Dear Anonymous Referee 1,

Thank you very much for your review and your constructive and detailed comments. The entire text of your comments (RC1) is shown below, followed by our responses (bold text). We hope that you will find the following explanations and proposed improvement to our manuscript a suitable response to the issues you have raised.

Throughout our responses, we refer to line numbers in the original manuscript and state revisions we have made to the manuscript. If we are invited by the editor to submit a revision, it will contain these changes.

Kind regards,

Nick Brown and Stephan Gruber

This study proposes and seeks to demonstrate the use of novel metrics derived from temperature observations to complement commonly used metrics such as mean annual ground temperature (MAGT) and active layer thickness (ALT), with the goal of improving the quantification of permafrost change, especially by better accounting for the effects of latent heat. The authors investigate the behaviour of existing and novel metrics using an ensemble of more than seventy 120 year simulations. They also apply these metrics to observations and discuss issues associated with sensor deployment and resolution. I think the topic is of interest and value to the community, and I strongly agree with the authors that new metrics need to be explored and proposed. I believe the real value and usefulness of new or improved metrics can only be assessed through their adoption, or lack thereof, in future studies. The work presented by the authors is detailed and appears correct with regard to the calculations and testing performed. I therefore have only minor comments.

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_l 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.