Review 1

The authors analyse a commonly used finite-difference method for estimating thermal diffusivity of supraglacial debris using vertical debris-temperature profile. The aim of the study is to understand the effects of temporal and spatial discretisation on the uncertainty in estimated diffusivity. Due to the importance of debris thermal properties in computing the dynamics of debris-covered glaciers, this topic is of importance. However, the study appears to have serious weaknesses in terms of the experimental design and the underlying assumptions. unless these issues are addressed with suitable (and doable) modification of the methods, the results and conclusions will remain weak.

We thank Dr Banerjee for his thoughtful and critical review of our work. In the light of this and the other reviewers comments we have substantially revised the manuscript. In particular we have:

- restructured and rewritten it to make our purpose and scope more clear.
- emphasised the utility of analysing this method, despite the developments of new approaches in e.g. Laha et al, 2022.
- clarified out experimental strategy
- checked and amended where needed the equations highlighted as in error

We have however not substantially modified our methods, but we believe with the revision we have justified our purpose and value of the analysis performed and the tool presented for other researchers to explore their own data.

We outline our responses and revisions that are included in a revised manuscript below.

Major comments

(1) I see a few possible careless mathematical errors and inappropriate assumptions, which could have been avoided. For example,

1. The expansion given in equation 20 is wrong. $(1+x)^{-2}=1-2x+3x^2-4x^3+5x^4+\ldots$..
2. Please check Eq 9. For example, according to https://en.wikipedia.org/wiki/Propagation_of_uncertainty, if $\sigma_f=a/b$,
   then $$\sigma_f \approx \sqrt{\sigma_a^2\left(\frac{1}{b}\right)^2+\sigma_b^2\left(\frac{-a}{b^2}\right)^2}.$$
3. In section 3.5, equation 19, you assume that the temperature values correspond to equispaced sensors and thus errors in the numerator vanishes. However, the spacings are not equal anymore due to the

random shifts (both in a real experiment and your simulations). The errors in the numerator must be considered.

We apologize for the equation errors, which are in the manuscript rather than the calculations and so do not affect the general results of the paper.

Section 3.5 has been rewritten and partially removed, which involved also removing equations 19 and 20. The published python tool uses the equations for unequal spacing of thermistors, so the equation shown is not relevant and has been revised. Furthermore, in revising the manuscript we noticed that Eq. 9 is not used further in the manuscript, and it was also removed. We have checked all other equations in the manuscript and cross-checked those behind the interactive tool and figures produced to ensure no remaining errors.

(2) A couple of recent publications (Laha et al. & Petersen et al.) went beyond the assumptions of a homogenous, source-free, purely conductive heat flux, as being considered here. You need to provide more compelling an argument about the motivation behind and significance of the present study than you do in L107. (Of course it is a different matter, if you were to actually analyse the existing data of thermal diffusivities reported in the literature, but that is not something you attempt here.)

We agree that the forward model and inclusion of a 2 layer consideration presented in Laha and others (2022) is demonstrated to outperform the Conway and Rasmussen (2000) method for cases where one is trying to calculate apparent thermal diffusivity from data in vertically varying debris and with irregularly spaced thermistors, however we maintain their is value in our analysis as:

- Much of the historical data upon which literature sourced values of debris thermal conductivity use the Conway and Rasmussen (2000) method, and understanding the limitations on the inter comparability of this data is important.
- Implemented correctly in the field the Conway and Rasmussen (2000) method still holds the potential to determine relevant thermal conductivity values representative for the bulk debris properties, as required for surface energy balance modelling. This is evidenced in Laha and others (2022) by the identical performance of this model to the newer methods for equal spaced thermistors and homogenous debris - with the implication that the method remains applicable to debris temperature profiles that confirm to these conditions, which is indeed how early studies applied it.
- We are interested in capturing the thermal conductivity throughout the debris layer to generate input parameterization for melt models and/or to identify stratigraphy of properties and/or non-conductive processes in the debris layer. While the method of Laha does account for two layers, it does not yield thermal conductivity values for each layer. This would be needed for application to a multilayer surface energy balance model where fluxes are driven from the surface rather than an internal temperature measurement.

To address this in the text, we introduce a section stating our purpose:

"*As the existing, and limited, sets of field data used to provide generalized values for the effective thermal conductivity of unmeasured glacier sites have been analysed based upon the simple Conway and Rasmussen method, rather than the later methods developed by Laha and others, there is value in a deeper exploration of the limitations to interpreting values derived by this method, in particular to better understand the meaning of comparison between sites at regional and global scales (e.g. Rowan et al., 2021; Miles et al., 2022). For example, the measurement parameters for temporal or spatial sampling intervals, thermistor spacings, and debris depths used in the application of the standard method presented by Conway and Rasmussen (2000) are selected ad hoc and differ from measurement site to measurement site (e.g. Juen et al., 2013; Chand and Kayastha, 2018; Rowan et al., 2021), and uncertainty estimates associated with this are missing. This means that baseline literature values that are subject to onward use on the literature may be differently influenced by sensor, installation and numerical truncation errors. This study explores the effect of the chosen temporal and vertical spatial temperature sampling interval and other systematic measurement errors originating in the measurement setup on the derived thermal diffusivity values. To explore the capabilities and limitations of this approach we apply this method to artificially generated data with a known value of thermal diffusivity, which allows us to individually quantify systematic and statistical errors by error source. We additionally present an online tool to allow interactive analysis of these combined errors for a given dataset and a best practice guideline on how to minimize the systematic errors inherent in the methods of Conway and Rasmussen (2000).*"

(3) While the space- and time-discretisation steps are important for setting up actual measurements, it is difficult to judge whether your analysis can actually lead to useful insights about real experiments, due to the idealisations involved in your study design. You have totally ignored the sources that are present due to the horizontal inhomogeneities and temperature gradients, water content, advected heat, latent heat transport, convection etc, and vertical variation in Kappa. If one were to incorporate the noise due to these effects in the forward model, which are present in the real system anyway, will your results hold?

We agree that in order to apply this method correctly in the field a site must be chosen that is minimally affected by these processes, and even then the full temperature profile should be investigated to understand the potential variation of debris properties and non-conductive processes being applied.

Nevertheless, where one can identify 'well behaved' sections of the debris temperature profile that conform to the underlying linearity of the heat equation for conductive systems, the approach remains valid. This is inline with the traditional application of the method which emphasised finding optimal conditions for thermal diffusivity determination in order to extract reasonable values of thermal properties for subsequent use in generalised energy balance modelling.

Once an optimal site for analysis is found the systematic error sources revealed by our study still apply and these should be additional considered in the ideal installation of an array of thermistors.

(4) Since you are solving the forward problem numerically, it is easy to create an ensemble of experiments where all the parameters and variables in your model are perturbed by appropriate space and time dependent noise, and the corresponding mean values are drawn from a distribution. If one does that, then all the lines in your plot will have an associated uncertainty band. When such uncertainty is considered, the differences between different curves that you have discussed throughout your manuscript may become insignificant. Till you do this exercise and demonstrate that your results/conclusions are robust against such the inherent variability and measurement noise, your conclusions are not on a firm footing.

Our goal is to show the fundamental behaviour and consequences of the choices of how to instrument and analyse thermistor data that are embedded within all the real world complexity - which we agree can be large. In fact the analysis you describe was already done in Laha and others (2022) for hypothetical data and, as the equations for the CRh model used in that study are the same as the ones we use, we see less community value in repeating that, and prefer to to focus on theoretical cases to understand the underlying behaviour. Nevertheless to address this comment we now add text to clarify the magnitude of the error sources we tackle in our analysis in comparison to the variability seen in published field observations to set our findings in context.

Other comments

I think there is scope and need for improving the writing. There are several sentences which either lacks justification, or are vague, or even incorrect. Also, it may be better to avoid subjective discussion when you set up/introduce the equations and symbols, eg, at the beginning of your methods section. Please just state the standard definitions, and provide units. Please revise/reconsider the sentences/phrases listed below.

L18: The regional-scale debris-covered effect was discussed in several papers eg, Scherler et al, Nature Geosci., 2011, Banerjee & Shankar, 2013. Even though Hock et al. 2019 ignored such effect, it was not previously thought to be unimportant - only there was ready-made way to incorporate it in large-scale models.

L33: Not sure how attenuation of daily signal controls heat flow. Of course, it is a consequence of the diffusive evolution of T(x,t).

L34: what is “thermal instability” in a conductive system?

L47: What processes? "The supply of melt energy” can never be “represented” by an effective thermal conductivity.

L57: Despite the long, general introduction, the question addressed in this paper is not motivated at all.

L87: Does this paragraph belong to methods? Seems to be more suitable for the introduction section.

L135: please demonstrate that it is enough to consider the intercept, to incorporate all the errors/uncertainties mentioned in major comment 3 and 4, for example.

L197: Limiting values are still well defined.

L206: Please provide mathematical justification or some relevant reference where such justification has been provided, for this average method (method 2). What you are calling skipping, is just a standard finite-difference method as explained in any textbook.

We have substantially reorganised, restructured and re-written the text so in that process all the problematic points below have been removed or corrected. In particular, we have clearly stated our goals, reduced and reorganised the introduction, ensured that methods are contained in the appropriate section, and added a conclusion that covers aspects of the general application of this method in both numerical and real-world terms. We are confident that this takes care of the concerns about the quality and clarity of the writing as well as the highlighted points above, but specifically we note re. L135: we have no systematic error in our artificial data but we also add a comment that for real world data the intercept contains error terms that are not represented in our analysis and re. L206: We would advise temperature sampling in time, but this averaging method was used for all field sampling of thermistor datasets from the Khumbu glacier in Rowan and others (2020) - as such we include it in our analysis to allow people to understand the implications of that. We add an explanation of this reason in the text.