Summary

A sampling strategy that includes collecting samples in a vertical or semi-vertical profile is an effective way to obtain more information about timing and rates of cooling using QTQt. In general, we expect all samples to have experienced more or less the same form of thermal history (e.g., timing of cooling events), and the paleo-temperature gradient can be either set to be constant or vary with time. However, as sensitivity testing shows, the resolution of the thermal history is dependent on the number of samples in the data set and quality of that data. Determining which samples to collect, how many, and what the distance between samples should be, will depend on the questions being asked, the types of analyses being done, and the availability or accessibility of material. Because we often do not have much control on those aspects, we need to do the best we can with the available data. The types of sensitivity tests we perform here help users to see how the output models may be affected by the number of samples included in a vertical profile. For example, removing the lowest sample reduces the information used to constrain the thermal history at higher temperatures that are inherently associated with an older part of the thermal history. Without the highest sample,



**Figure 10.** Left: ExTH model outputs and data predictions for vertical profile sensitivity tests. Middle: data predictions plotted against elevation (Elev.). Right: 1–1 data predictions. All apatite (U-Th-Sm)/He (AHe) ages are uncorrected ages. All runs include only AHe data, and tests show the results of removing the highest (B), lowest (C), and middle (D) samples within the vertical profile. When modeling a profile in QTQt, all the samples are input in the same run and linked by the information about the distance between each sample and the paleo-geothermal gradient.



TABLE 1. RESULTS AND INTERPRETATIONS FROM VERTICAL PROFILE SENSITIVITY TESTS

Model run name	Model initiation time (ca. Ma)*	Initial temperature range (~°C) <sup>†</sup>	Thermal history cooling rates (~°C/m.y.) and time (ca. Ma)	Position, number of samples, and closure time (ca. Ma)
Original vertical profile (seven samples)	14	60–130	3; 13–4 22; 4–0	Top three samples: 14 Middle two samples: 11–8 Bottom two samples: 4–3
Lowest sample removed (six samples)	10	35–90	0; 10–4 22; 4–0	Top five samples: 10 Bottom sample: <sup>§</sup> 8.5
Highest sample removed (six samples)	11.5	60–120	10; 11.5–9 0; 9–4 25; 4–0	Top three samples: 11 Middle two samples: 10–9 Bottom sample: 9–4
Highest sample removed (six samples)	41	55–125	1; 41–5 25; 5–0	Top sample: 36–31 Bottom sample: 3–2

\*Because we do not include an initial constraint box, the thermal history model begins at the time at which the input data begin to inform the model. This implies rapid cooling prior to model initiation.

<sup>†</sup>This temperature range is determined from the starting temperature of the highest and lowest samples, respectively.

<sup>§</sup>This is not the lowest sample in the original transect.

the model fails to precisely constrain the retention of higher elevations at relatively low temperatures before ca. 6 Ma, a key component of the geological interpretation. By only using the top and bottom samples in the profile, the resolution of the entire range of modeled thermal history across temperatures is sacrificed, thus making it difficult to constrain the timing or magnitude of the cooling event central to this study (Fig. 10).

QTQt provides the advantage of modeling multiple samples together, capitalizing on the temperature variability across a range of paleo-depths. Vertical profiles can be useful at capturing longer cooling histories or documenting paleo-PRZs and paleo-PAZs, when for example there is a large temperature offset due to large sample distances from high-relief settings. However, the functionality of multi-sample modeling is useful even with little to no vertical (and implied temperature) gradient. Thus, jointly modeling samples collected in a low-relief setting can be useful for finding thermal histories that fit all the data (e.g., House et al., 2001; Gallagher, 2012).

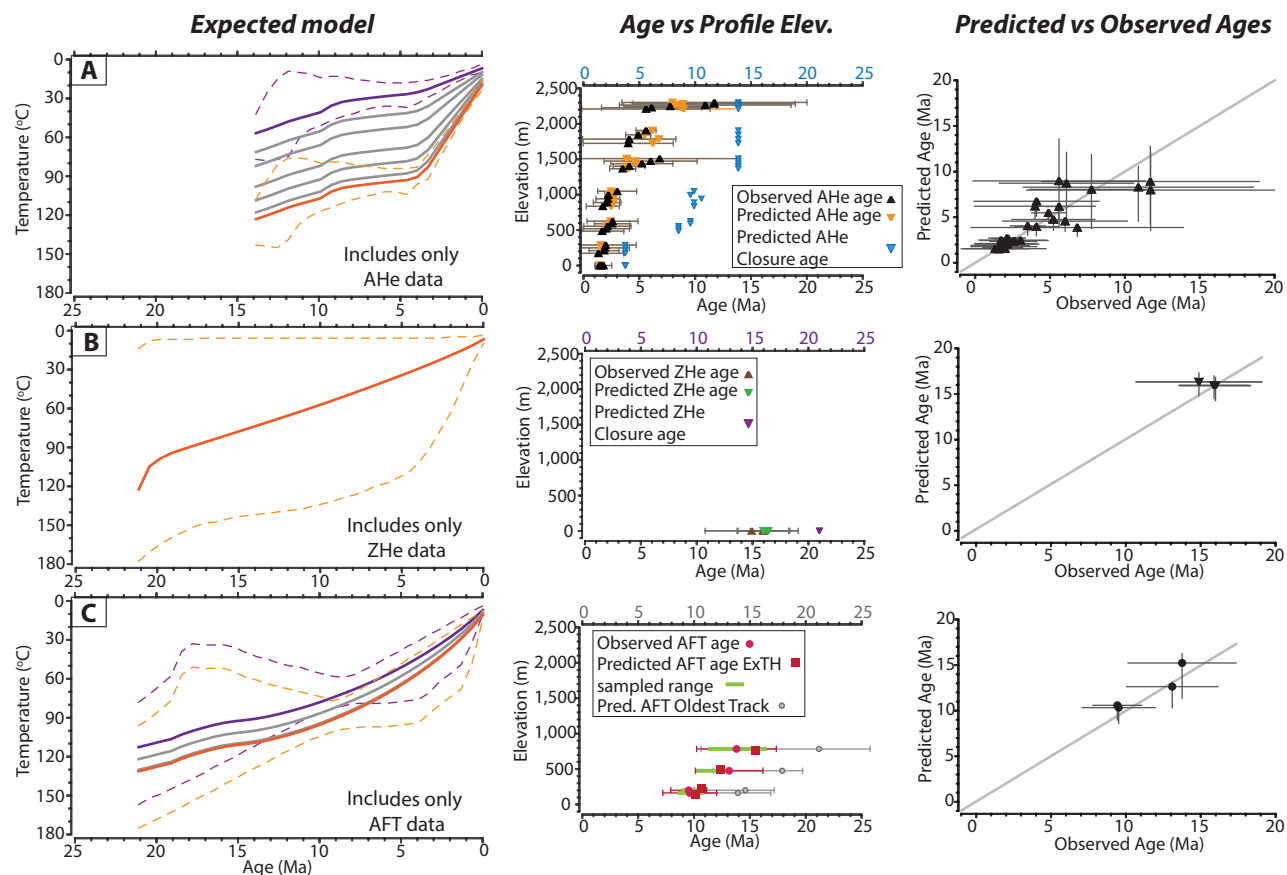
### Sensitivity Testing: The Use of Multiple Chronometers

Another way to complement the vertical profile method is to increase the number and/or types of thermochronometric analysis done on each individual sample. Similar to vertical profile sampling strategies, using multiple thermochronometers can increase the range of temperatures covered and can provide more information about the timing and rates of cooling events. Here we use the same example from the northern Rio Grande rift in Colorado, but instead of focusing on one data type and taking out samples based on location in the vertical profile, we perform sensitivity tests by selecting samples analyzed using multiple thermochronometric methods. In addition to the seven AHe samples, a sample was analyzed for ZHe (lowest in modern elevation and

deepest in paleo-distance space; Abbey and Niemi, 2018), and four samples in the middle of the vertical profile were analyzed for AFT (Kelley et al., 1992). The power of these tests helps users to determine when and where multiple thermochronometers serve as an effective strategy to extract thermal information to better inform a specific geologic question.

First, we compare model outputs for each of the individual thermochronometric methods used in the sample set (Fig. 11; Table S4, footnote 1). Although the number of samples analyzed using each method is different, and the elevation spanned by those samples is also different, the chronometric-specific outputs provides a baseline for identifying how different data types and combinations control the output model. The seven samples analyzed for AHe imply information about temperatures higher than the AHe PRZ (>90 °C) because of the determined paleo-depth relationships (Fig. 11A; Abbey and Niemi, 2018). Cooling appears to begin gradually and then increases to rapid cooling from 4 Ma to present (Fig. 11A; Table 2). When modeling only the ZHe data, which come solely from the lowest elevation sample, we see protracted monotonic cooling from model initiation to present (Fig. 11B; Table 2). When we model only the AFT data, which all come from four mid-profile samples closer to the lower elevations in the profile, we see that a distinct cooling event is not well captured, although there may be a slight increase in cooling rate at ca. 11 Ma (Fig. 11C; Table 2). We note that the credible intervals are significantly wider prior to 15–10 Ma, implying lack of resolution prior to any cooling around that time.

Combining chronometers either from the same sample or from other samples that geologically make sense to be modeled together (e.g., spatial relationships) adds more data and potentially more detail (Table S5, see footnote 1). When combining all of the AHe and AFT samples together, we see a cooling event beginning at 4 Ma with different cooling rates depending on sample location within the vertical profile (Fig. 12A; Table 3). The model output does



**Figure 11.** Left: ExTH model outputs for single-chronometer sensitivity tests. Middle: data predictions plotted against elevation. Right: 1–1 data predictions. All apatite (U-Th-Sm)/He (AHe) ages are uncorrected ages. The original data set (Abbey and Niemi, 2018) included one sample with ZHe and AHe data, two samples with AFT and AHe data, two samples with only AFT data, and four samples with only AHe data. These tests show the modeling results of using only one type of chronometer in each run, thus there are four samples in the AFT only run (C), one sample in the zircon (U-Th-Sm)/He (ZHe) only run (B), and seven samples in the AHe only run (A).

not give much more information besides an indication that the lower samples may have been as hot as ~130 °C ca. 20 Ma, and more rapid cooling began at ca. 4 Ma.

When AHe and ZHe are combined, the model begins earlier and at hotter temperatures (Fig. 12B; Table 3). However, this model is only able to capture the 4 Ma cooling event with any detail and confidence and predicts the simple solution of monotonic cooling prior to 4 Ma (Fig. 12B; Table 3).

Combining ZHe and AFT, without AHe, loses the resolution on the 4 Ma cooling event (Fig. 12C), and the output is similar to the model with only AFT

data (Fig. 11C), although the possible timing of cooling events shifts. Modeling ZHe and AFT data may also identify an additional cooling trend from model initiation to ca. 20 Ma, followed by residence within temperatures of the AFT PAZ until ca. 8 Ma (Fig. 12C; Table 3). At 8 Ma, the AFT and ZHe model shows that there is perhaps another cooling event, although, as with the AFT only model, the initiation is broad and difficult to pinpoint.

A combination of all three chronometers (ZHe, AFT, and AHe) illustrates the most constrained ExTH model output including two distinct cooling events (Fig. 12D). The first cooling event occurs from ca. 24 Ma to ca. 17 Ma, followed

TABLE 2. RESULTS AND INTERPRETATIONS FROM SINGLE-CHRONOMETER SENSITIVITY TESTS

Model run name	Model initiation time (ca. Ma)*	Initial temperature range (~°C)†	Thermal history cooling rates (~°C/m.y.) and time (ca. Ma)	Position, number of samples, and closure time (ca. Ma)
Only AHe data (seven samples)	14	60–130	3; 14–4 20; 4–0	Top three samples: 14 Middle two samples: 11–8 Lowest two samples: 4–3
Only ZHe data (one sample)	21	120	4; 21–0	Lowest sample: 21
Only AFT data (four samples)	21	100–130	3; 21–11 9; 11–0	Top sample: 21 Middle sample: 18 Lowest two samples: 15–14

\*Because we do not include an initial constraint box, the thermal history model begins at the time at which the input data begin to inform the model. This implies rapid cooling prior to model initiation.

†This range is determined from the starting temperature of the highest and lowest samples, respectively.

Note: AFT—apatite fission-track; AHe—apatite (U-Th-Sm)/He; ZHe—zircon (U-Th-Sm)/He.

by isothermal holding until 4 Ma. The highest samples remain at temperatures below the retention and annealing zone temperatures of these mineral systems, while the lower samples in the profile are kept in the apatite PRZ and PAZ. The final cooling event begins at 4 Ma with samples cooling at different rates, depending on sample location in the vertical profile (Fig. 12D).

### Summary

Using different chronometers increases the temperature range that modeling programs such as QTQt or HeFTy can use to assess possible thermal histories. In this example, we used three different chronometers with a

combined temperature sensitivity range from 30 °C to 220 °C (e.g., Gleadow et al., 1986; Shuster et al., 2006; Guenther et al., 2013; Ault et al., 2019). Modeling a multi-chronometer data set with expanded thermal sensitivity extends the temporal resolution of the thermal history. In this example, the AHe data serve a critical role in clearly resolving the cooling event at ca. 4 Ma; however, the addition of AFT and ZHe data sets extends the thermal information. In particular, the addition of AFT data in this example is critical in identifying the timing and rate of an early Miocene cooling event. This suite of sensitivity tests on a geologic case study serves as a useful way to develop an intuition for identifying the control of thermochronologic data on output models and evaluating if models may be oversimplified without additional data or geologic information. Such tests also help users developing a new project to:

TABLE 3. RESULTS AND INTERPRETATIONS FROM MULTI-CHRONOMETER SENSITIVITY TESTS

Model run name	Model initiation time (ca. Ma)*	Initial temperature range (~°C)†	Thermal history cooling rates (~°C/m.y.) and time (ca. Ma)	Sample type and closure time (ca. Ma)
AFT and AHe data (nine samples)	21	70–140	2; 21–4 5–20; 4–0	AHe: 21–11 AFT: 21–16
AHe and ZHe data (five samples)	36	130–200	4; 36–4 25; 4–0	AHe: 19–4 ZHe: 28–27
AFT and ZHe data (five samples)	26	150–170	7; 26–20 5; 20–8 10; 8–0	ZHe: 24 AFT: 20–15
AFT, AHe, and ZHe data (nine samples)	31	140–210	11; 24–17 0; 17–4 8–25; 4–0	AHe: 21–4 ZHe: 23 AFT: 20–17

\*Because we do not include an initial constraint box, the thermal history model begins at the time at which the input data begin to inform the model. This implies rapid cooling prior to model initiation.

†This temperature range is determined from the starting temperature of the highest and lowest samples, respectively.

Note: AFT—apatite fission-track; AHe—apatite (U-Th-Sm)/He; ZHe—zircon (U-Th-Sm)/He.

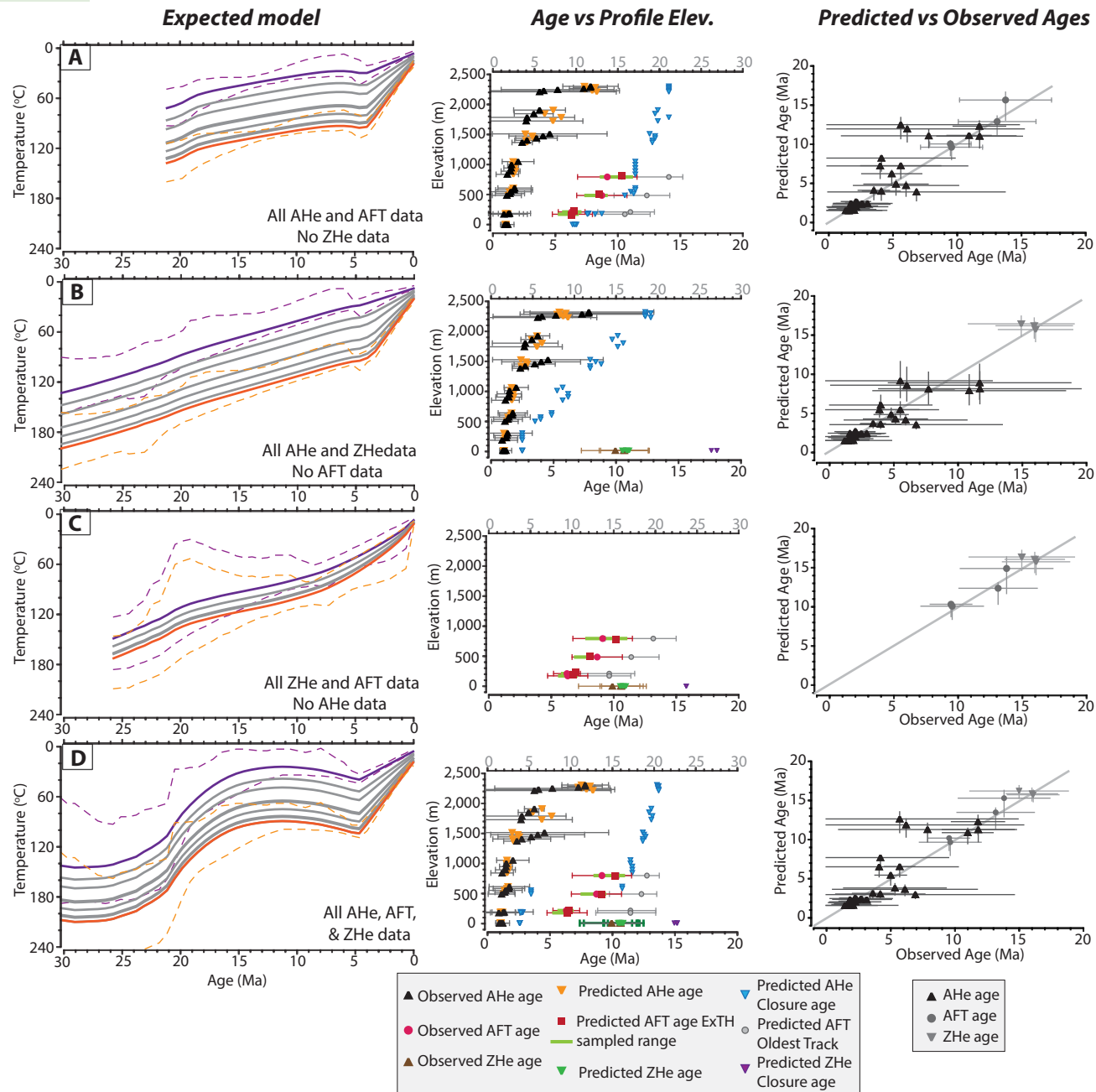


Figure 12. Left: ExTH model outputs and data predictions for multi-chronometer sensitivity tests. Middle: data predictions plotted against elevation (Elev.). Right: 1–1 data predictions. All apatite (U-Th-Sm)/He (AHe) and zircon (U-Th)/He (ZHe) ages are uncorrected ages. Models include all samples from the original data set (Abbey and Niemi, 2018) with each kind of data (AHe, ZHe, and AFT). Tests show the results of using a combination of different chronometers.

(1) collect multiple samples with a known spatial relationship (e.g., vertical profiles); (2) collect fewer samples but incorporate multiple dating techniques; or (3) combine a vertical profile sampling strategy with a multi-chronometer analytical strategy.

### Sensitivity Testing: Handling Complex AFT-AHe Data Sets

In this example, we use data from a single sample within the data set published by Wildman et al. (2016) from the southwest African continental margin. The study reports apatite fission-track and single-grain AHe data from Neoproterozoic basement rock at the high-relief transition zone between the low-relief, low-lying, coastal plain and the low-relief, elevated, continental interior. The data presented by Wildman et al. (2016) are challenging to model because the inferred thermal history does not simultaneously fit the AFT and single-grain AHe age data, and there is no robust independent geological information to constrain the Mesozoic and Cenozoic thermal history. Here, we attribute the complex relationship between the AHe ages and AFT data to the dispersion in the AHe ages and note the limited ability of current He diffusion models to reproduce this dispersion. This phenomenon has been observed elsewhere, particularly at ancient passive continental margins, cratons, and old orogen settings (e.g., Hendriks and Redfield, 2005; Kohn et al., 2009; Flowers and Kelley, 2011; McKeon et al., 2014; Wildman et al., 2016; Morón et al., 2020), and its causes have been thoroughly discussed (Brown et al., 2013; McDannell et al., 2018; Ault et al., 2019; Fox and Carter, 2020). Although the causes of AHe age variability are well appreciated (predominantly radiation damage and grain size), observed ages are often not fully explained with commonly used proxies for these causes, and, in this case, some grains are considered “incompatible” with current diffusion models. This scenario invites consideration into what is an appropriate uncertainty to assign to the single-grain AHe ages. It should also be recognized that the fission-track annealing model is unlikely to be perfect, and developments in our understanding of the contribution of non-thermal annealing phenomena, track etching efficiency, and chemical substitutions on annealing kinetics will help to refine and improve these annealing models (Ketcham, 2019).

The sample chosen for our example consists of AFT and AHe data. The AFT data include single-grain count data, horizontal confined track lengths with angles relative to c-axis for anisotropic annealing corrections (e.g., Ketcham et al., 2007), and a sample mean  $D_{par}$  to account for compositional effects on track annealing (Table S6, see footnote 1). The AHe data consist of four two-termination single grains of varying geometries and eU contents; the original data from Wildman et al. (2016) included several one-termination grains, which have been excluded for this example. The AHe ages appear to have a strong positive relationship with equivalent spherical grain radius ( $ESR = (3 * volume)/surface\ area$ , based on the true 3D geometry of the measured grain) consistent with larger grains having higher closure temperatures (e.g., Reiners and Farley, 2001), but the single-grain ages show a strong negative correlation

with eU, a proxy used for the amount of radiation damage experienced by the grain. This negative correlation contradicts the age-eU relationship predicted by radiation damage and annealing models (e.g., Flowers et al., 2009; Gautheron et al., 2009). However, Brown et al. (2013) show that competing factors change the AHe age, and simple bi-variate correlations may not be easily observable, particularly when interpreting a small number of single grains, as is the case in our example. Moreover, negative age-eU relationships are observed in ZHe data sets (Guenther et al., 2013; Baughman and Flowers, 2020), albeit these grains achieve much higher alpha doses, and investigations into radiation damage accumulation and annealing in apatite and individual-grain diffusion kinetics more generally is ongoing (Gerin et al., 2017; Willett et al., 2017; Guo et al., 2021). Our first decision is, therefore, not to exclude any single-grain AHe ages and instead aim to find a model that is mutually consistent with the observed AHe and AFT and the uncertainty of these techniques while evaluating to what extent a particular component of the data, and our model input decisions, influence the output and interpretation.

### Modeling Individual Data Sets

We initially explore what thermal history information is held in each component of the data set (i.e., a single AHe age or AFT data) without any additional constraints to assess the compatibility between the individual data sets (Table S6, footnote 1). The ExTH generated using only the AFT data infers a moderate rate of cooling through the PAZ between 130–80 Ma, which then progressively slows from 80 Ma to present through temperatures <60 °C (Fig. 13). For the AHe data, we first use the analytical error ( $1\sigma = 1\%$ ) as our uncertainty on the observed age and run an inverse model for each single grain (e.g., Sousa and Farley, 2020), as well as for all grains combined together as a complete AHe data set from a single sample. The ExTH model infers that each grain should initially cool through its specific PRZ at different times spanning from ca. 215–115 Ma, and each AHe age is reproduced well (Fig. 13). Combining all the AHe ages from the sample, the ExTH infers rapid cooling between 140 and 135 Ma and residence below 30 °C from 120 Ma until present day. Cooling through ~60–30 °C between 140 and 120 Ma appears to be well constrained with a tight band defined by the 95% credible intervals (Fig. 13). Although a model combining all grains is more appropriate given each grain is from the same rock, the ExTH has compromised on the data-fit with no single-grain age being reproduced well.

Our next decision addresses the uncertainty that is assigned to the helium ages. We initially used the analytical error. However, given that the data show unexplained dispersion, this is likely an under-estimate of the uncertainty we should allow for when modeling the AHe ages. Wildman et al. (2016) add an additional 10% of the measured age, which is the standard deviation (reproducibility) of the AHe ages of Durango apatite standards measured during their analytical sessions (Table S6, see footnote 1). The consequence of this decision is that all of the AHe models have broader 95% credible intervals



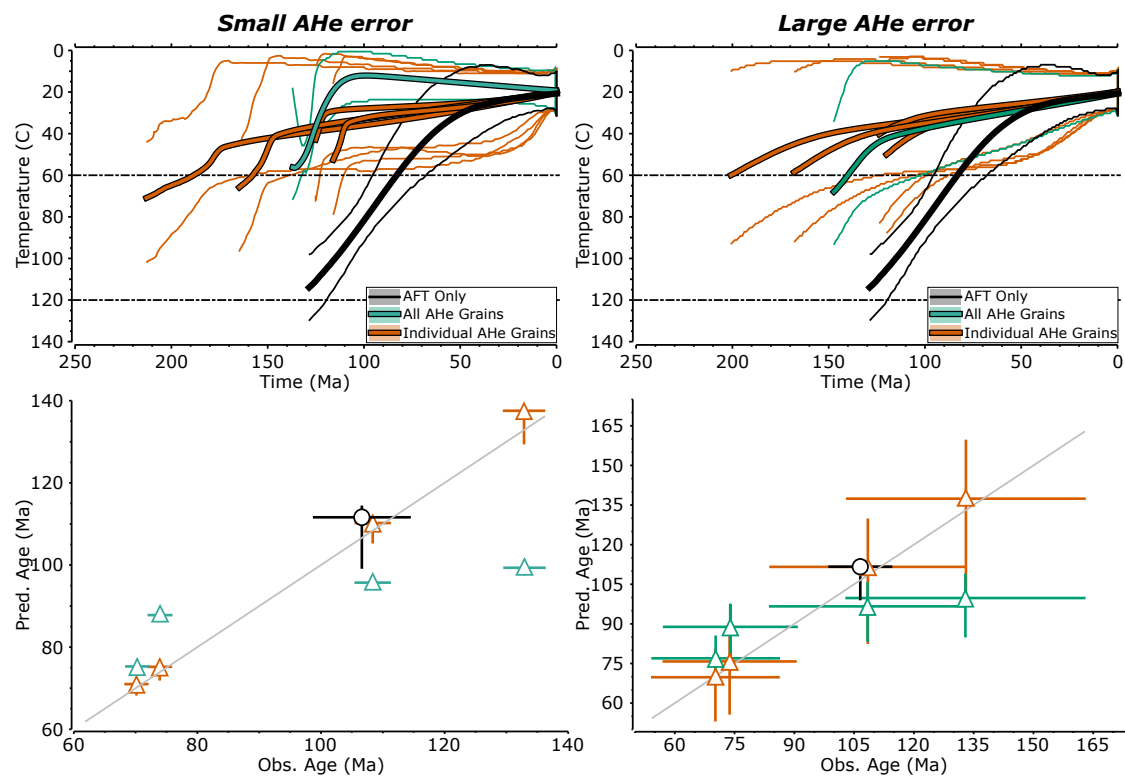
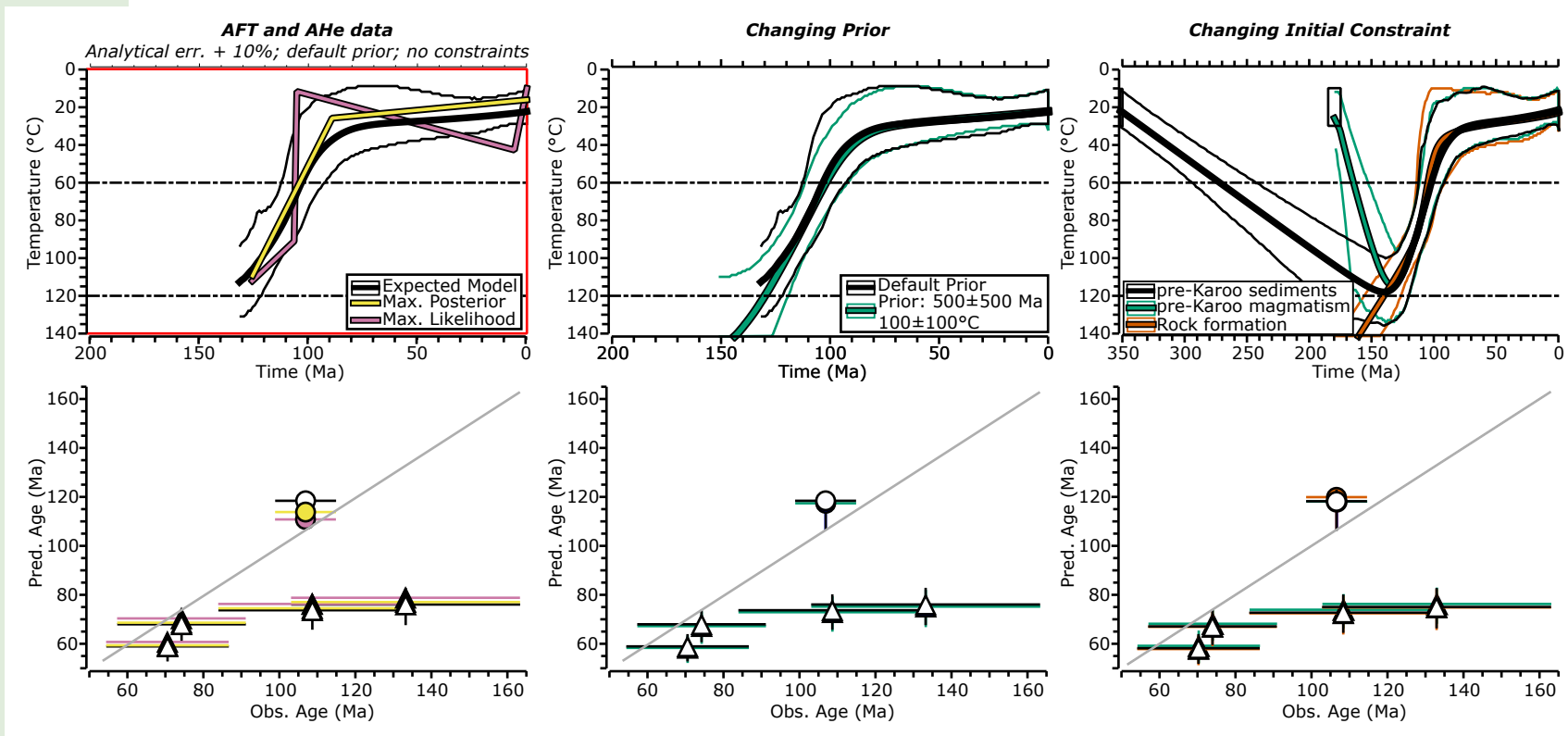


Figure 13. Top: ExTH results from inverting (1) Apatite fission-track (AFT) data only (black); (2) all AHe grains together as a single sample (green), and (3) each apatite (U-Th-Sm)/He (AHe) age individually (orange). Left plots show the results when a small error (1%) is assigned to the AHe ages. Right plots show the results when a larger error (11%) is assigned to the AHe ages. Bottom: Data predictions for inverse model results for (1), (2) and (3). Error bar on the y-axis (i.e., uncertainty on the predicted age) is the range of sampled predicted ages for the models within the 95% credible intervals. Predicted (Pred.) and observed (Obs.) ages are central AFT ages and uncorrected AHe ages.

and more protracted cooling through temperatures lower than  $\sim 70^\circ\text{C}$  (Fig. 13). Although the models for individual grains still imply quite disparate timings for initial cooling, there is greater overlap between the credible intervals. The ExTH that combines all of the AHe grains similarly shows broader 95% credible intervals and slower cooling, but, with the larger uncertainty, three out of four grains are reproducible at  $2\sigma$ , compared to zero out of four when only the analytical error is used. The conclusion from this series of tests is that when inverting the AHe data only, using the Flowers et al. (2009) RDAAM, a single thermal history capable of reproducing the observed AHe ages at  $2\sigma$  cannot be found. Given what is observed in our tests (Figs. 3 and 6), it is also likely that inverting the AFT and AHe data together will not produce a thermal history capable of reproducing all components of both data sets. This is problematic because all the apatite crystals were from the same granitic-gneiss host rock, and therefore their ages should be predictable with a common history. This will only be the case if our measurements are precise, and our annealing and diffusion numerical models are entirely representative of the natural processes over geologic timescales.

### Modeling Joint Data Sets

Following our observations from the individual data set modeling, we make the decision whether to exclude the AHe grains that yield a thermal history that is the most inconsistent with thermal histories produced by the other AHe grains—or are most poorly reproduced when the AHe grains are modeled together—or we continue to include all components of the data set and search for a mutually consistent thermal history within the scope of the uncertainty on each component of the data set. Here, we follow the approach of Wildman et al. (2016) and do not exclude any data (Table S6, footnote 1) and obtain the thermal history and data predictions shown in Figure 14. We observe that cooling through the base of the PAZ starts just before 130 Ma and continues at a rate of  $1.75^\circ\text{C}/\text{m.y.}$  until 90 Ma, at which point the cooling rate slows, and temperatures are  $<40^\circ\text{C}$  through to 0 Ma. As anticipated, the ExTH does not produce a satisfying visual data fit for all components of the data. Two young AHe ages are reproduced with  $2\sigma$ , and the AFT is almost reproduced at  $2\sigma$ . The MaxPost and MaxLike models are shown in Figure 14



**Figure 14.** Models results (top) and data predictions (bottom) from the joint inversion of apatite fission-track (AFT) and apatite (U-Th-Sm)/He (AHe) single-grain data. Predicted (Pred.) and observed (Obs.) ages are central AFT ages and uncorrected AHe ages. Left: Model set-up based on no initial constraints, using the default “General ranges for prior” and a large uncertainty on the AHe ages reflecting analytical uncertainty plus an additional 10% uncertainty to reflect standard reproducibility and uncertainty in the method (i.e., radiation damage effects). Black—ExTH (with 95% credible intervals); yellow—MaxPost model; pink—MaxLike model. Middle: Black—ExTH (with 95% credible intervals) for a model with the same initial conditions as “Left.” Green—ExTH (with 95% credible intervals) for a model that has been given a larger time range of  $500 \pm 500$  Ma and larger temperature range of  $100 \pm 100$  °C. Right: Black—ExTH (with 95% credible intervals) with the same initial conditions as “Left” and an additional constraint at  $(355 \pm 5$  Ma,  $20 \pm 10$  °C). Green—ExTH (with 95% credible intervals) with the same initial conditions as “Left” and an additional constraint at  $(1000 \pm 100$  Ma,  $1000 \pm 100$  °C). Orange—ExTH (with 95% credible intervals) with the same initial conditions as “Left” and an additional constraint at  $(1000 \pm 100$  Ma,  $1000 \pm 100$  °C).

to highlight that within the range of sampled thermal histories, from which the ExTH model is determined, there are individual models (both simple and complex) that do reproduce the AFT age within  $2\sigma$ . The older AHe ages are not well reproduced in any of the models, and, given what was observed in the previous series of tests on the individual data sets, it can be said that the form of the thermal history is being controlled by the AFT data and the two younger AHe ages.

Next, we test the effect of our assumption on the general range for the prior and our choice of initial constraint (or lack thereof) has on our expected thermal history (Table S6, footnote 1). Our default prior was a t-T space defined by the

oldest observed age  $\pm$  oldest observed age and  $70 \pm 70$  °C. As shown above, increasing the size of the prior will tend to produce simpler thermal histories, and therefore, we explored the effect of using a large prior of  $500 \pm 500$  Ma and  $100 \pm 100$  °C. We observe that this change has a negligible effect on our thermal history with both models yielding similar ExTHs and data predictions, suggesting that the data are dominating the inference of the thermal history solutions. Using the default prior, we then test different initial constraints and compare this model to our initial “unconstrained” model. We test constraints that reflect: (1) the basement being at the surface immediately before deposition of the Karoo Supergroup sedimentary rocks ( $355 \pm 5$  Ma,  $20 \pm 10$  °C);

(2) the basement being at the surface immediately before emplacement of lavas associated with Karoo magmatism ( $180 \pm 5$  Ma,  $20 \pm 10$  °C); and (3) crystallization of the basement rock ( $1000 \pm 100$  Ma,  $1000 \pm 100$  °C). Regardless of the constraint, we observe that all models cool through the base of the PAZ at the same time (ca. 130 Ma) and experience the same cooling history between 130–0 Ma as the sample passes through 110–20 °C (Fig. 14). Because the sample experienced temperatures greater than the range of temperature sensitivities for the AFT and AHe systems immediately prior to 130 Ma, the thermal history before this time is not informed by the observed data and is entirely dependent on the user's decision for the initial constraint.

The ExTH shown in Figure 14 is our preferred model because (1) it includes all AFT and AHe ages, and the model and model uncertainty (i.e., 95% credible intervals) is a product of the complex observed data, their analytical uncertainties, and uncertainties from the annealing and diffusion models used to reproduce the data; (2) it is solely driven by the data in the absence of any well-constrained geological constraints, and it does not include speculative constraints before the time of cooling through the base of the PAZ, which the data provide no information about; (3) the thermal history reproduces the AFT age, TLD, and two of the AHe ages well and identifies that the older AHe ages may be problematic and cannot be explained with the diffusion model used; (4) the thermal history obtained is consistent with the geologic model of the region, where cooling follows the rifting and break-up of southwest Africa from eastern South America as observed previously along this margin and along other passive continental margins (Wildman et al., 2019a). A different user of this data set may have made alternative decisions in data handling and screening, model set-up, and in testing different geologic scenarios (i.e., by imposing different constraint boxes). Similarly, another user may have decided to use and interpret a different model output (e.g., the posterior model), or use the information learned from the inverse modeling approach to propose a forward model they feel better fits the geological history and observed data. Regardless of what these decisions were, they should be discussed in terms of their influence on the thermal history chosen for interpretation, and the satisfactory and unsatisfactory parts of the model should be acknowledged. In our case, this is clearly the poor data fit to two out of four of our AHe ages.

### *Handling Complex Apatite (U-Th)/He Data in QTQt*

QTQt offers tools to explore the influence of these potentially erroneous data on the thermal history and to explore alternative diffusion models that may better reproduce the observed data (Table S6, footnote 1). If we decide that our AHe ages are complex and likely to be challenging to explain with current diffusion models, we can choose not to fit the observed age exactly and instead resample the observed He age using random sampling of a normal distribution, centered on the observed age, with a standard deviation equal to the uncertainty on the observed age. In doing this, we implicitly allow for more uncertainty elsewhere (e.g., in the diffusion model) and compare our

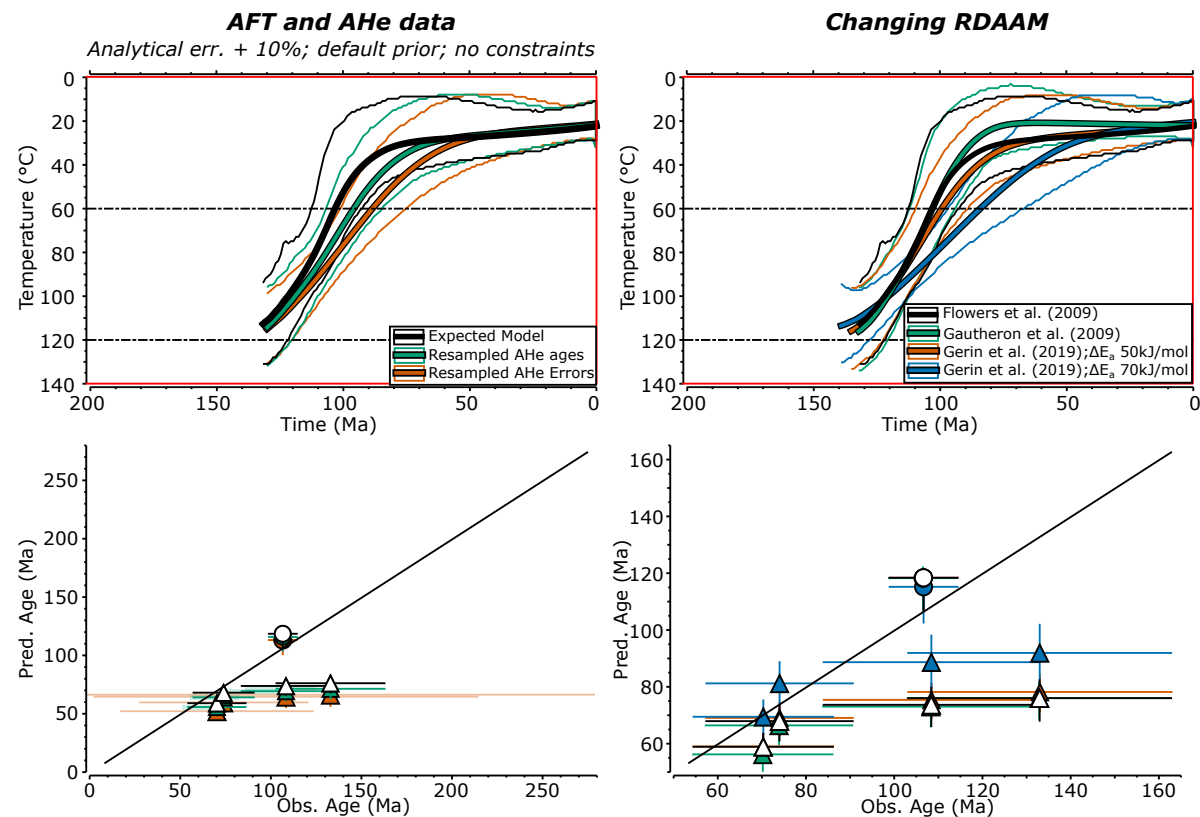
predicted ages to a sampled estimate of the true AHe age when deciding to accept a given thermal history model. Alternatively, we can resample the error we have assigned to our He ages such that AHe data that are reproduced well by the thermal history will have a small error scaling, whereas the AHe data that are difficult to reproduce (e.g., outliers, and AHe ages seemingly incompatible with chosen diffusion model) have a larger error scaling. In other words, the data with lower error scaling will tend to dominate the solution, while data with the higher error scaling have reduced importance.

We observe that applying either of these options causes the AHe data overall to have less influence on the thermal history, and the ExTH is controlled by the AFT data (Fig. 15). If we decide that our observed ages are a good estimate of the age, we could explore alternative radiation damage models that have been proposed (e.g., Gautheron et al., 2009; Gerin et al., 2017; Willett et al., 2017). In this case, we note that the Gautheron et al. (2009) model and Gerin et al. (2017) model (with  $D_0 = 3 \times 10^{-7}$  m<sup>2</sup>/s, activation energy in defect-free crystal,  $E_a = 100$  kJ/mol, additional energy to escape a void,  $\Delta E_a = 50$  kJ/mol) produce similar thermal histories, and differences to the data fit are very small (Fig. 15). However, for the Gerin et al. (2017) model with a  $\Delta E_a = 70$  kJ/mol, we obtain a thermal history with slower cooling and now reproduce a third AHe age within  $2\sigma$ . The higher  $\Delta E_a$  is attributed to “vacancy clustering” (Gerin et al., 2017) due to long-term accumulation of radiation damage.

A full assessment of the influence of radiation damage models is beyond the scope of this work, but we highlight it briefly here to show that as our understanding of the diffusion of helium in apatite develops, complex AHe data sets may be better understood. In the interim, the user will need to decide whether to allow the AHe data to have more influence on the form of the thermal history with an appreciation that the choice of radiation damage model may influence the form of the thermal history or adopt a strategy where uncertainty in the method is included in the model and therefore depend on other sources of information (e.g., from other dating techniques or geological data) to help constrain the thermal history.

## ■ EXPLORING PATH STRUCTURE TO INTERPRET MODEL RESULTS

So far, we have presented inverse thermal history modeling and sensitivity tests as examples of how to determine the full range of plausible t-T paths that are consistent with the available thermochronologic data and geologic observations. However, in cases for which there is significant t-T path diversity in inverse model results (i.e., the ExTH is averaging between many thermal histories that are quite different and yet fit the data just as well), it can be difficult to interpret a geologic history. There is not one single way to handle these challenges; however, here we present a workflow for users to follow and a strategy for handling a diverse set of QTQt inverse model results (Fig. 16). We call this strategy the “Path Structure Approach,” which can be used to evaluate model results that show disparate thermal paths, with essentially the same likelihood (e.g., MaxLike and MaxPost paths), from a geologic process-based



**Figure 15.** Left: ExTH results (top) and data predictions (bottom) from inverting: (1) Apatite fission-track (AFT) and apatite (U-Th-Sm)/He (AHe) data (AHe data modeled with 11% uncertainty) (black line); (2) AFT and AHe data (AHe data modeled with resampling the observed ages within the uncertainty) (green line); (3) AFT and AHe data (AHe data modeled with resampling the error on the AHe age) (orange line). Right: ExTH results (top) and data predictions (bottom) from inverting: (1) AFT and AHe data, with AHe data modeled with the Flowers et al. (2009) radiation damage accumulation and annealing model (RDAAM) (black line); (2) AFT and AHe data, with AHe data modeled with the Gautheron et al. (2009) radiation damage model (green line); (3) AFT and AHe data, with AHe data modeled with the Gerin et al. (2017) radiation damage model and a “trapping” energy of 50 kJ/mol (see text for details) (orange line); and (4) AFT and AHe data, with AHe data modeled with the Gerin et al. (2017) radiation damage model and a “trapping” energy of 70 kJ/mol (see text for details) (blue line). Predicted and observed ages are central AFT ages and uncorrected AHe ages.

perspective. Our approach is similar in nature to the Path Family Approach outlined in Murray et al. (2022) for looking into different plausible models in HeFTy and builds on similar strategies for exploring conditional probabilities described by Fox and Shuster (2014).

We propose a workflow that users can follow to identify a “preferred” thermal history to interpret in a geological context (Fig. 16). The workflow involves an iterative process of modeling, evaluating what is learned from the model, revising the modeling approach, if necessary, based on what has been learned,

and selecting a “preferred” model. Through this process, a user should learn what parts of the model are best resolved by the data, where complexity may exist but is not necessarily required to fit the data, and how the available data limit the thermal history resolution that can be achieved. This workflow may be a very quick process if the ExTH model fits the data and is consistent with independent geological information. However, the workflow may be more laborious if data are complicated and/or there are a number of hypotheses that need to be explored. Often the ExTH is a good choice for the “preferred”

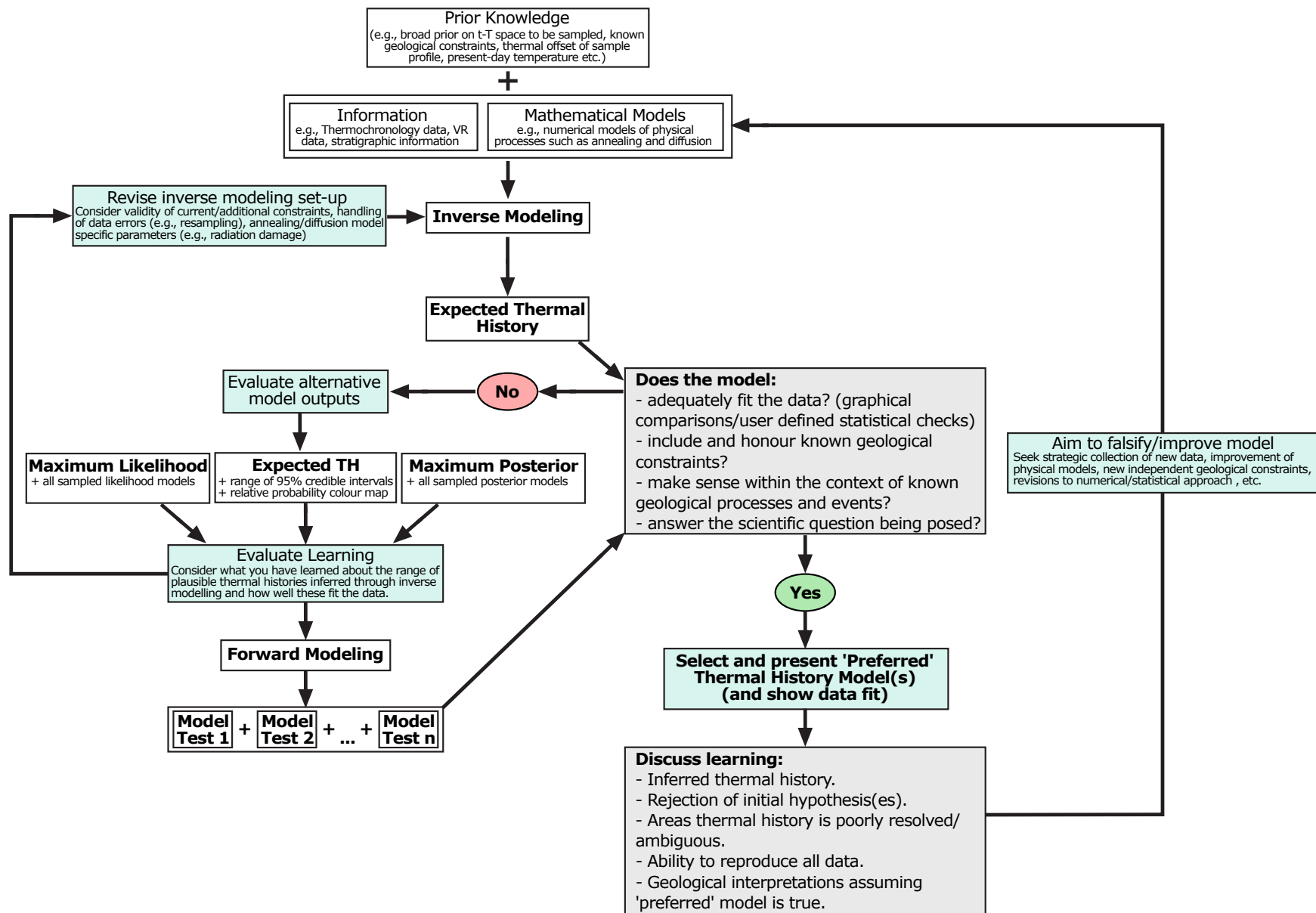


Figure 16. Workflow for the “Path Structure Approach,” a model assessment strategy introduced in this paper. VR – vitrinite reflectance.



thermal history because it conveys the posterior distribution of all accepted models and highlights the level of consistency across individual models (for example, with broader or tighter 95% credible intervals). We emphasize that the MaxLike model should always be interpreted with caution due to the complexity it may infer to fit the data.

In this paper, we give an example of progressing through this workflow using four samples with just the AFT ages from the vertical profile discussed in the previous section (see Sensitivity Testing: A Vertical Profile Sampling Strategy). We pose the hypothetical scenario (as an exercise only) of controversial evidence for overlying sedimentary rocks of unknown thickness that would require burial heating of our samples during the early to middle Miocene. In this scenario, we are still interested in resolving the timing and rate of fault-driven exhumation of the profile during the late Cenozoic. Specifically, we would like to know if exhumation has been at a steady rate since the Oligocene, or whether there was a younger, more rapid exhumation event in the Miocene.

Initial modeling of the AFT data without any imposed constraints produces an ExTH with slow cooling between 20–10 Ma before an increasing cooling rate to present day (Fig. 17). We observe that the 95% credible intervals are broad at 18 Ma and skew to much colder temperatures, suggesting rapid cooling of the samples before the data are used to resolve the younger part of the thermal history. We also note that although the 95% credible intervals narrow to constrain the thermal history ca. 10 Ma, they broaden and skew toward higher temperatures ca. 5 Ma. These patterns in the 95% credible intervals indicate that the ExTH is averaging multiple different thermal histories, which may be equally plausible with this specific given data set (e.g., the AFT data were not published with track lengths).

Next, we take a closer look at the data fit (Fig. 17A). The data predictions for the ExTH do reproduce the four samples within  $2\sigma$  uncertainty. The sampled range of predicted AFT ages (i.e., the AFT ages predicted by all models) also fit the observed AFT ages. This output (the ExTH path) may be entirely satisfactory to the user if there is no known additional geologic information that can be incorporated. However, in our hypothetical scenario, we have reason to believe there may have been sedimentary cover in the early to middle Miocene, and the ExTH inversion does not reflect this independent information; thus, additional modeling work is required (Fig. 16).

The objective of the next step is to find a model that (1) better resolves the history between ca. 20–10 Ma, the timing of possible Miocene burial heating, and (2) better fits the data. To this end, we choose to examine the noticeably wide credible intervals of ~18 m.y. using additional model outputs including the MaxLike and MaxPost models. We generate plots showing representative accepted thermal histories, color coded by the relative likelihood and posterior (Fig. 17A). We observe that the MaxLike and MaxPost models are entirely contradictory but reproduce all of the data within  $2\sigma$  uncertainty (Fig. 17A).

The contradictory structures of the MaxLike and MaxPost outputs suggest that potential solutions might be lost in averaging to the ExTH output; we re-evaluate our inverse modeling approach using three tests. Inverse Test (1) manually incorporates an initial constraint to explore whether the inference

of rapid cooling at the beginning of the unconstrained model is constrained by the data. Inverse Test (2) examines the influence of a constraint that forces the model to near-surface temperatures at ca.  $18 \pm 2$  Ma to reflect hypothetical sedimentation and burial in the middle Miocene. Inverse Test (3) uses both constraints from (1) and (2) (Fig. 17B).

Inverse Test (1), with an initial high-temperature constraint, yields an ExTH that shows that monotonic cooling fits the data reasonably well, and this is reinforced by the MaxPost model, although the 95% credible intervals influenced by the MaxLike models suggest the likelihood that <5 Ma reheating will also fit the data (Fig. 17B). From Inverse Test (2), we see that the ExTH produced, which is conditional on the low-temperature constraint at  $18 \pm 2$  Ma (Fig. 17B), fails to predict the two lower-elevation AFT ages within  $2\sigma$  uncertainty. The MaxLike and MaxPost models do fit the data well, but the solutions show variable timing and temperatures for maximum heating post-10 Ma and different heating and cooling rates (Fig. 17B). Inverse Test (3) shows that all models (ExTH, MaxLike, and MaxPost) can fit the data, and that if the 18 Ma constraint is valid, then there must be reheating following that phase of cooling, but the timing and rates of that reheating and subsequent cooling are variable (Fig. 17B).

If we now evaluate the results of the three inverse tests, we can conclude that two thermal history scenarios appear to adequately fit the data within  $2\sigma$ : a simple monotonic cooling history as well as a more complex thermal history with cooling to low temperatures at 18 Ma, followed by re-heating and final cooling. However, the rates of cooling and heating and the time and maximum temperature of the reheating episode are not well resolved. In this scenario, the ExTH presents an average of the possible solutions and itself does not reproduce the data as well as the other possible solutions. This approach tests the impact of including hypothetical constraint boxes during the inversion and assesses the posterior probability conditional on these constraints. This strategy draws comparisons to that employed by Fox and Shuster (2014), where, following the inversion of the data, Fox and Shuster (2014) filter models from the posterior solution of the inversion based on a specified constraint and assess the conditional probabilities across the t-T space.

This approach makes it readily apparent that the data and model design alone are not sufficient to distinguish between the monotonic cooling and cooling-reheating scenarios. Therefore, if the user needs to select a preferred model, they must justify why it is appropriate and discuss the limitations of their interpretations.

A round of forward modeling may help extract new useful information from this generally inconclusive inversion result and to identify a “preferred” model that predicts the data (e.g., Fig. 17C), both of which can then be carried into broader discussion of the geological implications. We present examples of two forward modeling iterations that test (1) the timing for a reheating episode where the top sample is brought to 100 °C post-15 Ma, and (2) the temperature to which the samples cool at 18 Ma. We model all four samples in the profile with the thermal history experienced by the top sample in the profile shown in Figure 17C, the profile offset between the samples before

Path Structure Approach: using the workflow (Fig. 16)

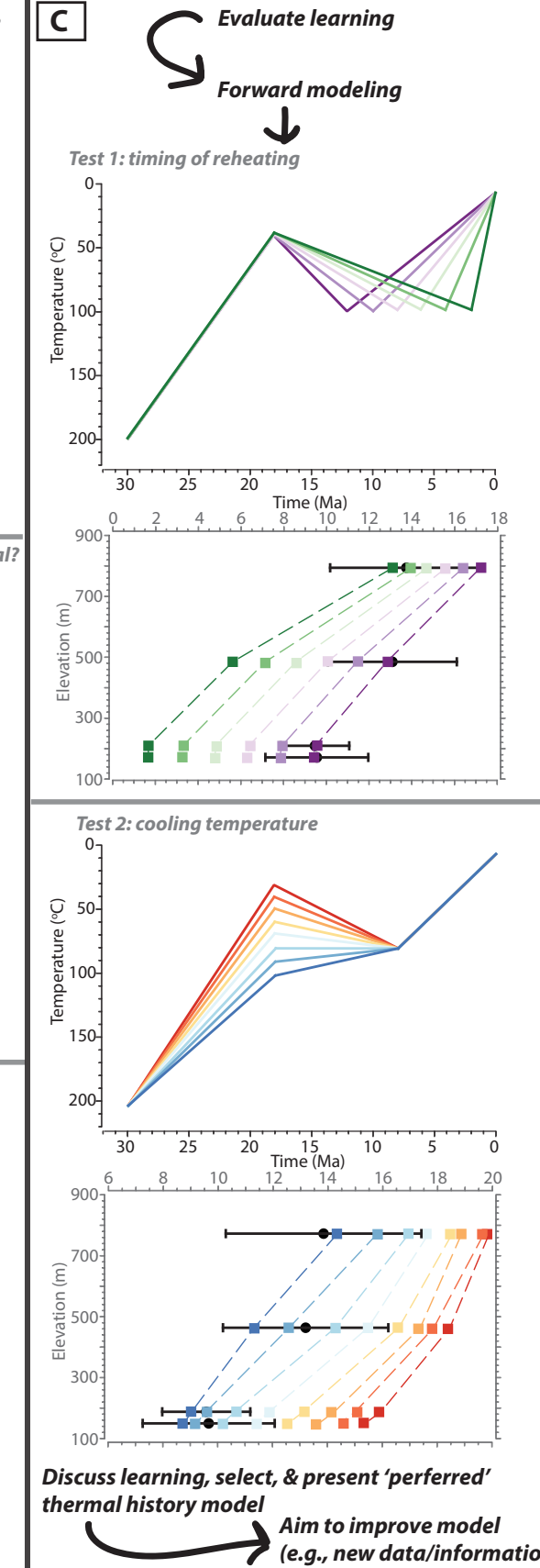
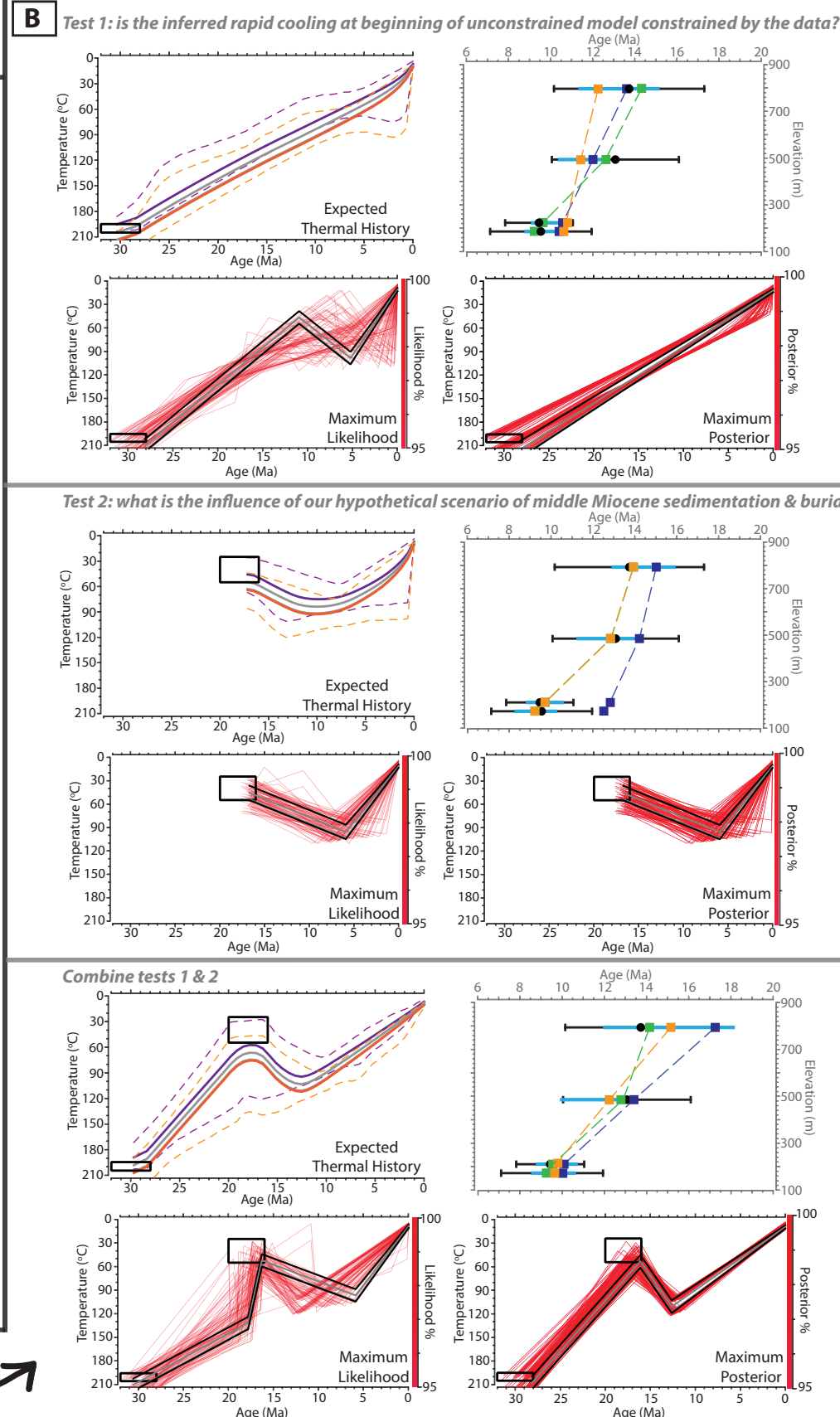
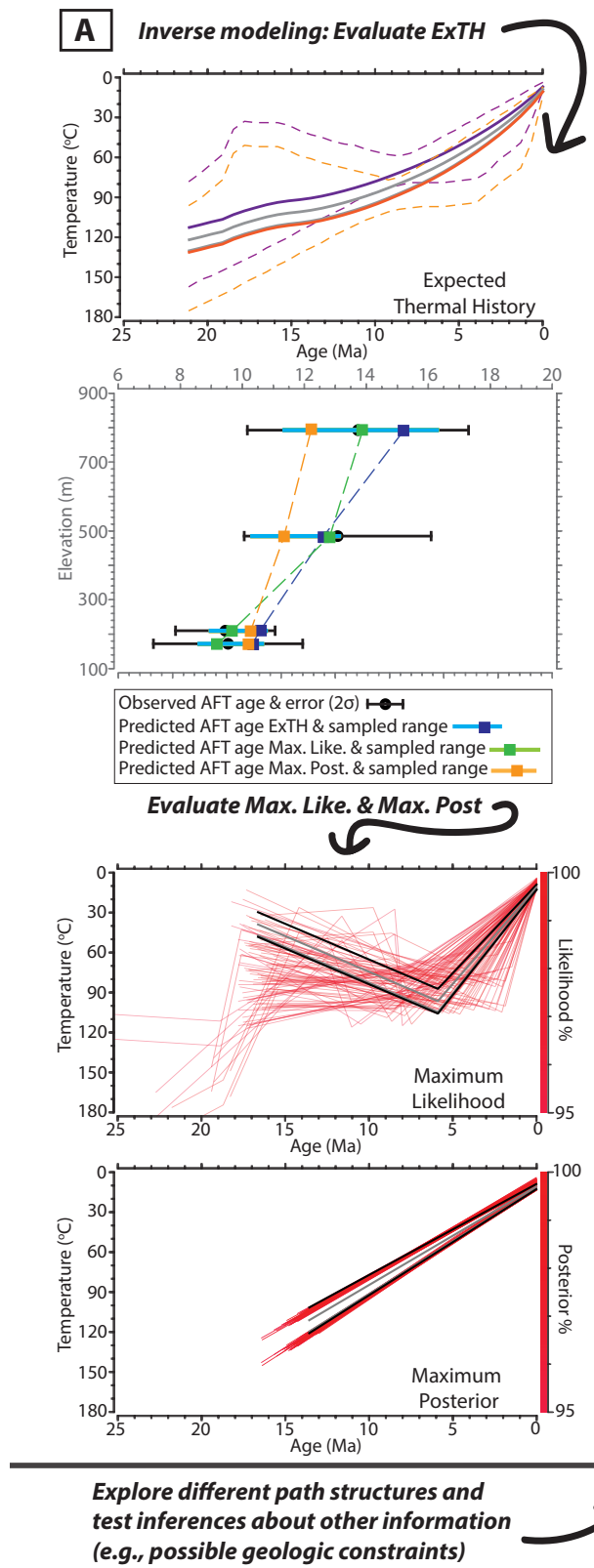


Figure 17. Example using the “Path Structure Approach” to explore the ExTH output from an inversion of apatite fission-track (AFT) data. Panel A (left) shows the initial inverse model result and data predictions. The 95% credible intervals around the ExTH suggest that different thermal histories, which are being averaged, skew toward lower temperatures ca. 18 Ma and warmer temperatures ca. 5 Ma. Those different histories can be seen in the lower two plots of panel A showing the thermal history paths in the top 95% for both the MaxLike and MaxPost outputs. Panel B (middle) exemplifies three inverse tests that explore the structure of the MaxLike and MaxPost outputs seen in panel A, using constraint boxes. Panel C (right) uses two forward modeling tests to examine the path structure seen in the MaxLike and MaxPost outputs from panel A. Bold black labels and arrows indicate where and what parts of the workflow (Fig. 16) are in progress.

present-day is defined by the 30 °C/km geothermal gradient, and the vertical distance between the samples. If the top sample cools to temperatures below the AFT PAZ at 18 Ma, then the only way to fit all the data is if the samples are reheated by 12–10 Ma. If reheating happens more recently than 10 Ma, the lower samples in the profile do not fit the data. However, the data fit better if the cooling episode at 18 Ma does not cool the samples below ~70 °C (Fig. 17C). Thus, if one views the early-mid Miocene burial as required by the geologic record, a “Path Structure Approach” for assessing these data places new bounds on the timing and magnitude of the associated rock heating and cooling, even if the inverse models cannot strongly support or refute that this hypothetical “controversial” geologic event actually occurred. Alternatively, if a user has a clearly expressed philosophical preference for using the simplest history resolved by an inverse model for geologic interpretations, then the monotonic cooling supported by the MaxPost models could be an appropriate “preferred” model result. Regardless of one’s preferred result, all the possible thermal histories should be discussed as a part of the geologic interpretation to lay a strong foundation for future work.

Iterating between thermal history model results and testable, geologically based thermal signatures provides not only an effective workflow for assessing inconclusive inverse model results but also a roadmap for which additional geologic observations or data could narrow the inverse model results. Knowing that the data in this “Path Structure Approach” example are not as informative as they could be (i.e., only AFT data with no track-length information), we could then decide to take the final step in the workflow and aim to improve the model (Fig. 16) by adding new data or geologic observations. These particular data are part of a vertical profile in which zircon and apatite He data also exist. The ExTH with all the data included (Fig. 12D) shows some parts of the thermal history that were inferred using just the AFT data and the “Path Structure Approach” (e.g., cooling to lower temperatures by 18 Ma), plus minor reheating or isothermal holding and more recent cooling. These structures are now reinforced by the He data, agree with the forward model tests, and provide evidence against the slow constant monotonic cooling thermal history we derived from the AFT ages alone. Much of the data are reproduced by the ExTH in this new model (Fig. 12D), but some of the AHe ages are reproduced less well. Again, we can iterate, evaluating what we may have learned about our observed data and assessing whether the shortcomings are from the analyzed grains or the diffusion model that is used to predict the AHe ages. The ExTH may then be chosen as a preferred model to interpret geologically, but it remains ready to be falsified or made more robust with the addition of new thermochronologic data (e.g., AFT data with track lengths) and/or other new geologic information. We reiterate that throughout this process, model predictions should be compared to the data and can be used to inform further model testing.

Even in a situation where all of the data are predicted well by a thermal history, the thermal history is not a unique solution. Any advancement in our understanding of the geology of a region, new data collection, and helium diffusion and/or fission track annealing has the potential to impact an existing

thermal history model. The “Path Structure Approach” offers an iterative workflow that allows users to explore complex or ambiguous model results. This is especially useful for a program such as QTQt, where the commonly interpreted thermal history output (ExTH) is not an actual thermal history path, but instead is a weighted average of many different possible thermal histories. By using this approach and our workflow as a guide (Fig. 16), users will not only be able to examine different features of a thermal history but can do it in a way that is easy to record and document all steps of the process.

## CONCLUSIONS

Thermal history modeling allows us to leverage the richness of low-temperature thermochronologic data to make more robust interpretations in a range of geologic contexts. As new research projects generate and model such data, an explicit and ongoing discussion of modeling approaches and interpretation strategies is needed to highlight the opportunities and limitations of thermal history modeling to resolve geologic problems. Although this contribution features examples using QTQt, many of the strategies are transferable to forward and inverse thermal history modeling exercises using alternative programs or codes. Our contribution is designed to:

- (1) Cultivate an intuitive approach to thermal history modeling. The synthetic data sets applied in this study are “perfect” data with known “true” thermal histories providing an opportunity to develop an intuitive understanding of the capacity of thermochronologic data sets to distinguish among distinct thermal histories using both forward and inverse modeling. This awareness is useful both during research planning—for example, in deciding which types of data to collect for testing competing hypotheses—as well as in the modeling and data interpretation stages of a project.
- (2) Suggest best practices for model design and execution. We use both synthetic data and examples from measured (geologic) data sets to outline strategies for sensitivity testing thermal history models. Sensitivity testing can be used to interrogate what factor(s) control model behavior, including, but not limited to, the model design or inversion algorithm, a geologic constraint, a data trend, the spatial relationship between samples, or the choice of kinetics model. Our examples illustrate why sensitivity testing is an important exercise to avoid subsampling thermal history solutions for a given data set and inadvertently misrepresenting the thermal history model results.
- (3) Emphasize the importance of reporting all of the decisions that go into modeling, including which modeling program and version was used, which data are modeled, and what additional model parameters were defined. Documenting the effects of modeling decisions on model output provides context for geologic interpretations. Importantly, reporting each decision will help external users use or reproduce results. There are several templates available for workers to use when recording their sensitivity tests and when publishing modeling decisions (e.g., Tables S2–S7, see footnote 1; Flowers et al., 2015; Wildman et al., 2019b).

(4) Introduce interpretation strategies including a detailed workflow for what we term the “Path Structure Approach,” well suited to evaluate thermal history model results from QTQt. Describing, interrogating, and interpreting thermal history model results can be challenging. We provide examples of working with profile data, multiple chronometers, and data sets with seemingly unexplainable age variation. As researchers, we know our data are not perfect, nor are the predictive models for diffusion and annealing; therefore, we should not expect a perfectly resolved thermal history model. Consequently, we highly recommend performing similar sensitivity tests, with both inverse and forward modeling approaches, and we encourage the development of additional useful interpretation strategies for all types of thermochronometric data.

#### ACKNOWLEDGMENTS

The authors thank the participants of the HeFTy and QTQt workshops at the 17th International Conference on Thermochronology and Kaylin Luciani (California State University Long Beach undergraduate) for feedback on the tutorial provided in the Supplemental Material (footnote 1). We thank Rich Ketcham and Kerry Gallagher for support and encouragement for both this contribution and the HeFTy-focused companion manuscript. We thank Kerry Gallagher for providing feedback on ideas and early drafts of this manuscript. We thank Matthew Fox and one anonymous reviewer for comments that improved this work. The authors welcomed Charlie (MW), Theodore (KEM), and Timothy (ASG) into the world during the submission, review, and revision of this manuscript. We thank the editors and reviewers for their patience and accommodation through the review process.

#### REFERENCES CITED

- Abbey, A.L., and Niemi, N.A., 2018, Low-temperature thermochronometric constraints on fault initiation and growth in the northern Rio Grande rift, upper Arkansas River valley, Colorado, USA: *Geology*, v. 46, no. 7, p. 627–630, <https://doi.org/10.1130/G40232.1>.
- Ault, A.K., Gautheron, C., and King, G.E., 2019, Innovations in (U-Th)/He, fission track, and trapped charge thermochronometry with applications to earthquakes, weathering, surface-mantle connections, and the growth and decay of mountains: *Tectonics*, v. 38, no. 11, p. 3705–3739, <https://doi.org/10.1029/2018TC005312>.
- Baughman, J.S., and Flowers, R.M., 2020, Mesoproterozoic burial of the Kaapvaal craton, southern Africa during Rodinia supercontinent assembly from (U-Th)/He thermochronology: *Earth and Planetary Science Letters*, v. 531, <https://doi.org/10.1016/j.epsl.2019.115930>.
- Braun, J., 2003, Pecube: A new finite-element code to solve the 3D heat transport equation including the effects of a time-varying, finite amplitude surface topography: *Computers & Geosciences*, v. 29, no. 6, p. 787–794, [https://doi.org/10.1016/S0098-3004\(03\)00052-9](https://doi.org/10.1016/S0098-3004(03)00052-9).
- Brown, R.W., Beucher, R., Roper, S., Persano, C., Stuart, F., and Fitzgerald, P., 2013, Natural age dispersion arising from the analysis of broken crystals. Part I: Theoretical basis and implications for the apatite (U-Th)/He thermochronometer: *Geochimica et Cosmochimica Acta*, v. 122, p. 478–497, <https://doi.org/10.1016/j.gca.2013.05.041>.
- Carlson, W.D., Donelick, R.A., and Ketcham, R.A., 1999, Variability of apatite fission-track annealing kinetics: I. Experimental results: *The American Mineralogist*, v. 84, no. 9, p. 1213–1223, <https://doi.org/10.2138/am-1999-0901>.
- Djimbi, D.M., Gautheron, C., Roques, J., Tassan-Got, L., Gerin, C., and Simoni, E., 2015, Impact of apatite chemical composition on (U-Th)/He thermochronometry: An atomistic point of view: *Geochimica et Cosmochimica Acta*, v. 167, p. 162–176, <https://doi.org/10.1016/j.gca.2015.06.017>.
- Dodson, M., 1973, Closure temperature in cooling geochronological and petrological systems: Contributions to Mineralogy and Petrology, v. 40, p. 259–274, <https://doi.org/10.1007/BF00373790>.
- Donelick, R.A., Ketcham, R.A., and Carlson, W.D., 1999, Variability of apatite fission-track annealing kinetics: II. Crystallographic orientation effects: *The American Mineralogist*, v. 84, no. 9, p. 1224–1234, <https://doi.org/10.2138/am-1999-0902>.

- Flowers, R.M., and Kelley, S.A., 2011, Interpreting data dispersion and “inverted” dates in apatite (U-Th)/He and fission-track datasets: An example from the US midcontinent: *Geochimica et Cosmochimica Acta*, v. 75, no. 18, p. 5169–5186, <https://doi.org/10.1016/j.gca.2011.06.016>.
- Flowers, R.M., Ketcham, R.A., Shuster, D.L., and Farley, K.A., 2009, Apatite (U-Th)/He thermochronometry using a radiation damage accumulation and annealing model: *Geochimica et Cosmochimica Acta*, v. 73, no. 8, p. 2347–2365, <https://doi.org/10.1016/j.gca.2009.01.015>.
- Flowers, R.M., Farley, K.A., and Ketcham, R.A., 2015, A reporting protocol for thermochronologic modeling illustrated with data from the Grand Canyon: *Earth and Planetary Science Letters*, v. 432, p. 425–435, <https://doi.org/10.1016/j.epsl.2015.09.053>.
- Flowers, R.M., Farley, K.A., and Ketcham, R.A., 2016, Response to comment on “A reporting protocol for thermochronologic modeling illustrated with data from the Grand Canyon”: *Earth and Planetary Science Letters*, v. 441, p. 213, <https://doi.org/10.1016/j.epsl.2016.02.024>.
- Flowers, R.M., Zeitler, P.K., Danišik, M., Reiners, P.W., Gautheron, C., Ketcham, R.A., Metcalf, J.R., Stockli, D.F., Enkelmann, E., and Brown, R.W., 2022a, (U-Th)/He chronology: Part 1. Data, uncertainty, and reporting: *Geological Society of America Bulletin*, <https://doi.org/10.1130/B36266.1>.
- Flowers, R.M., Ketcham, R.A., Enkelmann, E., Gautheron, C., Reiners, P.W., Metcalf, J.R., Danišik, M., Stockli, D.F., and Brown, R.W., 2022b, (U-Th)/He chronology: Part 2. Considerations for evaluating, integrating, and interpreting conventional individual aliquot data: *Geological Society of America Bulletin*, <https://doi.org/10.1130/B36268.1>.
- Fox, M., and Carter, A., 2020, Heated topics in Thermochronology and paths towards resolution: *Geosciences*, v. 10, no. 9, p. 375, <https://doi.org/10.3390/geosciences10090375>.
- Fox, M., and Shuster, D.L., 2014, The influence of burial heating on the (U-Th)/He system in apatite: Grand Canyon case study: *Earth and Planetary Science Letters*, v. 397, p. 174–183, <http://doi.org/10.1016/j.epsl.2014.04.041>.
- Gallagher, K., 1995, Evolving temperature histories from apatite fission-track data: *Earth and Planetary Science Letters*, v. 136, p. 421–435, [https://doi.org/10.1016/0012-821X\(95\)00197-K](https://doi.org/10.1016/0012-821X(95)00197-K).
- Gallagher, K., 2012, Transdimensional inverse thermal history modelling for quantitative thermochronology: *Journal of Geophysical Research: Solid Earth*, v. 117, <https://doi.org/10.1029/2011JB008825>.
- Gallagher, K., 2016, Comment on ‘A reporting protocol for thermochronologic modeling illustrated with data from the Grand Canyon’ by Flowers, Farley and Ketcham: *Earth and Planetary Science Letters*, v. 441, p. 211–212, <https://doi.org/10.1016/j.epsl.2016.02.021>.
- Gallagher, K., and Ketcham, R.A., 2018, Comment on “Thermal history modelling: HeFTy vs. QTQt” by Vermeesch and Tian, *Earth-Science Reviews* (2014), 139, 279–290: *Earth-Science Reviews*, v. 176, p. 387–394, <https://doi.org/10.1016/j.earscirev.2017.11.001>.
- Gallagher, K., and Ketcham, R.A., 2020, Comment on the reply to the Comment on “Thermal history modelling: HeFTy vs. QTQt” by Vermeesch and Tian, *Earth-Science Reviews* (2014), 139, 279–290: *Earth-Science Reviews*, v. 203, <https://doi.org/10.1016/j.earscirev.2019.102878>.
- Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., and Stephenson, J., 2009, Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems: *Marine and Petroleum Geology*, v. 26, p. 525–535, <https://doi.org/10.1016/j.marpetgeo.2009.01.003>.
- Gautheron, C., Tassan-Got, L., Barbarand, J., and Pagel, M., 2009, Effect of alpha-damage annealing on apatite (U-Th)/He thermochronology: *Chemical Geology*, v. 266, no. 3–4, p. 157–170, <https://doi.org/10.1016/j.chemgeo.2009.06.001>.
- Gautheron, C., Barbarand, J., Ketcham, R.A., Tassan-Got, L., van der Beek, P., Pagel, M., Pinna-Jamme, R., Coffignal, F., and Fialin, M., 2013, Chemical influence on  $\alpha$ -recoil damage annealing in apatite: Implications for (U-Th)/He dating: *Chemical Geology*, v. 351, p. 257–267, <https://doi.org/10.1016/j.chemgeo.2013.05.027>.
- Gerin, C., Gautheron, C., Oliviero, E., Bachelet, C., Djimbi, D.M., Seydoux-Guillaume, A.M., Tassan-Got, L., Sarda, P., Roques, J., and Garrido, F., 2017, Influence of vacancy damage on He diffusion in apatite, investigated at atomic to mineralogical scales: *Geochimica et Cosmochimica Acta*, v. 197, p. 87–103, <https://doi.org/10.1016/j.gca.2016.10.018>.
- Gleadow, A.J., Duddy, I.R., Green, P.F., and Hegarty, K.A., 1986, Fission track lengths in the apatite annealing zone and the interpretation of mixed ages: *Earth and Planetary Science Letters*, v. 78, no. 2–3, p. 245–254, [https://doi.org/10.1016/0012-821X\(86\)90065-8](https://doi.org/10.1016/0012-821X(86)90065-8).
- Guenther, W.R., Reiners, P.W., Ketcham, R.A., Nasdala, L., and Giester, G., 2013, Helium diffusion in natural zircon: Radiation damage, anisotropy, and the interpretation of zircon (U-Th)/He thermochronology: *American Journal of Science*, v. 313, no. 3, p. 145–198, <https://doi.org/10.2475/03.2013.01>.



- Guo, H., Zeitler, P.K., Idleman, B.D., Fayon, A.K., Fitzgerald, P.G., and McDannell, K.T., 2021, Helium diffusion systematics inferred from continuous ramped heating analysis of Transantarctic Mountains apatites showing age overdispersion: *Geochimica et Cosmochimica Acta*, v. 310, p. 113–130, <https://doi.org/10.1016/j.gca.2021.07.015>.
- Hager, C., and Stockli, D.F., 2009, A new Matlab®-based Helium Modeling Package (“HeMP”) for thermal history recovery from single and multi-thermochronometer (U-Th)/He data and data arrays: *Geological Society of America Abstracts with Programs*, v. 41, no. 7, p. 487.
- Hendriks, B.W.H., and Redfield, T.F., 2005, Apatite fission track and (U-Th)/He data from Fennoscandia: An example of underestimation of fission track annealing in apatite: *Earth and Planetary Science Letters*, v. 236, no. 1–2, p. 443–458, <https://doi.org/10.1016/j.epsl.2005.05.027>.
- House, M.A., Wernicke, B.P., and Farley, K.A., 2001, Paleo-geomorphology of the Sierra Nevada, California, from (U-Th)/He ages in apatite: *American Journal of Science*, v. 301, no. 2, p. 77–102, <https://doi.org/10.2475/ajs.301.2.77>.
- Kelley, S.A., Chapin, C.E., and Corrigan, J., 1992, Late Mesozoic to Cenozoic cooling histories of the flanks of the northern and central Rio Grande rift, Colorado and New Mexico: *New Mexico Bureau of Mines and Mineral Resources*, v. 145, 40 p.
- Ketcham, R.A., 2005, Forward and inverse modelling of low-temperature thermochronometry data: *Reviews in Mineralogy and Geochemistry*, v. 58, no. 1, p. 275–314, <https://doi.org/10.2138/rmg.2005.58.11>.
- Ketcham, R.A., 2019, Fission-track annealing: from geologic observations to thermal history modeling, in Malusà, M.G., and Fitzgerald, P.G., eds., *Fission-Track Thermochronology and Its Application to Geology*: Cham, Switzerland, Springer, p. 49–75, [https://doi.org/10.1007/978-3-319-89421-8\\_3](https://doi.org/10.1007/978-3-319-89421-8_3).
- Ketcham, R.A., Donelick, R.A., and Donelick, M.B., 2000, AFTSolve: A program for multi-kinetic modelling of apatite fission-track data: *Geological Materials Research*, v. 2, p. 1–32.
- Ketcham, R.A., Carter, A., Donelick, R.A., Barbarand, J., and Hurford, A.J., 2007, Improved measurement of fission-track annealing in apatite using c-axis projection: *The American Mineralogist*, v. 92, no. 5–6, p. 789–798, <https://doi.org/10.2138/am.2007.2280>.
- Ketcham, R.A., Tremblay, M., Abbey, A., Baughman, J., Cooperdock, E., Jepson, G., Murray, K., Odium, M., Stanley, J., and Thurston, O., 2022, Report from the 17th International Conference on Thermochronology: *American Geophysical Union*, article 10.1002, <https://doi.org/10.1002/essoar.10511082.1>.
- Kohn, B.P., Lorencak, M., Gleadow, A.J., Kohlmann, F., Raza, A., Osadetz, K.G., and Sorjonen-Ward, P., 2009, A reappraisal of low-temperature thermochronology of the eastern Fennoscandia Shield and radiation-enhanced apatite fission-track annealing, in Lisker, F., et al., eds., *Thermochronological Methods: From Palaeotemperature Constraints to Landscape Evolution Models*: Geological Society of London, Special Publication 324, p. 193–216, <https://doi.org/10.1144/SP324.15>.
- McDannell, K.T., Zeitler, P.K., Janes, D.G., Idleman, B.D., and Fayon, A.K., 2018, Screening apatites for (U-Th)/He thermochronometry via continuous ramped heating: He age components and implications for age dispersion: *Geochimica et Cosmochimica Acta*, v. 223, p. 90–106, <https://doi.org/10.1016/j.gca.2017.11.031>.
- McKeon, R.E., Zeitler, P.K., Pazzaglia, F.J., Idleman, B.D., and Enkelmann, E., 2014, Decay of an old orogen: Inferences about Appalachian landscape evolution from low-temperature thermochronology: *Geological Society of America Bulletin*, v. 126, no. 1–2, p. 31–46, <https://doi.org/10.1130/B30808.1>.
- Morón, S., Kohn, B.P., Beucher, R., Mackintosh, V., Cawood, P.A., Moresi, L., and Gallagher, S.J., 2020, Denuding a craton: Thermochronology record of Phanerozoic unroofing from the Pilbara Craton, Australia: *Tectonics*, v. 39, <https://doi.org/10.1029/2019TC005988>.
- Murray, K.E., Reiners, P.W., Thomson, S.N., Robert, X., and Whipple, K.X., 2019, The thermochronologic record of erosion and magmatism in the Canyonlands region of the Colorado Plateau: *American Journal of Science*, v. 319, p. 339–380, <https://doi.org/10.2475/05.2019.01>.
- Murray, K.E., Stevens Goddard, A.L., Abbey, A.L., and Wildman, M., 2022, Thermal history modeling techniques and interpretation strategies: Applications using HeFTy: *Geosphere*, v. 18, no. 5, p. 1622–1642, <https://doi.org/10.1130/GES02500.1>.
- O’Sullivan, P.B., and Parrish, R.R., 1995, The importance of apatite composition and single-grain ages when interpreting fission track data from plutonic rocks: A case study from the Coast Ranges, British Columbia: *Earth and Planetary Science Letters*, v. 132, no. 1–4, p. 213–224, [https://doi.org/10.1016/0012-821X\(95\)00058-K](https://doi.org/10.1016/0012-821X(95)00058-K).
- Reiners, P.W., and Farley, K.A., 2001, Influence of crystal size on apatite (U-Th)/He thermochronology: An example from the Bighorn Mountains, Wyoming: *Earth and Planetary Science Letters*, v. 188, no. 3–4, p. 413–420, [https://doi.org/10.1016/S0012-821X\(01\)00341-7](https://doi.org/10.1016/S0012-821X(01)00341-7).
- Ricketts, J.W., Kelley, S.A., Karlstrom, K.E., Schmandt, B., Donahue, M.S., and van Wijk, J., 2016, Synchronous opening of the Rio Grande rift along its entire length at 25–10 Ma supported by apatite (U-Th)/He and fission-track thermochronology, and evaluation of possible driving mechanisms: *Geological Society of America Bulletin*, v. 128, no. 3–4, p. 397–424, <https://doi.org/10.1130/B31223.1>.
- Shuster, D.L., Flowers, R.M., and Farley, K.A., 2006, The influence of natural radiation damage on helium diffusion kinetics in apatite: *Earth and Planetary Science Letters*, v. 249, p. 148–161, <https://doi.org/10.1016/j.epsl.2006.07.028>.
- Sobel, E.R., and Seward, D., 2010, Influence of etching conditions on apatite fission-track etch pit diameter: *Chemical Geology*, v. 271, no. 1–2, p. 59–69, <https://doi.org/10.1016/j.chemgeo.2009.12.012>.
- Sousa, F.J., and Farley, K.A., 2020, A framework for evaluating variation in (U-Th)/He datasets: *Minerals*, v. 10, <https://doi.org/10.3390/min10121111>.
- Stephenson, J., Gallagher, K., and Holmes, C.C., 2006, Low temperature thermochronology and strategies for multiple samples: 2: Partition modelling for 2D/3D distributions with discontinuities: *Earth and Planetary Science Letters*, v. 241, no. 3–4, p. 557–570, <https://doi.org/10.1016/j.epsl.2005.11.027>.
- Tamer, M.T., Chung, L., Ketcham, R.A., and Gleadow, A.J., 2019, Analyst and etching protocol effects on the reproducibility of apatite confined fission-track length measurement, and ambient-temperature annealing at decadal timescales: *American Mineralogist: Journal of Earth and Planetary Materials*, v. 104, no. 10, p. 1421–1435, <https://doi.org/10.2138/am-2019-7046>.
- Vermeesch, P., and Tian, Y., 2014, Thermal history modelling: HeFTy vs. QTQt: *Earth-Science Reviews*, v. 139, p. 279–290, <https://doi.org/10.1016/j.earscirev.2014.09.010>.
- Vermeesch, P., and Tian, Y., 2018, Reply to Comment on “Thermal history modelling: HeFTy vs. QTQt” by K. Gallagher and R.A. Ketcham: *Earth-Science Reviews*, v. 176, p. 395–396, <https://doi.org/10.1016/j.earscirev.2017.11.015>.
- Vermeesch, P., and Tian, Y., 2019, Reply to the Comment on the Reply to the Comment on Vermeesch and Tian (2014): *Earth-Science Reviews*, v. 203, <https://doi.org/10.1016/j.earscirev.2019.102879>.
- Wildman, M., Brown, R., Beucher, R., Persano, C., Stuart, F., Gallagher, K., Schwanethal, J., and Carter, A., 2016, The chronology and tectonic style of landscape evolution along the elevated Atlantic continental margin of South Africa resolved by joint apatite fission track and (U-Th-Sm)/He thermochronology: *Tectonics*, v. 35, no. 3, p. 511–545, <https://doi.org/10.1002/2015TC004042>.
- Wildman, M., Cogné, N., and Beucher, R., 2019a, Fission-track thermochronology applied to the evolution of passive continental margins, in Malusà, M.G., and Fitzgerald, P.G., eds., *Fission-Track Thermochronology and Its Application to Geology*: Cham, Springer, p. 351–371, <https://doi.org/10.1016/j.earscirev.2019.103044>.
- Wildman, M., Webster, D., Brown, R., Chardon, D., Rouby, D., Ye, J., Huyghe, D., and Dall’Asta, M., 2019b, Long-term evolution of the West African transform margin: Estimates of denudation from Benin using apatite thermochronology: *Journal of the Geological Society of London*, v. 176, no. 1, p. 97–114, <https://doi.org/10.1144/jgs2018-078>.
- Willett, C.D., Fox, M., and Shuster, D.L., 2017, A helium-based model for the effects of radiation damage annealing on helium diffusion kinetics in apatite: *Earth and Planetary Science Letters*, v. 477, p. 195–204, <https://doi.org/10.1016/j.epsl.2017.07.047>.
- Wolf, R.A., Farley, K.A., and Silver, L.T., 1996, Helium diffusion and low-temperature thermochronometry of apatite: *Geochimica et Cosmochimica Acta*, v. 60, no. 21, p. 4231–4240, [https://doi.org/10.1016/S0016-7037\(96\)00192-5](https://doi.org/10.1016/S0016-7037(96)00192-5).
- Wolf, R.A., Farley, K.A., and Kass, D.M., 1998, Modeling of the temperature sensitivity of the apatite (U-Th)/He thermochronometer: *Chemical Geology*, v. 148, no. 1–2, p. 105–114, [https://doi.org/10.1016/S0009-2541\(98\)00024-2](https://doi.org/10.1016/S0009-2541(98)00024-2).
- Zeitler, P.K., 2004, Avert 4.0.1. Inversion of <sup>40</sup>Ar/<sup>39</sup>Ar age spectra: User’s Manual: [https://eesarchive.lehigh.edu/EESdocs/geochron/downloads/arvert/arvert\\_41\\_manual.pdf](https://eesarchive.lehigh.edu/EESdocs/geochron/downloads/arvert/arvert_41_manual.pdf).