We thank reviewer David Whipp for his constructive and helpful comments on our manuscript. Below, we provide some responses and outline how we incorporate the comments (reviewer comments in blue, our response in black; changes made to the manuscript in *italics*):

Summary

van der Beek and Schildgen presents an overview of a numerical model written in Matlab for the purpose of rapidly calculating estimates of rates of exhumation from regional thermochronology datasets. Their new software uses a steady-state analytical solution to the heat transfer equation with fixed temperature boundary conditions in combination with analytical solutions for estimating thermochronometer ages, which allows for efficient calculation of effective closure temperatures to find exhumation rates that produce the input age(s). The model aims to fill a void in the available options for interpreting thermochronometer age data, providing estimates of exhumation rates for large thermochronometer age datasets (in contrast to models designed for calculating only thermal histories), but with less complexity and computational overhead than other popular software. They demonstrate both the models use for exploring exhumation rates in large datasets and its sensitivity to changes in the input parameters using a large age dataset from the Himalaya, showing major observations from more detailed, local studies are also captured in their model. Overall, the manuscript is clear and well written, and it will certainly appeal to readers of Geochronology. However, there are a few places where the text may be able to be improved in revision, which are detailed below.

Main comments

1. There are several places in the text where a choice in the design of age2exhume is suggested to be better than equivalent choices in other models, such as for the boundary conditions. While the authors do justify their reasoning for making such suggestions, model design choices are often based on what is most valid for the physical system being studied. Choosing to use a half-space thermal model with a fixed gradient surface boundary condition will ensure that irrespective of the input exhumation rate, the temperature gradient will remain fixed and honor, for example, borehole temperature measurements / surface heat flow observations. While this is a challenge when working with regional thermochron datasets, it is also an observation that can be used to define the range of allowable solutions that fit the observed age dataset. Because thermochronometer ages are non-unique, it can be otherwise difficult to exclude models that produce the correct ages, but violate other observations. Thus, it would be good to include some text in the discussion that notes some of the limitations of the choice of design in age2exhume for completeness.

We agree that model design choices are (and should be!) based on what is most valid in the particular physical system being studied. A thermal half-space model makes sense when studying wholesale lithospheric thinning, for instance (although note that currently most authors would probably prefer thermal boundary-layer models, see for instance the recent review by Richards et al., PEPI 2020). However, one should distinguish between using a thermal half-space model versus a model with a fixed temperature at the base, and an input stable geothermal gradient versus an input present-day surface geotherm/heat-flow value. These two choices are independent of each other. We agree that, where such information is available, using the present-day heat flow or surface geothermal gradient will provide additional constraints on possible model outcomes, but reiterate that such constraints are exceedingly rare in mountain belt environments (zero heat flow measurements available in the entire Himalaya for instance). *We have modified the text to make these issues a bit clearer*.

2. Another consideration is the use of Dodson's method to estimate effective closure temperatures. While it is simple and efficient for such calculations, there are also behaviors that are observed in some thermochronometer systems that Dodson's approach does not capture, possibly resulting in difficulties interpreting some age data. For example, it is expected that the effective closure temperature for zircon (U-Th)/He ages will decrease with increasing cooling rate for low-eU zircons (e.g., Guenthner et al., 2013; Whipp, Kellett et al., 2022). This limitation is not discussed, and similar to the comment above, would be a good thing to mention in the discussion section somewhere.

Again, we agree with the reviewer on this point; Dodson's method is a simplification that is strictly valid only in a context of continuous monotonic cooling where temperature evolves inversely with time. This clearly limits applications of this approach to settings where rocks currently at the surface can be reasonably assumed to have cooled monotonously and continuously. The reviewer is entirely correct that more complex settings will not be predicted correctly by the model, also due to the steady-state thermal assumption. *We have included a short discussion of this limitation in the manuscript*.

3. Based on the two comments above, I would suggest adding a section to the Discussion and Conclusions, perhaps titled "Caveats (or assumptions or limitations) and recommended use cases". Here the authors could combine the points related to the limitations of the models to a single section, and also make suggestions for the best applications of the models by future studies. Some of the text from the current section 4.2 could be moved here (lines 317-330), as well as some additions to the text based on the comments above. Perhaps the authors could also comment on the lack of heat production in the geotherm calculation and how that might affect the estimates of exhumation rates. This is not strictly necessary, but may make it easier for users looking for a model to apply to their data to know what the limitations are and be able to easily evaluate whether age2exhume is suitable for their needs. I suspect age2exhume will be a valuable modelling contribution, and more users may choose to utilize the code if they clearly see it would be a good fit for interpreting their data.

We appreciate this suggestion, and have followed it in revising our manuscript.

4. Finally, I appreciate the authors making the software source code available but also feel it is worth noting that the choice to provide the software as a Matlab script is somewhat unfortunate. Matlab is a commercial software product, and if the software was available in another language (such as Python or Julia), then it would accessible to a larger user base at no cost. Similarly, it would be nice to have instructions for calculating delta h using freely available GIS software such as QGIS. I am not requesting the authors rewrite their code, but wanted to emphasize there is a limitation in providing Matlab code.

We have now written a version of age2exhume (the version that takes an input file as an argument) in Python, and will include that, together with an appropriately formatted .csv *input file, in a new Zenodo repository referred to in the revised manuscript. We also added a paragraph to the Appendix explaining how users can calculate ∆h values and automatically extract them at the location of all sample points in QGIS.*

Specific remarks (L = line number)

L15: I know it may not be easy, but is there some kind of clarification that could be used to distinguish a "limited area" from a regional dataset? Is there a certain spatial scale or dataset size that could be used to emphasize when your model should be considered?

A rule-of-thumb guideline for a maximum area that could reasonably be modeled using Pecube with normal modern computing power would be in the order of 100x100 km. *We have added this quantification to the text, but rather in the more detailed description of Pecube (lines 115-120 of the original manuscript) than in the abstract.*

L16: I suppose "largely underutilized" may be somewhat subjective, but perhaps the point here is that it can be challenging to interpret these larger datasets. People have done it (e.g., Herman et al., 2013; Thiede and Ehlers, 2013), but there are not so many tools that are easily applied for such a task (as you say in the introduction). Would it be better to emphasize the challenge and lack of tools here too?

We have modified the writing here to focus on the challenge of using such large-scale datasets in thermochronology.

L85: While I can see the authors' point here, particularly for large age datasets, it is also true that failing to consider the constraint the surface heat flow could provide in terms of allowable models that fit the data is also problematic. Heat flow is an observation that can be used to constrain the model parameter space, and it may be practically easier to exclude this due to the paucity of measurements, but that is not necessarily a better solution.

A limitation of our model is that it only calculates steady-state geotherms since the time of closure of the associated thermochronometer, meaning that any changes to the thermal properties of the crust since closure are not considered. Nevertheless, *we have modified the default output of the main version of the code (the one that takes an input file as an argument) to include the surface geotherm*. With that, a user could at least easily compare the calculated steady-state geotherm with an observed modern geotherm.

L89-90: I have not used the code cited here, but based on the plots in the manuscript I cannot see an obvious reason why more rapid exhumation rates would be problematic for code stability. Could you please clarify what the issue is here and possibly why it arises?

That text was not accurately written; *we have clarified now that the age2edot code does not return a solution if the input modern geotherm is far from what would be expected for a high exhumation rate*.

L101-102: This is again an issue about choice of model design and boundary conditions. While the point raised here about a half-space thermal model perhaps not reproducing constant lag times for regions with a constant rate of exhumation is valid, it is also true that it may be difficult to define a reference temperature in the crust or lithosphere that is fixed over the timescale of orogenic growth and exhumation. For instance, Moho temperatures can change significantly during orogenesis, which is an effect that requires a user to choose an initial geothermal gradient in order to produce a Moho temperature they feel is applicable to the study region. This alone may result in the need to have different initial geothermal gradients for different thermochronometer systems and effective closure temperatures.

We agree with the reviewer and have added a couple sentences to our new section on "Limitations and caveats" that explains this issue.

L106: This is a nit-picky thing, but HeFTy is not exclusively an inverse model, but one that can be run as a forward or inverse model.

Correct; *we have adapted the phrasing here*.

L116: "neighbourhood algorithm (Sambridge, 1999)" rather than "natural-neighbor algorithm".

We corrected this in the text.

L145: Perhaps it could be clarified that the closure temperature estimate is a value the user should provide, not a calculated value. The text says this, but it may help to note the user should select this as an input value (right?).

The initial estimate of the closure temperature actually has little to no impact on the final solution, as the model rapidly converges to a closure temperature that is consistent with the input age and thermal properties. So, although it can be a user input, we have included default values that we believe suffice. But *we have modified Figure 1 to show input parameters in the workflow, including an initial estimate of closure temperature*.

L148: Would it be better to say that "an estimate of Ts is calculated..." rather than "Ts is estimated..."?

We have modified this to "Ts is calculated …".

L168: Is there any justification for the threshold value of the exhumation rate change here?

We experimented with the code to see how many iterations are required for convergence. We found that it is quite fast (most results converge in 7 iterations or less). This enabled us to maintain rapid calculation times even with a rather small threshold for exhumation-rate change. That small value in turn is helpful in ensuring that results are reproducible.

L174: Somewhere in the text above it may be helpful to clarify precisely how the initial geothermal gradient is used, to avoid any possible confusion about it being treated as an initial condition. As it is to me as a reader familiar with modelling, this is clear to me, but it is possible some less familiar readers could be confused. Thus, it may help to state it is used only to estimate zc at the start of the simulation and to define the basal boundary condition. Just a suggestion, but one that might help readers better understand how the model works.

We had hoped the text and Figure 1 were sufficiently clear on this point but *we have added a phrase to avoid any potential confusion on this point*.

L195-197: This seems like a very useful point and possible way in which new users would really apply the code. Would it be possible to somehow emphasize this point more in the text?

We have done this by adding a phrase to the abstract and the concluding remarks.

L216-217: Again, this is somewhat nit-picky, but the model does not make assumptions, the programmer does. In this case, I would suggest rephrasing to say "...time in the model." rather than "...time made by the model".

Agreed, *we made this change*.

L243-244: Is the reference here citable?

Not yet, the manuscript is now in review. However, a pre-print is now available, and we will check the journal guidelines regarding how to appropriately cite it (if we can).

L283-284: This effect is something that has been shown previously in, for example, Mancktelow and Grasemann (1997) and Stüwe et al. (1994). Perhaps it would be note the consistency with earlier work?

While both Stüwe et al. (1994) and Mancktelow and Grasemann (1997) provided important new insights in how to interpret thermochronology data, neither of these included a direct comparison with real-world data. The point we are trying to make here is that the thermal effect of exhumation will affect thermochronometer ages to a degree that exceeds the uncertainty on these ages (i.e., “significantly”).

L303-304: I would guess that some readers might not immediately understand the point here about the Peclet number and how changes in L affect the exhumation rate estimates. Could you add a sentence to two to clarify why this happens?

We have rewritten this sentence to more explicitly state how L affects the modelled thermal structure and thereby the inferred exhumation rates.

Figures

Figures 2 and 3: Would it be possible to produce versions of these figures with a colored contour fill and black contour lines? It may make them easier to read, as the yellow color on a white background is somewhat hard to see. Also, the contour lines in Figure 2 get quite dense along the left side of each panel. Would it be possible to remove some to make it easier to see the remaining lines? Finally, it may be helpful to have a reference gridline for delta h = 0 on the plots.

We appreciate these suggestions to make these two figures more easily readable, and *are experimenting with implementing them*.

Figure 4: Would it be possible to include an orogen-scale inset map somewhere in this figure showing the extents of the panels?

We have added an inset map that shows where the zoomed panels plot over the scale of the whole orogen. We have made some other slight modifications to the figure to improve readability.