Response to second reviews

We thank both reviewers for their time in considering our paper again, and are happy to hear that the paper has improved. As the reviews essentially have the same concerns we treat them together here, leading with the comments of Reviewer 1.

Reviewer 1 asked for us to add random temporally constant but space variant noise in the forward model to set our relative error analysis in the context of likely uncertainty associated with real world debris cover, specifically referring to the ways that natural debris might not meet the assumptions of the method we are testing. In the end, we still prefer to not do this beyond what we have already done in the paper, for the reasons explained below. We have however included further analysis exploring the role of noise within the method to highlight the limits at which the method breaks down and is no longer applicable, which has resulted in 2 new figures (Figures 10 and 11).

Firstly, the thermistor location accuracy experiment that we perform already includes the analysis requested - a temporally constant but space variant locational offset sampled randomly from a Gaussian distribution. This case represents both 'poorly known location' and also non-conformity to the assumptions of the method, because in effect this resampling applies the method to a temperature profile that is no longer that of a purely conductive system. What we see is that, as the Gaussian error is symmetrical, the mean of the ensemble returns the target Kappa value, aside for impractically close thermistor spacing (see Figure 9 in our paper). The issue with adding Gaussian noise in general to the forward model to represent all the potential error of non-conductive processes, is that the nature of a distribution that represents the published real world debris cases is impossible to determine meaningfully, as it could be very small for cases that closely approximate the assumptions of the method, or very (indefinitely) large if the sampling site is poorly chosen and suffers from inhomogeneity and significant non-conductive processes. So we would be adding a potentially unlimited uncertainty to our analysis. Instead, by isolating only methodological errors and error behaviour we can make specific recommendations for deployment and analysis strategies. If we add noise to the target Kappa then we can no longer determine the source of error produced. Our study follows the structure of synthetic experimentation performed in Laha et al, 2022 to assess the numerical error of uneven thermistor spacing, expanding that analysis to a wider range of numerical errors related to measurement choices. Laha et al., 2022 also did not apply Gaussian noise to their synthetic cases, rather only to the field datasets, which we agree seems appropriate and desirable. By isolating error sources independently our approach allows us to identify that the error response of adding Gaussian noise to the measurement position varies with vertical sampling interval, from that we can see that while the Gaussian noise added to real temperature data from sites 4 and 7 on Santopath glacier in Laha et al (2022) show similar error response, this would be expected as both of the sites analysed use a sampling interval of 6cm, and this error response might however, not be representative for sites on the same glacier with different vertical sampling interval.

Secondly, as the point of our study is to identify meaningful best practice guidelines for measurement set up, we must focus on the methodological errors inherent in the Conway and Rasmussen method. For this it is only fair to test the method in cases for which it is valid - i.e. in which the debris is homogeneous and closely approximates a purely conductive system. Laha et al 2022 already nicely demonstrate the weaknesses of applying this method when these assumptions are not met, and provide an alternative method that outperforms the one we are testing in those circumstances, and other studies highlight how to identify when these conditions are not met in the real world datasets (Nicholson and Benn, 2013, Petersen et al, 2022). We have included an assessment of statistical noise to showcase the behaviour of the method when applied to spatial-temporal gradients that can no longer be meaningfully assessed by the method to yield Kappa values. As the vast majority of published data points are based on the original Conway and Rasmussen method, we still feel it is useful to present its associated numerical errors alongside a tool aimed to allow colleagues to easily determine an optimal strategy for future field deployments.

Reviewer 2 suggested that we should follow the uncertainty assessment of Laha et al, 2022. However we note that these authors only applied the type of Gaussian noise uncertainty to their observed field datasets, which we agree is entirely appropriate and desirable. As we have done in our study, for their analysis of *synthetic cases* to assess the error associated with uneven thermistor spacing and choice of method , Laha et al., 2022, also did not apply noise, presumably because they also wanted to preserve the known target Kappa value in order to be able to assess methodological performance at recovering it.

To address the reviewer points within the manuscript we now:

- Emphasise that care should be taken to apply the analysis only to subsets of the data which can be shown to be predominantly conductive. These conditions have been established through previous work to be best met in a matrix-supported diamict, during stable meteorological conditions within the established ablation season, and that data from these conditions will yield the most reliable result of apparent thermal diffusivity for use in models.
- Emphasise throughout the paper why we wish to test only the inherent methodological errors of this method that has produced the vast majority of available published data points for debris thermal properties.
- Include an analysis of Gaussian statistical noise on the temperature timeseries to demonstrate the limits of the method in cases where the spatio-temporal gradients are degraded to the point where they are no longer interpretable by the method being tested. These results are shown in new Figures 10 and 11.

- Removed the background shading for the figures that had been used to distinguish between skipping and averaging temporal resampling, as requested by Reviewer 2.

Additionally we also:

- Stress that the examples we show are illustrative of the relevant parameter space that can be explored in its entirety with the interactive tool provided - as a means to more strongly promote the use of the provided tool to determine optimal field sampling strategies prior to sensor deployment.
- Relatedly we have added more information on the details of each case shown, as in places this was incomplete in the former manuscript.
- Include reference to the recently published DebDab dataset (Fontrodona-Bach et al., 2025), which includes the most complete database of published and non-published debris thermal conductivity values.

These changes entailed the inclusion of some additional text compared to the previous version, which also included some editorial