

Dear Editor,

Thank you for your invitation to resubmit following minor edits. We have included below the comments from the first (RC1) and second (RC2) anonymous reviewers alongside our responses in **bold text**.

Please note that all line numbers refer to the original manuscript. Changes to text can be seen in the included, marked up pdf version (red strikethrough = deleted, blue underlined = added). Please also note however that changes to figures are not explicitly highlighted in the markup.

Kind regards,

Nick Brown and Stephan Gruber

RC1 Minor comments:

RC1: In the abstract (lines 11–14), the text reads as if it presents a finding related to MAGT, but it does not discuss how other metrics may be more suitable, particularly with respect to uncertainty versus informative value. Later in the paper (L.389), there is some discussion of the value of the “annual thermal integral” in this context. Consider revising the abstract to more clearly highlight what the new metrics can contribute.

The abstract has been revised to more clearly state the benefits of the new metrics.

RC1: In the abstract and the introduction, consider clarifying that MAGT is generally inferred at a single depth and is sometimes compared across sites at different depths. I found this background somewhat unclear.

In the abstract, the following text has been added:

"Although MAGT is calculated for a single depth, associated warming rates are often compared or aggregated between sites with differing measurement depths. We observed depth-related differences in MAGT warming rates ..."

In the introduction, the following text lines have been changed to more clearly identify this point:

"(3) Differences in observed depth can further affect the relative timing and magnitude of the MAGT trends from different locations. There is no standardized measurement depth near d_{za} in the current practice of reporting MAGT and the potential magnitude of depth-related effects is unknown."

The new text reads as follows:

"(3) MAGT is inferred using observations from a single sensor. In current reporting of MAGT there is no standardized measurement depth (other than generally being at or near d_{za}) and the metric is sometimes compared across sites at different depths. Differences in the choice of observed depth may further affect the relative timing and magnitude of the MAGT trends from different locations, however the potential magnitude of depth-related effects has not been reported."

RC1: L.550: Few details are provided on the model. Consider adding a few sentences describing the processes represented in the active layer and talik, if present (for example, advection). Please also explain how freeze thaw energy is represented in the model. While some details are provided in the Supplementary Material, the main text lacks basic information on the model type.

We have expanded the model information in section 3.3 when the model is first introduced to describe relevant processes and freeze thaw energy:

"To simulate transient ground temperatures in a one-dimensional configuration, we use the model FreeThawXice1D—a numerical model of heat conduction with freezing and thawing in soils without water flow (Tubini and Gruber, 2025, Tubini et al. 2021). Energy conservation and model convergence is guaranteed at large time steps, making it suitable for ice-rich simulations requiring long spin up. Liquid water content, and thus freeze-thaw energy, is represented in the model using temperature-dependent soil freezing characteristic curves (SFCC). For our application we used the van Genuchten SFCC parametrization presented in (Dall'Amico, 2011). FreeThawXice1D represents the effects of subsidence caused by ground-ice loss and accurately tracks 0 °C isotherms via local mesh refinement. Although advective water transport is not included in the governing equations, liquid water volume caused by

excess ice melt and the associated latent heat is removed directly from the simulation.”

RC1: The use of a model may influence the results if the model has limitations in representing certain ongoing hydro thermal processes. Consider clarifying potential weaknesses and their implications (for testing the value of the metrics).

"This is an important consideration! The key issue isn't whether the model's limitations affect both the metrics and heat content, since if they're affected equally, the relationship between them should remain consistent, but rather whether the lack of advective water transport disproportionately impacts one or the other. If it does, the inferred relationships could be biased. Such is the case when we also consider the limitation of a 1D soil column.

We've added the following discussion to address this limitation:

Our use of FreeThaw1D means we do not consider spatial effects or the soil water balance. Taken together, these limitations could affect our results and the applicability of the metrics.

Advective water transport could lead to local redistribution of heat that does not correspond to an overall net change in the larger area. This could lead to an overestimation of change by the metrics because they only measure local changes.

In our simulations, there is also no water flow within or out of the simulation with the exception of melted excess ice. temporal variability in near-surface moisture can influence d_{za} by altering the soil's apparent thermal diffusivity independent of long-term change. The model's fixed moisture conditions may smooth out these variations, leading to more stable d_{za} trends than would occur in reality. This could make d_{za} appear less sensitive to changes in H_i than it truly is.

Despite these potential biases, our focus on the relative performance across metrics helps control for systematic errors. For example, if all metrics are similarly affected by missing processes advection, their relative performances remain valid.

RC2 Major points to address

RC2: Some of the proposed metrics are not completely clear to me. I have noted specific questions below.

Specific questions are addressed below

RC2: Several sections are overly lengthy and would benefit from more concise writing, particularly the Introduction. Throughout the text, there are some redundancies which should be addressed

In addition to specific recommendations below, the introduction has been shortened by removing redundant material and tightening sentence structure.

RC2: The manuscript relies mainly on simulated time series (ERA) forcing a heat flow model (GlobSim) to derive and present the new metrics and then includes a case study that does not add much new information. Consider removing Section 5 or integrating it into the Discussion.

We believe that including real data as a part of a case study helps ground the research in reality and acts as a ‘ground truth’ of the simulation-driven theoretical development of the metrics. It also helps illustrate some of the differences in trend rates caused by measurement depth.

We have shortened section 5 to tighten the focus and have added a map of locations in the supplementary material as Figure S37.

RC2 Detailed comments

RC2: p. 5, chapt. 2.4: I do not fully understand the discussion about the talik. Please rephrase for clarity.

This paragraph has been rephrased as follows:

In cold permafrost, our estimate of the ALT coincides with the depth to the permafrost table, or top of permafrost (TOP). However, if a supra-permafrost talik develops, this is no longer true. In that case, TOP continues to deepen independently of ALT causing the two metrics to differ by an amount equal to the talik thickness. Here, we neglect any

differences between the thermal (cryotic) and physical (freeze-thaw) definitions of talik or active layer boundaries.

RC2: p. 6: There are many equations — check whether all are necessary. Some measures are commonly known and do not need to be written out.

Four of the more obvious equations have been removed from sections 2.3–2.8.

RC2: p. 7, chapt. 2.8: I do not follow this paragraph; please clarify or rewrite.

Section 2.8 has been rewritten for clarity.

RC2: p. 10, last paragraph: Delete this paragraph, it is not necessary.

Paragraph has been deleted

RC2: p. 11, l. 275: Delete this paragraph. You do not need to continually preview what comes next.

Paragraph has been deleted

RC2: Figure 1: This is a nice figure. Consider indicating that the first six boxes represent “cold” sites and the next boxes “warm” sites. Alternatively, place cold sites in the left column and warm sites in the right column to facilitate comparison.

The image has been reordered so that the left column represents cold sites, and right column represents warm sites

RC2: p. 13: You have a subsection 4.3.1 but no 4.3.2. Either remove the subheading for 4.3.1 or add the missing subsection.

Subheading has been changed to 4.4

RC2: p. 4.5: This paragraph contains some redundancy — please tighten it.

Assuming this means subsection 4.5, the text has been reworked to reduce redundancy

RC2: Figure 3: I do not understand this figure; please improve the caption and figure clarity or simplify the presentation.

Figure has been simplified by removing coloured CDFs corresponding to different terrain types (these have been moved to supplementary materials). The retained CDF has added reference lines for 50th, 90th and 95th percentiles discussed in text. The caption has been reworded

RC2: p. 4.9: I do not follow this section completely. Why is this point important? Also, Fig. 4 is not readable — improve resolution and labeling.

This section aims to provides quantification for the reliability of metrics based on their signal-to-noise ratio (when used as indicators of changes to borehole heat content) and consistency. More specifically, we want to evaluate how much to *trust* that the direction and existence of change in borehole heat is predicted by change in a metric.

We recognize that calling this concept ‘predictiveness’ may contribute to some confusion and have replaced this with ‘reliability’ and SNR as defined above.

The introduction of this subsection has been significantly rewritten to clarify this point, and the remainder of the text has been updated for consistency.

The resolution of figure 4 and its labels has been increased to 600 DPI. Axis labels have been added to all bottom axes. Axis ticks and labels have been added to all left axes. The leftmost figure panel has been removed for clarity.

RC2: p. 4.10: Consider deleting this summary section; it seems unnecessary.

While this section does repeat some material from earlier parts of section 4, we argue that it serves a purpose by integrating the qualitative (sec 4.1-4.7) and quantitative (sec 4.8-4.9) analyses of the metrics to provide a final interpretation before moving on to the case study and discussion.

RC2: Chapter 5: See comment above about possibly removing this section. If you keep it, provide a map showing site locations — many readers will not know where “Schilthorn” and the other sites are located.

A map (Figure 5 in revised manuscript) has been added to show site locations and some text has been removed to shorten and clarify the section.

RC2: l. 571–572: The purpose of these lines is unclear, especially the sentence in line 572. Please reword or delete.

The lines have been deleted

Our results highlight the challenges of comparing trends between boreholes or regions. In our simulations, the behaviour of each metric generally follows a similar trajectory. For MAGT, this is: rapid warming, reduced warming, no warming, and finally rapid warming if thaw progresses to the observation depth. Most importantly, we do not change the magnitude of the warming...

RC2: Chapt. 6.3: The comment about bedrock sites echoes the original objective of the PACE boreholes, if I recall correctly — consider mentioning that here.

A sentence has been added pointing towards PACE as an exemplar of bedrock sites:

“The PACE project provides another example where boreholes in bedrock on mountain summits or plateaus were used for permafrost and climatic monitoring (Etzelmüller et al., 2020; Harris et al., 2009).”

RC2: Chapt. 6.5: The discussion seems to indicate talik development. If that is the intended message, state it clearly; if not, consider removing this paragraph.

Section heading changed to “Talik formation and accelerated permafrost degradation” removed. Two paragraphs of the section have been split out into a separate ‘implications for monitoring’ section for clarity.

RC2: Conclusions: Provide clear, concise conclusions, preferably as bullet points, rather than a long paragraph with references and summaries.

References and summaries have been removed. The conclusions have been reworked into bullet points, and some material has been moved into the discussion.