Author reply on RC1

Kalin T. McDannell and C. Brenhin Keller
Dartmouth College

October 8, 2024

We thank reviewer #1 Kendra Murray for comments on preprint gchron-2024-3. We appreciate the opportunity to clarify some points. Below we respond to each of the comments below.

RC3 denotes reviewer #1 comments and **AC** and bold text marks the authors' replies.

RC3: GENERAL COMMENTS

This manuscript describes a quantitative method for extracting the timing of rock cooling from a QTQt inversion, given a time period of interest that encompasses a simple cooling event. Specifically, this approach documents the timing of the maximum rate of cooling, which is different than other common metrics for characterizing the timing or magnitude of thermal features in inversion results (such as the timing of the start of cooling, or the rate of cooling). This timing of the peak cooling rate is inferred to occur at the midpoint of the cooling event, as tT histories pass through a 'half-maximum isotherm'. The distribution of times when each thermal history solution passes through this isotherm is termed the Full time-Distribution at Half-Maximum temperature (FDHM), i.e., the timing of peak cooling. The source code is available in GitHub and was originally developed and used to interpret data in a publication from 2022 by the same authors. To demonstrate how this tool works, the authors use several synthetic examples. Then, they discuss how the FDHM could be used to address geological questions using their previously published deep-time inversion result from the Canadian Shield and new modeling of published data that document the recent cooling of the Bergell pluton in the Alps.

There is a need for this kind of contribution, which extracts a inverse model analysis tool developed for use in a recent paper (McDannell and Keller, 2022) and makes it more accessible to the broader community. I think that with moderate revisions, this manuscript will provide a valuable entry point anyone who wants to add this method to their thermal history modeling toolkit. Below I give an overview of the specific areas where I think revision or additional discussion would strengthen this manuscript's description and demonstration of the method; in the accompanying PDF, I flag a number of additional areas for the authors to consider additional rephrasing and revision. I hope the authors find this feedback useful. I look forward to taking this method out for a spin!

SPECIFIC COMMENTS (see also PDF)

RC3: What is the geological significance of "the timing of peak cooling"? The mathematical elegance of the FDHM is clear, but that doesn't mean it has inherent or intuitive geological significance. To me, this metric doesn't intuitively relate to a part of a geological process, and so it is much less compelling than the starting, ending, or rate of cooling; perhaps this is because it is a novel way of thinking about cooling events. Although the geological examples help somewhat, I find the synthetic histories and accompanying discussion significantly less helpful on this front, for a few reasons. Revising how the synthetic examples are leveraged and discussed could support a more compelling argument for how and when to use the FHDM metric.

AC: We appreciate the reviewer's thoughts on this, which are also shared by the other reviewers. We are not sure what is meant by the start/end/rate of cooling being more "compelling," since those features are what they are, and we can usually only quantify approximations of the start/end of cooling. We do not inherently know cooling rate directly from thermochronology. The 'start/end' of a cooling path in a thermal history are still interpreted features, not being really much different than the 'middle' of cooling in that sense. One somewhat unspoken assumption is that our model is accurately capturing a cooling signal, but how much of, and how well we reproduce that signal depends on how well our thermochronometric data can resolve detail. Sometimes we are prevented from resolving the 'start' of a cooling path if a thermochronometer is only resolving a partial signal (as discussed in the m.s.)—so the start of cooling in such a case is only the start of thermochronometric sensitivity, which may be meaningless geologically. The 'end' of cooling (in the sense of reaching low temperatures encountered at Earth's surface) can be outside the sensitivity bounds of most commonly applied low-temperature chronometers, and in terms of "system response" can also be difficult to constrain without external information (e.g., Green and Duddy, 2021; doi: 10.1016/j.earscirev.2020.103197).

AC: Our explanation of the FDHM method could be better. It is a metric of the midpoint of the cooling event, and the least biased estimate of the time of an event because 'start' and 'end' are subject to uncertainties inherent in thermochronology, whether kinetic or otherwise. Whatever specific geologic event a cooling signal is related to, or caused by, is an interpretive step beyond what the FDHM tool is meant for. The time of half-maximum cooling can be considered an approximation of the time when the first derivative of the cooling curve is maximized if the cooling curve is smooth, symmetrical, and unimodal (under those caveats it's also the inflection point, or the time when the second derivative changes sign). Interpreting cooling as exhumation—if the onset and termination of cooling are geologically meaningful, wouldn't the middle of a cooling event represent the 'middle' of an exhumation event? (also referenced in reply to RC2). In such a case, it represents the time at which half the total exhumation has occurred.

RC3: Line 155 described the cooling rates chosen in the synthetic tT histories as "arbitrary" (see comment in pdf about clarifying the meaning of this sentence). If that is the intended meaning, choosing 'arbitrary' cooling rates is unsatisfying, because it isn't clear if and how they are relevant to real cooling histories rocks commonly experience in nature, and thus how to connect these synthetic examples to the real world. Synthetic tT histories have the most value if they are designed to mimic something in nature (exactly what depends on the context, of course).

AC: **Use of the word "arbitrary" is imprecise language and we will change this upon revision. The use of cooling rates of 10°C/Myr or 1°C/Myr as examples are very common in published thermochronological literature, and in this case, the exact trajectory of those paths is less meaningful than their purpose of introducing a 'real' dataset that includes some uncertainty to test the FDHM. Geologic setting isn't really a concern in these synthetic examples, since they are simply cases to see how well we can recover the cooling midpoint from a 'known' t-T path. Whether those are, for example, volcanic or cratonic t-T paths makes little difference in terms of FDHM outcome. Nonetheless, we could add 'typical volcanic cooling' and 'typical cratonic cooling' labels to the paths if that helps.**

RC3: The synthetic examples in Fig 2 have constant cooling rates during the time interval of interest, so describing the FDHM as the timing of the peak cooling rate during 200°C of cooling is a bit misleading, because there rate is actually constant during the entire event. In this way, the FDHM value will not always document the timing of the "peak" cooling rate even in the simplest circumstances, like when rates that are constant for long periods of time; instead, it is the maximum rate, but also the sustained rate, which has different geological implications. Is there a particular style of tT history that the FDHM approach is well-suited for? If so, the manuscript would be strengthen by more discussion, and perhaps demonstration, of this.

AC: **It is unclear what the reviewer is discussing here and there seems to be some confusion. We never described the FDHM explicitly as the "timing of the peak cooling rate." The cooling rate is defined by the slope of the cooling path, as shown by our synthetic examples. We are not overly concerned with *rate*—only to say that very slow rates over long time intervals will yield a FDHM that will have less meaning if we are relating this cooling to some geologic event (see Fig. 4 and text discussion in the m.s.). We don't see that as much of an issue, since it would not make sense to interpret or apply the FDHM in that way. If someone were to suggest that 0.5°C/Myr monotonic cooling rates occurred over the past 500 Myr, and that there is a 'peak' cooling time associated with such a history—we would all naturally have questions about what that means exactly. Fast cooling rates will result in a case where the FDHM mathematically approaches or approximates the 'start' of cooling as well (i.e., at instantaneous cooling; also discussed in the text). In the course of this manuscript review/discussion it has become apparent that deep-time applications are the most suitable for the FDHM approach to avoid complications such as time lag, etc. (see RC3 comments/replies).**

RC3: Figure 1 nicely demonstrates why three cooling curves can be described mathematically by the same FDHM value, and the synthetic tT histories (Fig 2) demonstrate how this works with synthetic data and inversion results. But... how are the FDHM values derived from these models helpful for answering geological questions, if we get the same values for different thermal histories? Don't we want to be able to clearly distinguish among tT histories with such differences? Let's take the two synthetic tT histories. Cooling 200°C over 20 Myr (10°C/Myr cooling rate) suggests a radically different process, operating over an order of magnitude shorter timescale, compared to cooling 200°C over 200 Myr (1°C/Myr). But they yield the same FDHM value (and it's clear why). But, I'm left wondering: what do I learn from this FDHM 'cooling time'?

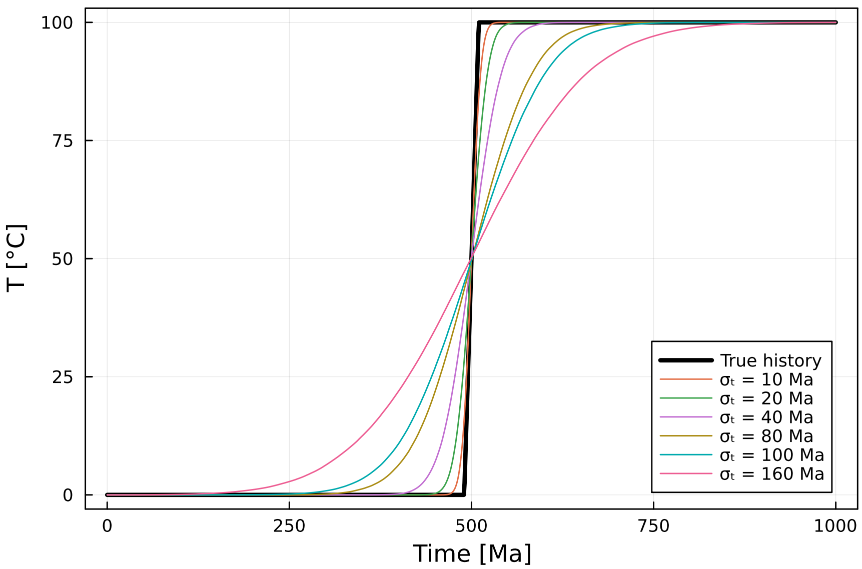
AC: That is because all three histories are characterized by the same total magnitude of cooling (but different rates). The message we intended to convey in Figure 1 is that 'cooling onset' which is usually the primary metric by which cooling paths are interpreted, is or can be problematic due to uncertainties in recovering the timing for the onset of cooling (refer to Fig. 4), as discussed previously. However, in all three cases shown in Figure 1, we can avoid this issue by focusing on the midpoint of cooling, which is the least affected by [thermochronological] uncertainty. We will address this to make it more clear. We understand the reviewer's point about different geologic processes with different inherent rates—but in what sort of real scenario would we be assessing 3 different 'real' thermal histories with different rates over the same time interval? If such histories were part of a group of modeled t-T paths from an inversion, then dramatically different paths with dramatic step/rate changes could possibly signify poor resolution of a coherent t-T signal, or other issues with input data, model assumptions/boundary conditions, or even make us question the need for any interpretive tools whatsoever.

AC: The importance of the uncertainty related to cooling history is shown in an example figure 1 below, which we can add to the revised manuscript. The cooling event in Fig. 1 occurs over 20 Myr (510 to 490 Ma) and 100°C. Uncertainty is shown by assigning a time for 'start of cooling' and time of 'end of cooling' but with Gaussian uncertainty (σ_t) on each path from 10 Myr to 160 Myr (Fig. 1a). Since the 'start' can't be after 'end,' the effects are asymmetric. In Fig. 1b, the uncertainty is estimated by randomly drawing 10,000 monotonic-cooling paths (only 250 shown; light blue) assuming that $\sigma_t = 100$ Ma. The mean of those paths is shown (equivalent to the cyan path $\sigma_t = 100$ Myr in panel A). This also shows that the effect of uncertainty is equivalent to convolving with a Gaussian.

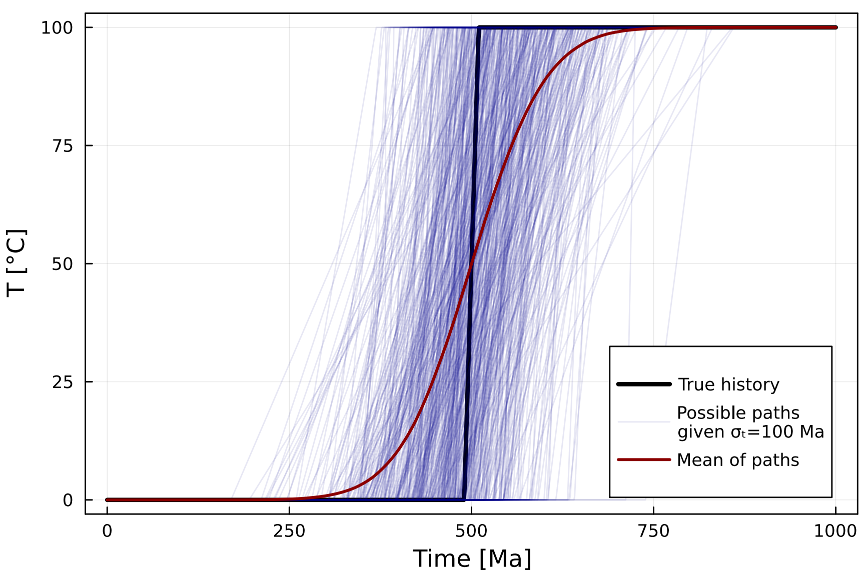
RC3: As discussed starting at line 60, cooling rates are one of the key ways we relate tT model results to geological processes. So, how is an metric that is agnostic about cooling rate, such as FDHM, helpful?

AC: We are a bit confused. The reviewer's previous comment above focused on cooling rate(s), but here they say the FDHM is 'agnostic' with respect to rate [which is true]. Thermochronological data inherently contain information that indirectly constrains rates, so 'rate' need not be a focus with the FDHM. The FDHM is just meant to quantify a 'time' based on t-T path behavior during an interval of cooling and the uncertainty related to estimating timing from t-T models. Since the reviewer focuses on start/end of cooling—we could also ask how is determining 'onset of cooling' helpful or useful if the onset is actually some fraction of a longer cooling path (i.e., 100°C resolved in model but 150 or 200°C in reality)?

RC3: A more compelling pitch would be useful. Arguably the 700 Ma FDHM 'peak cooling time' from the synthetic history with 1°C/Myr cooling rate has little geological significance; its just the middle of a protracted steady slow cooling event. In contrast, I can see how for more rapid cooling events, this 'peak cooling time' could be geologically significant, e.g., be used to relate cooling to some geological event (as in the real examples). However, the sentence at lines 70-72 seems to suggest that the FDHM approach is a good fit for regions that have experienced slow cooling; it is not clear to me why this would be true. Addressing some of these questions can be in part accomplished by describing what other thermal information one would need in order to make a robust geological interpretation of the FDHM 'peak cooling time', in addition to



(a)



(b)

Figure 1: The impact of variable time uncertainty for the start/end of cooling with respect to a true t-T path. Refer to text for explanation.

what is discussed in Section 3.4. The cooling rate? Other metrics for quantifying a cooling event from a model result? Is this only useful for rapid cooling? Addressing some of these questions would really help future users avoid using this as a shiny black box that spits out 'cooling times'.

AC: The middle of a protracted cooling event can still be geologically meaningful—but again, that is dependent upon context and what slow cooling rates are related to (see reply above). Arguably, 'slow exhumation' is one of the few ways to interpret 'slow cooling'. It could also be asserted that if our goal is to interpret total exhumation magnitude across some time interval,

then why would the midpoint of a cooling path be any less significant than the beginning or end?

AC: With respect to deep-time thermochronology, we sometimes don't have the resolution to know duration—but if we do know the duration of an event, then 'start + end' is the same information as 'midpoint + duration.' If cooling duration is interpreted as an "event" then the midpoint is arguably the most meaningful single number. People frequently discuss geologic events as a single 'age' and implicitly that's often a midpoint. We provide a few Earth history examples below (quotes shortened).

- "*The Wopmay Orogen, located in the Canadian Arctic, is considered to have formed within a relatively short time period, spanning roughly between 1.97 and 1.84 billion years ago ...* – Google search; Wikipedia and Brittanica sources
- "*The 1.9 Ga Wopmay orogen contains a deformed continental margin prism ...*
  – Hoffman and Bowring (1984) Geology v. 12
- "*... the Trans-Hudson Orogen ... represents the largest and most completely preserved segment amongst a number of Palaeoproterozoic collisional belts ... during the interval 2.0–1.8 Ga ...* – Corrigan et al. (2009) Geol. Soc., London, Spec. Pub. v. 327
- "*After the ca. 1.85 Ga Trans-Hudson Orogen, the Laurentian core (present-day North America), underwent a prolonged period of arc terranes accretion ...*"
  – Volante and Kirschner (2024) Scientific Reports 14; 6483

RC3: Finally, in a case where the FDHM cooling time is geologically significant and one wants to relate this cooling time to some event or process... how exactly should we be thinking about this relationship? The challenge of translating thermal information back into a geological framework has been a core interpretive challenge for low-T thermochron for decades now—cooling start, end, and rate are also not entirely intuitive to relate to processes like erosion, but the community has just spent a lot of time on this problem such that the fundamentals are well established—and the readers are going to be accustomed to thinking about this in the common approached used to characterizing thermal events. For example, hypothesis testing with thermochronology can leverage start/stop times of processes that produce cooling signals, for example 'cooling started after 5 Ma in the model result, and therefore the start of process X, which we know started at 15 Ma from independent information, cannot be responsible'. So, what should we do with 'peak cooling timing' information? It seems to require a different approach to hypothesis testing and conceptual models about the thermal consequences of various processes. I think we have much less practice thinking about this, and some guidance from the authors would provide a solid foundation to build upon. Also, importantly, what do that authors think we should not do with FDHM peak cooling timing?

AC: Relating aspects of thermal histories to geologic processes is certainly foundational to thermochronology. Certain fundamentals are established but there is room for different ways of thinking, changing methodologies, or improvements. We aren't saying that the FDHM method is groundbreaking, but it is a different way of approaching how we think about thermal history interpretation. The idea of "hypothesis testing" (strictly with respect to thermal history

modeling and interpretation) has become problematic, but is beyond the scope of this reply. In the example mentioned about 'cooling started in the model by X Ma' the FDHM approach is potentially just a more accurate way to estimate when cooling is happening, especially in deep time. We can't really speculate on possible abuses of the FDHM before they occur.

RC3: The way the half-maximum isotherm is determined from an inversion result is not explained, and it also appears to lack the quantitative rigor of the other elements of the FDHM calculation and the consequences of this are not discussed.

AC: This is discussed in the m.s. but we will elaborate in the revised text (see below).

RC3: A key piece of the FDHM approach is the need to quantify the 'total cooling magnitude' of the cooling event of interest, in order to identify the half-maximum isotherm that is then used as the reference point for quantifying the timing of cooling through that temperature. This manuscript would be more impactful, and the method more accessible to potential users, if this part of the workflow and its potential limitations were discussed in additional detail. The results of modeling synthetic data produced from known ('true') tT paths presented in Figure 2, and the discussion of those results at lines 191-204, exemplify how this part of the workflow could be better explained. Line 196 simply says: "A key feature of the inversions is that the total cooling magnitude was accurately recovered," but it does not then compare some total cooling magnitude retrieved from the inversion result to the 'true' cooling magnitude. Instead, the analysis goes on to use the total cooling magnitude from the 'true' tT histories (exactly 200°C, from 240-40C) to determine the isotherm of interest (140°C). However, if one looks at the actual inversion results, and pretends the true history is unknown, the total cooling magnitude is much less clear and does not seem to be not well represented by a single value with no uncertainty. For example, at the end of cooling, the high relative probability region in both examples spans at least 20°. At the start of cooling, the slow-cooling example is a similarly narrow range of high relative probabilities, but the fast cooling example has no high relative probability region around 240°C. Of course, such features are not unique to these results; even the best-constrained inversion result with perfect kinetics and a model design tuned to successfully reproduce the 'true' history will not resolve that exact tT history with no uncertainty.

AC: We respectfully disagree with the reviewer on the characterization of the Figure 2 results. The whole reason for the synthetic tests was to produce a known t-T path with some inherent uncertainty* and demonstrate only that the method 'works'. Of course we could make the midpoint isotherm estimation (with respect to total cooling magnitude) as simple or as complex as we want, but we would prefer to it remain simple. The high relative probability region of the models spans 10–15°C at most, and high probability region 'absent' in the fast cooling example is an artefact of interpolation (and smearing) due to fast cooling. This may be something that needs addressed, at least visually. Nevertheless, the *highest* probability regions in both models spans 240–40°C. For the t-T path ensemble, the apparent turning points (min/max) and the high relative probabilities are used together.

AC: Lines 190–195 discuss how the half-maximum isotherm was calculated. For the 10°C/Myr

*in line with the reviewer's last sentence of this comment block "...perfect kinetics and a model design tuned to successfully reproduce the true history will not resolve that exact tT history with no uncertainty" As we stated in our reply to RC2 comment, if we know the 'true' path perfectly, we don't need the FDHM.

example, if we pretend and say that cooling is instead from 230°C to 45°C (not 240–40°C), then total cooling is 185°C ÷ 2 = 92.5°C half maximum cooling amount—the FDHM at the 137.5°C isotherm = 700.0 +6.6/-5.1 Ma, or within uncertainty of the true value of 700 Ma and nearly identical to the FDHM quoted in the manuscript. Alternatively, 260°C (max) to 35°C (min) is 225°C of cooling, making 112.5°C half of the total. A FDHM at 147.5°C (or 260-112.5°) = 704.0 +3.8/-7.2 Ma. We could make a workflow revision where perhaps three estimates of the total cooling magnitude are made to account for any uncertainty (FDHM calculated for each), and then those temporal distributions could be averaged—similar to the discussion in Section 5. If we do this for the scenarios discussed above and in the manuscript we get a mean FDHM of 701.4 ± 3.6 Ma (1σ) or 701.4 +7.4/-7.6 Ma (95% CI). Obviously even 10–20°C of 'half-max isotherm' uncertainty will be trivial in fast cooling scenarios, since the slope of the cooling curve (i.e., rate) dictates the apparent offset in FDHM timing if different half-maximum isotherms are chosen. The opposite would be true for slow-cooling examples. This may be a good thing to add and show visually in a figure cartoon.

In addition, the user can report or specify what the min/max temperatures (and total cooling magnitude) they are using to calculate the FDHM. This could be assessed and agreed/disagreed upon by anyone inspecting the results. Of course we could make this more complex by trying to find some average of the turning points at high/low temperatures in the model, but this is a nontrivial set of calculations.

RC3: So, several question arise: How exactly does one figure out the total cooling magnitude from an inversion result as a part of the FDHM method? Visual estimation? assessment of the relative probability space shown in Figure 2? Assessment of the ensemble of individual tT histories used to build that relative probability heat map? Does a difference of a few tens of degrees matter for the half-max isotherm, and under what circumstances? If the FDHW value isn't that sensitive to the choice of half-maximum isotherm, then that is important to demonstrate and discuss why. In any case, the approach that the authors used to determine the cooling magnitude and isotherms from the inversion results for all the examples in this manuscript needs to be more completely explained so their approach can be understood and replicated by others. And, perhaps most critically for the future use of this tool, why is uncertainty on this cooling magnitude not accounted for in this method? The inversion result for the 10°C/Myr cooling rate would be a natural vehicle for discussing this.

AC: See reply above, we can address this in revision by more clearly showing and explaining how the calculations are performed and accounting uncertainty on the cooling magnitude by averaging multiple FDHMs.

RC3: The importance of paying careful attention to the temperature sensitivities of the chronometers being used to produce the inversion result is nicely emphasized in Section 3.4, which uses one of the synthetic examples to explore what happens to a FDHW assessment when a cooling event starts at temperatures higher than the chronomter(s) are sensitive to. However, the same considerations should apply at the cold, low-temperature end of cooling events, too, and this is not discussed. A similar assessment of a lack of low-T sensitivity could, for example, be demonstrated by only modeling a high-T system in the same synthetic example. More pointedly, I think the Bergell pluton example highlights the potential perils of not exploring this more completely. 6.45 Ma AHe ages are input into the model, and the FDHM approach is used with a

30°C isotherm to infer peak cooling at ca. 1 Ma. However, 30°C is well below the temperature sensitivity of the AHe system, and it is not clear how or why the AHe system would be sensitive to the exact shape of the thermal history after the rock exits the PRZ at 50-40 °C; to me this seems like a function of the QTQt model design, not the data. Although the authors simply say this is an 'interesting' result and note the temporal correspondence to previously hypothesized accelerated cooling at a similar time, it is very unclear whether this FDHM is a rigorous result that is capable of supporting such a hypothesis (or rejecting the other proposed times of peak cooling). Some additional discussion of this is warranted as a part of demonstrating this new method.

AC: We should note that we are focusing on low-temperature thermochronology and not high-temperature methods, however, low-T sensitivity is primarily a concern for estimating the total cooling magnitude, as previously discussed above. We discussed in the m.s. (lines 290–300) that the Bergell example is potentially showing a certain cooling path due to waning sensitivity and the requirement of being at the surface at time zero. Importantly this is not a "a function of the QTQt model design" but simply a function of any t-T model software requiring surface temperatures for present day. If anything, the QTQt model design was entirely data driven, with the exception of the high-T constraint (surface T being an explicit boundary for nearly all t-T models). A sidenote is that while what the reviewer states is true about sensitivity of the AHe system, rocks nonetheless are at the surface today—so while there is certainly complexity surrounding what is potentially controlling the cooling path trajectory, those rocks are currently at the surface. Plenty of published papers interpret thermal histories either 'above' or 'below' the limits of thermochronometer sensitivity, so we do not think that is something somehow inherent to the models discussed in our manuscript or due to any features of QTQt or any other model software. One of the main thrusts behind the FDHM is to put a little more quantitative framework around estimating and interpreting features of thermal histories that are often done entirely by eye or estimated in some unknown manner. At least with the FDHM, all calculation steps can be traced and explicitly shown and the process is reproducible by anyone with the t-T model output. We have decided to remove the Bergell example, partially for the reasons the reviewer mentions, but also because the FDHM is more suited to problems in deep time. This is discussed more in response to RC3 (potential complications due to thermal lag etc).

RC3: Several other aspects of the FDHM method and workflow would benefit from being more completely explained. This information is very succinctly described in lines 129-135, tucked into section 3.1. A few suggestions: The text comparing FDHM and "full width at half-maximum" FWHM (lines 120-124) is confusing. How is the FWHM's "width of a distribution" different from "the full distribution rather than merely a width" used for the FDHM? The FWHM inspiration for this approach is an intuitive metric because it is easy to visualize the connection between the shape of a (spectral) curve and the FWHM. I suggest further embracing this 'spirit' by using at least one figure to more clearly illustrate the connection between [1] what one sees on an inversion result in tT space (such as a cooling event of interest in an Expected Thermal History plot) and [2] the FDHM results (such as a PDP histogram), as well as [3] what the code (and user) does to get from [1] to [2], emphasizing the role of variables that may need to be "tuned" or chosen by the user vs. what is build into the program.

AC: We appreciate this suggestion and will make these changes to provide better context and detail.

RC3: Which QTQt outputs are used in this method? How are they loaded into the FDHM code? Certainly not all related details are necessary, but a quick description of the workflow and specific direction about where the reader can find additional documentation is needed.

AC: The QTQt thermal history output is read into the FDHM code. Simply take a QTQt model file (.txt) and it gets read in by the Julia code. The Julia code only needs the file name. We encourage the reviewer to try this and see if more explicit and/or clear directions should be annotated in the Julia file on Github.

RC3: "Peak cooling" vs. "time of cooling" vs. "peak cooling time" are all used interchangeably in the title and abstract, but their meaning is vague without a more clear definition at the very beginning of what exactly this approach quantifies. A more explicit and comprehensive description is important because the FDHM approach is quantifying something different than features of tT inversion results that are typically quantified. I suggest re-writing the sentence that starts in line 5 "This study presents a straightforward methodology wherein we ascertain the time of peak cooling for the entire cooling signal within a thermal history model." to include a clarification of what is meant by 'peak cooling' and 'the entire cooling signal'. Additionally, the authors may consider incorporating this detail into the manuscript title.

AC: We can make those changes and provide more detail.

RC3: The paragraph starting at line 74 (section 2.2) claims that the commonly used ""onset of cooling" for a thermal history is an unreliable metric," but provides no clear support (references, model results that demonstrate the specific "unreliability" described, etc) for this characterization. The last sentence in the caption of Figure 1 offers a similarly unsupported assessment, which also doesn't seem to be related to or demonstrated by Figure 1. I suggest that the authors either clearly support claims about the limitations of the "onset of cooling" (line 80) and other common metrics, or simply remove these claims from the paper and focus instead on better articulating the strengths and limitations of the FDHM method, with specific demonstrations of how FDHM compares to other methods only as needed.

AC: Figure 1 and Figure 4 clearly show this idea that the onset of cooling can be 'unreliably' recovered—if maximum temperatures exceed the sensitivity of the thermochronometer, then what is the onset of cooling in a t-T model (this was discussed previously as well in our reply)? The AFT retention age analogy (lines 136–148) is a good reference for the idea of sensitivity onset, along with Figure 4 in the manuscript. We are also going to add a figure/panel to that showing Gaussian blur of a path and how the start/end of a path are most heavily affected, whereas the midpoint is not. We also could pull plenty of material from the published literature, but that isn't really something we want to do, nor is it what this paper is about. The last sentence of Figure 1 is simply referring to the convergence of the posterior in QTQt, which is evident because the model well reproduces the true t-T path. If the Markov chain did not run long enough, this wouldn't be the case. We can remove this sentence if necessary.

RC3: Although I broadly agree with the sentiments expressed in the paragraph that starts at line 28, I suggest some moderate revisions that would better represent previous work and situate this new contribution in that context. I think it would be useful to briefly distinguish between thermal history models (like QTQt and HeFTy, when they are used for single-sample analysis) and thermokinematic models (such as Pecube). The latter have thoroughly discussed methods for quantifying "the timing of cooling within inversions"; see for example the method papers lead by Braun cited in line 24. "a systematic approach to defining and quantifying the time of cooling within inversions has to our knowledge never been thoroughly discussed in the literature." It depends what is meant by "thoroughly discussed," and I suggest tweaking this language to better articulate the problem at hand. Many studies present and use (commonly one-off) quantitative approaches to quantify a particular feature of a tT inverse model result, for a particular scientific question. I absolutely agree that such previous approaches are almost always presented as a part of papers that are focused on geology, not on modeling methods, and are certainly ad hoc in the sense that they are commonly designed for a specific study, not more broad use. As a result there has been little discussion of the merits of various approaches for extracting information from inversion results, how they compare and contrast, etc. And as a community it can feel like we are constantly reinventing the wheel. So, this new manuscript can provide a substantial contribution, by extracting such an approach from a recent paper (McDannell and Keller, 2022) and demonstrating how it works more broadly in a stand-alone manuscript. I think this is fantastic, and we need more of this kind of paper to help document these tT inversion interpretation tools and make them more accessible in our modeling toolkits.

AC: This bit of text was meant simply to mean most t-T model approaches are ad hoc and "one-offs" -to use the reviewer's words. That is fine and often we have to tailor tools to specific problems, and not every tool will work in all situations. The FDHM is included in that. The main issue with going into great detail about using X or Y software for single-sample analysis is that this is a very contentious issue in the thermochron community in terms of models/methods and interpretation. The real sticking point is interpretation. We think everyone can appreciate that if you model a single, low-T thermochronometer, in many cases you can't say much about high temperatures (without any other info)—but this does not prevent such cases from making their way into the published literature. That is sort of the point of Figure 4 as a cautionary example. While we would like to weigh in on these topics, this isn't the forum for that and we do not want to get into discussions about modeling software (HeFTy vs. QTQt) or Pecube. That ground has been trodden with mixed results. We can potentially cite these sorts of papers as examples, but examples of what exactly, without getting too deep into the weeds? Even the mention of the influence of t-T "constraint boxes" in our manuscript seemingly set off alarms. We think, as the reviewer mentions above, that the thermochron community can generally understand, and knows that there are no real systematic approaches for t-T interpretation in inverse models that are universal and they are mostly (as the reviewer states) "certainly ad hoc in the sense that they are commonly designed for a specific study, not more broad use". The FDHM is still ad hoc to a degree but it can be systematically applied in situations that meet certain, specified criteria, outlined in one of our first replies above. We agree with the reviewer's sentiments about getting more widely applied, systematic tools out there for community use.

RC3: "The ad hoc method most often used is a visual estimate—where cooling initiation is typically framed with respect to a specific temperature at the time of maximum reheating preceding cooling—in other words, the point where there is a change from heating to cooling." I agree, many "ad hoc'" methods do have a visual component—indeed, a qualitative visual assessment in tT space is a common starting point, and it is also common to illustrate tT features qualitatively in figures (in fact, above I am suggesting this paper do more of this illustration). However, such approaches can be (and in many cases are) accompanied by a quantitative component. The authors may find it useful to refer to Murray, K.E., Goddard, A.L.S., Abbey, A.L., and Wildman, M., 2022, Thermal history modeling techniques and interpretation strategies: Applications using HeFTy: Geosphere, v. 18, p. 1622–1642, doi:10.1130/ges02500.1. (for example, Fig. 8).

AC: More qualitative examples are a good idea and those will be included. We have looked at Murray et al. Figure 8 and it does blend qualitative and quantitative elements. Arguably, the framework or manner in which t-T paths are generated in HeFTy (i.e., number of nodes/segments allowed and gradual/intermediate/etc path changes) can highly influence results—either directly or indirectly. But more importantly, those model run options are explicitly set by the user a priori, which can be problematic or even biased due to 'interventions.' Panel C of that figure is interesting in the context of our manuscript. Their Figure 8C zooms in on the last 20 Myr of a longer >100 Myr set of t-T histories to examine the "range of uncertainty for cooling initiation" over about 100°C –showing that 7 Ma is the "earliest possible start of required cooling" and 3.5 Ma is the respective "latest required cooling must initiate." That is obviously a snapshot and could change with more model run iterations. This is a reasonable approach but does illustrate what we talked about in terms of 'cooling onset' estimation bias. One thing we can envision is that if the number of HeFTy path segments is changed (i.e., decreased could force the model to potentially start a path turning point earlier, or later, depending on the temperature at which the data are reproduced) or if the number of model iterations were also changed, it will most certainly change the range of times for cooling initiation. HeFTy models in the literature are also typically run for a fraction of the total iterations compared to QTQt (10^4 vs. 10^5 or 10^6, respectively), and model termination is usually based only on X # iterations reached, rather than X number of Good/Acceptable paths found. The latter being the better option to produce a robust t-T ensemble that may potentially mitigate the effects of variable path segmentation. Segmentation of t-T paths is of course random, but in general, if histories are made coarser with fewer allowed t-T points, then there will be compensation for that enforced simplicity somewhere else in the model—likely in regions of t-T space where there is low resolving power†—or in this case, just above the sensitivity of the low-T chronometer. This just so happens to be where 'cooling onset' would normally be identified and interpreted. However, what will be unlikely to change is the actual cooling path itself, more specifically the midpoint of that cooling path. It is clear in the model that ca. 60–50°C is the most stable and well constrained portion of their t-T ensemble.

RC3: TABLES Table 1: Please report the Ft-corrected He ages here. Corrected ages are the geologically meaningful He ages; they are directly connected to the timing of cooling in the 'true' forward paths, the inversion results, and FDHW timing discussed in the text and presented in the figures. Currently, it is confusing to see a table of AHe ages <500 Ma when these rocks have been colder than 40°C since 600 Ma. I also suggest reporting the grain sizes for the He systems, since they were not kept constant.

†this can occur in QTQt as well but the reversible jump MCMC is not constrained by an explicitly fixed number of t-T nodes or path segments

AC: We would argue that the uncorrected ages are those actually 'measured' as well as what are modeled in the software — alpha-corrected ages are a further step and not really more/less meaningful in a geological context, since cooling ages do not necessarily carry intuitive t-T information upon visual inspection. This is discussed further in the reply to RC2. Nevertheless, we are going to remove Table 1 from the paper and make it an appendix; grain sizes are reported. It is not necessary for understanding the work and was included only for completeness.

RC3: FIGURES All Figures: please use letters to identify panels in figures that have more than one plot

AC: We will fix this.

RC3: Figure 2. This and other similar plots would be more readable if the legend was expanded to include: a label on the color bar indicating the heat map is for relative probability, and an entry indicating the dotted line is the 'true' forward model path. Consider overlaying the 95% credible intervals on these expected thermal history plots, especially if these CI are used to define the total cooling magnitude or FDHM

AC: We can label and adjust the legend. The 95% CI isn't currently used but could be incorporated in some way, but also isn't necessarily required for the t-T modeling interpretation. We are going to remove the 'younger' case study examples. Again, those models are just a more realistic reference that include some uncertainty to test the FDHM method.

AC: The separate comments in the PDF annotation made by RC1 primarily echo many of those here and are addressed in the replies in this document. We are removing the sections on the Bergell case study that make up the majority of the other comments.