

We thank Kendra Murray for the thoughtful, supportive, and constructive comments on our manuscript. In this document, reviewer comments are replicated in black, and our replies are in blue. Following the request from *GChron*, at this stage we indicate only the changes we intend to make in the revised manuscript, which will be submitted soon.

This manuscript presents a new open-source Python package for predicting cooling ages for AHe, ZHe, and AFT chronometers from time-temperature (tT) histories extracted from the results of geodynamic numerical simulations. In particular, it addresses the challenge of the large number of tT paths that are produced by such simulations that need to be forward modeled in order to validate geodynamic simulations with thermochronologic observations. Using established approaches for numerically integrating diffusion and annealing behaviors along a particle tT path, this modeling package can take in outputs from common geodynamic modeling software. It also addresses some of the known problems with extracting specific particle tT paths from such numerical results using an interpolation scheme. This contribution explicitly presents an open-source starting point for developing tools that are capable of jointly leveraging the power of geodynamic simulations and thermal history analysis of low-temperature thermochronologic data. As such, some features (such as the implementation of thermochronometer kinetics models) are overly simplified, with the invitation for community-generated innovation and updates that can implement more complex and realistic behaviors as needed.

The central challenge that this new Python package is designed to address is interesting, important, and relevant within the scope of *GChron* and the Special Issue it was submitted to. The accompanying online documentation and demonstrations are a laudable companion resource. Creating accessible tools to bridge the divides between the geodynamic and thermochronologic communities has the potential to support advances in multiple geoscience disciplines. However, as a thermochronologist, there are a few things about the design and description of the proof-of-concept examples presented here that I think would benefit from some adjustment, which I describe below. I think with revision, this will be an excellent contribution.

We warmly thank the reviewer for their support of our contribution for inclusion in this special issue for *GChron*.

MAJOR COMMENTS

1. Before walking through specific examples, I have one general comment for all figures and corresponding text: the axes on many graphs are presented in the opposite sense from what is typical for thermochronological data and exhumation studies, which overcomplicates the presentation of the numerical results. In most cases, this is because the depth (y position) and time (elapsed model time) information is being directly exported from the simulations without being converted to and presented in a geological reference frame. These conversions are simple, but for thermochronological data, this matters! It matters because depth below the surface corresponds to closure behavior, and cooling ages are in Ma (millions of years before present) not model time (Myr elapsed). Users (and readers) need to be able to efficiently evaluate if the numerical results “make sense” thermochronologically. I provide some details below, but overall I suggest that

unless it is essential to present the modeling reference frame, all the results for each example be reframed as depth = depth below the surface and time = geological time (Ma). I realize that following these suggestions may require parallel updates to the very nice online documentation the authors have prepared; however, I also anticipate that just a few simple lines of code are required to flip axes or convert model depth to geological depth.

We will be happy to revise all figures that employ Y position so that they indicate depth below the surface. We will also reframe all references to model runtime in Myr to be geologic ages in Ma.

2. Forward modeling example (Section 2.3)

1. The forward modeling demonstration has a strange starting condition from a thermochronologic perspective that is not explained. The modelled tT path starts at 100 C, but it is used to predict ages for thermochronometers that are completely (ZHe), partially (AFT), or not at all (AHe) closed at 100 C. The manuscript should either: be explicit about the assumptions of such a model design and describe clearly how the consequences of these assumptions produce the model results; or, have a starting condition that is hotter than the temperature sensitivity of all systems being modelled. This is not sufficiently discussed in the current text. For example, at line 95, the text mischaracterizes the reason why the ZHe age is predicted to be 34.8 Ma. “For the zircon (U-Th)/He (ZHe) system, this path results in an age of 34.8 Ma, corresponding to nearly the full duration of the 35 Myr history, given that the sample remains at a temperature *largely below the ZHe partial retention zone* throughout this history.” (emphasis added). The starting condition (and isothermal hold T) is not “largely” colder than the ZHe partial retention temperatures for the Reiners et al. 2004, it is much much colder (at least 50C colder) than even the coldest part of the PRZ for these kinetics (and for the current ZRDAAM kinetics too, for such as tT history). From a thermochronological perspective, this zircon grain simply “appears” at 100C at 35 million years ago, and accumulates He that whole time. Having the model start in the midst of the AFT PRZ is even more complicated and potentially problematic from my perspective. Unless this is intentional? If so, why?

This choice of example model was intentional because of the behavior the reviewer describes, and we will be sure to revise the manuscript to make the rationale for this choice more explicit.

We chose this example because it produces ages in each system that intuitively make sense given the variations in their partial retention zones (PRZ) or partial annealing zones (PAZ). As the reviewer states, ZHe is completely closed at 100°C and accumulates He throughout entire the modeled t-T path, AFT is partially closed and so has a complicated age that does not correspond to a particular thermal event, and AHe is not at all closed and so has an age that corresponds to the prescribed cooling event. As a result, all 3 systems have distinctly different

ages, and the explanations for those distinctly different ages are intuitive given what is known about each system's kinetics.

We will remove the word “largely” from the discussion of the ZHe PRZ.

2. When I put this tT history into HeFTy (v 2.3.1), using the Ketcham 1999 annealing model, a Dpar of 1.75, and c-axis projection (and Ketcham 2003 c-axis projection) it predicts an AFT age of 19.8 Ma, not 21.3 Ma. (The predicted track-length distribution (Fig. 1c) is not reported in numbers so I cannot compare that to HeFTy's prediction.) Section 2.2 describes that Ketcham's 2005 approach (i.e., HeFTy) is being used for the AFT system here, but was this new code benchmarked against HeFTy? If I'm not setting up the AFT correctly in HeFTy to mirror what is being done here, then more information is needed. Such a difference in ages may not be geologically important, but from a numerical calculation perspective it is essential to know why this discrepancy arises. Starting the history from within the AFT PRZ makes this discrepancy additionally difficult to diagnose.

GDTchron is benchmarked against HeFTy, using the example t-T paths that are provided by Ketcham (2005), and there is very good agreement between GDTchron and HeFTy for these examples. In fact, these examples are currently used as unit tests within GDTchron, such that no changes to the source code can be implemented if they result in ages that do not match the HeFTy ages. We will revise the manuscript to indicate this benchmarking.

We agree that there is a small discrepancy between HeFTy and GDTchron for this particular AFT age in our example t-T path, and that the history beginning within the AFT PAZ likely plays a role. It is difficult to diagnose the exact numerical reason for the discrepancy because HeFTy is not open-source software, and so we cannot directly compare our routines with its source code. One possibility is that the time intervals we input into GDTchron for this this example do not precisely align with the time intervals HeFTy uses for its calculations of annealing behavior. As described in our response to Reviewer 1, we are aware of a separate community effort to create a Python repository of thermochronology kinetics that we think would make it easier to assess the causes of such discrepancies.

3. Simple “uplift”

1. This scenario is repeatedly described as “uplift,” but thermochronologic ages never directly document uplift. They only document rock cooling, which in this scenario is driven by rock exhumation that happens when the imposed *rock* uplift occurs without any corresponding *surface* uplift (in other words, rock exhumation). Thus, this is a “simple exhumation” scenario, and should be described as such, starting in the abstract. I suggest never using the word “uplift” without specifying “surface” or “rock”, to avoid this confusion. See England and Molnar, 1990 (Surface uplift, uplift of rocks, and exhumation of rocks: Geology, v. 18, p. 1173–1177)

This is a good point, and we will be sure to revise the manuscript to be very explicit about when we mean surface uplift and when we mean rock uplift. The reviewer is correct that this model is better described as “simple exhumation” since there is rock uplift without surface uplift, and we will revise the text to reflect this throughout.

2. The model design description at lines 168-170 states: “At 10 Myr, the bottom of the box is pushed upwards at a rate of 1 mm/yr, allowing material to flow out the top of the box while the temperature structure of the box remains constant,” If I understand correctly, in the geodynamic model used to generate the tT paths modeled here, a simple linear geothermal gradient was held constant despite a massive transient change in rock exhumation rate, from 0 km/Myr to 1 km/Myr and then back to 0 km/Myr. By imposing rock uplift and exhumation without any corresponding change in the geothermal gradient, this example is an extraordinarily unrealistic departure from how we know the Earth works. Especially considering the imposed rapid exhumation rate: the Peclet number for this scenario [$Pe = (\text{exhumation rate} * \text{lithosphere thickness}) / \text{thermal diffusivity}$, which describes the competition between heat advection and conduction] is close to 1 (assuming a standard thermal diffusivity of 25 km²/Myr), and so this is nearly an advection-dominated system that would not even be well represented by using a steady-state solution of the conductive geotherm for a 1 km/Myr exhumation rate. These considerations and their consequences for cooling age patterns have been quantified and evaluated by the thermochronologic community for decades (see for example the “Quantitative Thermochronology” book by Braun et al published in 2006). Given (my admittedly under-informed sense of) the abilities of geodynamic models, I’m surprised at this oversimplification, and surprised that the thermal limitations of this geodynamic result (and the corresponding age predictions) are not acknowledged. The main point of this manuscript is to demonstrate how the new tool addresses some of the major challenges of transferring particle-path tT predictions from geodynamical tools to thermal history analysis tools. And there will necessarily be some simplifications. However, given that the audience of this Technical Note spans both thermochronologists and geodynamicists, I think it is essential that the examples use simplified, but still physically reasonable, thermokinematic scenarios.

We agree that this is an unrealistic thermal regime for the reasons the reviewer describes and will revise the manuscript to make that more explicit. We note that this example model is not intended to demonstrate the full capabilities of geodynamic models, which typically do employ much more complex and realistic temperature models, but rather to illustrate how the results of a geodynamic model, regardless of complexity, can be output to GDTchron to predict ages. In this case, a model with a static linear geothermal gradient is useful precisely because the expected thermochronometric ages are easy to predict and understand for those who are not experts in thermochronology (which is also a component of our target audience). Such users may not have thought about Peclet numbers, but

they can understand that if you know roughly at what depth particular isotherms occur, then you can connect cooling ages to advection of material from particular depths.

We note that the subsequent rift-inversion model is a much more complex geodynamic model with a more realistic temperature model like the one the reviewer describes. We will make this more explicit in the revised manuscript.

3. Is 10 Myr sufficient for these thermochronometers to achieve equilibrium at partial retention/annealing temperatures for each system? In other words, how arbitrary is this choice of duration, and does that matter for this demonstration?

10 Myr is a fairly arbitrary choice that is mainly used so that there is an opportunity for He/FT to accumulate both above and within the PAZ/PRZ prior to the rapid “exhumation” event. We will make this rationale explicit in the revised manuscript.

4. The text should be more clear about reporting ‘model time’ (Myr elapsed since the start of a simulation) vs. ‘geological time’ (Ma, what the cooling ages are documenting and how we think about geodynamic histories); see also Major Comment #1 above. The use of Ma vs Myr is consistent and correct, but the practical difference between these concepts is not explained. For example, Figure 5 is mixing these two opposite concepts. My suggestion: Unless it is essential for the reader to see the ‘model time’ framing of the design of these numerical simulations, I suggest the authors simply convert everything into ‘geological time’ for us. This matters because thermochronologists always start examining data (whether synthetic or real) by comparing cooling ages to intervals of geological time that are of interest. In this simple example, I’m looking to see how the predicted cooling ages at the end of the model run (“20 Myr – After Quiescence” in Fig. 5) document (or not) the imposed exhumation event in order to assess the rigor and utility of this approach. But, rather than just looking at Figure 5, I have to do the extra work of converting the geodynamic model time to geological time, or vice versa. (Geologic time = How long ago did the exhumation event start and end?) I also have to know, as a reader, that this conversion is necessary.

We will revise the manuscript to report model runtimes as geological times in Ma to reduce the potential for confusion.

4. Complex model, Figure 6

1. For the “AHe Ages” plots, why is the result shown for 150 km thickness? It seems like He ages are only calculated for the top 50 km, and even that is an order of magnitude greater lithospheric thickness than these ages are sensitive to. The relatively near-surface details (between the surface and the effective closure of each system) is what is important for assessing the utility of this tool, but that cannot be seen here.

These plots are shown at the same scale and covering the same area of the model domain as the top row, so that it is easy to directly compare other outputs of the model (such as location of shear zones/faults) with the distribution of thermochronometric ages. Because that makes it difficult to see the patterns at the surface, we include the bottom row of plots to show the surface ages across the model domain.

He ages are only calculated for the top 50 km because, as the reviewer states, anything deeper would have no change of every actually accumulating He. We use 50 km to conservatively ensure that we capture any material that might have entered the PRZ sometime during the 36 Myr model runtime.

2. For the “Surface Ages” plots, why is age increasing down the y-axis? This is the opposite of how such data are usually visualized, including in similar tools (see for example, Fig. 6 from this paper by McQuarrie et al: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018TC005340>)

We will revise these plots so that age is decreasing down the Y axis.

5. Description and simplification of the thermochronometer kinetics models
 1. Sections 2.1 and 2.2 would benefit from reporting the nominal closure temperature ranges for each system, given the kinetics being used in this version of GDTchron.

We will revise these sections to include approximate closure temperatures for these systems.

2. Simplifying the diffusion and annealing kinetics models being implemented in this first version of GDTchron makes sense, and I support the approach taken here as a first step (so long as it is not the last). But, I think this choice warrants more discussion up front, in the methods section of the paper, to help convince thermochronologists of the basic utility of the current GDTchron package despite the limited kinetics, highlight its potential for future adaptation, and describe the consequences of simplifying kinetics in this way for those who are not familiar with chronometer kinetics. For He diffusion kinetics, for example, the 2000 AHe kinetics parameters and 2004 ZHe kinetics parameters being used here are extremely out of date; from the perspective of thermal history analysis, these kinetics models have been superseded by the radiation damage accumulation and annealing models and the old “static” kinetics parameters should never be used to interpret real thermochronologic data, especially for scenarios with certain styles of thermal histories. But of course, the choice of kinetics model only ends up being important in some contexts. So, I suggest that the authors move much of the information presented at the beginning of section 5 (lines 250-260) into section 2.1 and 2.2. The authors could support this choice by discussing how other numerical tools (e.g., Pecube, age2exhume) have navigated this challenge, and in

what types of thermal histories / geological scenarios not using the radiation damage kinetics models matters most.

We appreciate the reviewer's acknowledgement of these kinetics as a good first step, and we agree that employing more updated kinetics in future versions will increase the utility of GDTchron. As described in the response to Reviewer 1, we anticipate that a parallel community effort to compile kinetics in a Python repository will provide an opportunity to integrate these kinetics into GDTchron without a large duplication of effort. We will move information from section 5 into section 2 in the revised manuscript to better clarify the consequences of this initial choice of kinetics, as well as the scenarios where these kinetics are still appropriate. We do not necessarily agree that these kinetics should never be used to interpret real data in cases where there is reason to expect monotonic cooling and where there is a strong correlation between grain size and age, for example, but we acknowledge that in many scenarios they are not the preferred choice.

MINOR COMMENTS

Figure 1b: Can the user adjust the grain size and composition? Can different kinetics be chosen?

The user can assign grain size and composition, though in the current version they apply only to the entire model domain. We will make that explicit in the revised manuscript. Different kinetics could be chosen once they are implemented in future versions.