Reviewer comments are in *black italics,* our responses are in blue font.

Reviewer 2

Summary

This study explores how both seasonal and annual mean temperature variations influence permafrost dynamics during past interglacial periods, specifically the Mid Holocene (MH) and Last Interglacial (LIG). The authors employ simulations from the AWI-ESM-2.5 Earth System Model in conjunction with the CryoGridLite permafrost model to analyze ground thermal characteristics such as permafrost extent, thaw depth, and thermal contraction cracking. The key finding is that permafrost extent is primarily controlled by annual mean temperatures, whereas thaw depth and thermal contraction cracking are predominantly influenced by the amplitude of seasonal temperature variations. Importantly, the study concludes that past interglacial climates are not suitable analogues for projecting future permafrost thaw, as modern climate change trends involve rising mean temperatures coupled with decreasing seasonality, a pattern not observed in the past.

We thank the reviewer for the overall positive evaluation of our manuscript and the various comments and suggestions which will improve the revised article. Below, we provide point-by-point replies to the comments.

Major Comments:

1. Missing Literature on Borehole Temperature Reconstructions: *The manuscript omits an important body of work related to reconstructing past climate from borehole temperature data. Borehole climatology offers critical insights into ground surface temperature (GST) changes over the past centuries and millennia. For instance:*

- *Liu et al. (2021) applied Tikhonov regularization to infer ~1.8°C of ground warming over the past ~308 years, aligning with instrumental records.*
- *Mareschal & Beltrami (1992) and Beltrami & Mareschal (1991) pioneered SVD-based inversions of borehole data, identifying historical warming in regions like eastern Canada.*
- *Shen & Beck (1991) used least-squares inversion with regularization for stable GST reconstructions.*
- *Pollack & Huang (2000) presented a global borehole climatology, documenting multi-century surface warming. These studies offer a complementary line of evidence that should be discussed to provide a more complete context for past permafrost dynamics.*

We thank the reviewer for pointing us to this body of work. Indeed, there is a considerable body of literature on how past climate, in particular ground surface

temperatures, can be reconstructed on the basis of (permafrost) borehole temperatures, and part of our team has been involved in such efforts (Groenke et al., 2024). These approaches allow for partly assessing permafrost characteristics during the past decades to centuries. However, we do not see that this body of work is of immediate relevance to our present study, in which we want to assess and compare permafrost characteristics between different interglacial climates, i.e. on orbital time scales of multiple millennia. While the borehole temperature reconstructions allow insights into the recent transient evolution of the climate over the past decades to centuries, we are primarily interested in comparing permafrost characteristics between different quasi-equilibrium climate states. Thus, the borehole temperature reconstructions are aiming at a much higher level of detail and shorter time scales, than our approach based on climate modelling (combined with speleothem proxies) is targeting. We therefore appreciate the suggestion of the reviewer, but would prefer to deliberately omit discussing this body of work, as it is only very remotely related and not essential for interpreting our results.

2. Clarification of Paleo Proxy Records: *A paragraph introducing and defining speleothem and pollen records is needed for readers unfamiliar with these proxies. While valuable, these records are spatially limited and prone to interpretive uncertainty, which affects their robustness in validating model outputs.*

We agree with the reviewer that these proxies should be introduced in more detail and we will add respective introductory paragraphs in the revised article.

3. Model Bias and Alternatives: *The AWI-ESM-2.5 model has a documented cold bias in high-latitude regions, which likely leads to an overestimation of permafrost extent. The authors should discuss whether other Earth System Models (e.g., CESM, MPI-ESM) have been considered for comparison, especially those with better temperature performance at high latitudes.*

We agree with the reviewer that the AWI-ESM-2 has certain biases in its simulated climate state and that these could affect the simulated permafrost characteristics. Specifically, the model version AWI-ESM-2-1-LR has a documented warm bias for the PI in northern high latitudes (60-90N), and a cold bias in northern mid latitudes (30-60N), and its predecessor AWI-ESM-1-1-LR a warm bias in both of mid and high latitudes (Otto-Bliesner et al., 2021). Similar biases have been documented for other CMIP6-PMIP4 models including MPI-ESM-1-2-LR and CESM2.

In the present study, our objective is not primarily to assess how accurately the AWI-ESM-2.5 can simulate the climate and permafrost characteristics during the PI and past interglacials. Instead, our primary objective is to assess the effects which different mean annual temperatures and seasonal amplitudes during these climate states have on permafrost characteristics. Therefore, additional simulations driven by the climate forcing of other ESMs would likely not lead to qualitatively different results with respect to the main drivers of permafrost dynamics. We therefore consider the

added value of such additional simulations with other ESMs as relatively small compared to the substantial additional work that would be required.

Nevertheless, we take this as a valid comment and have extended the discussion of possible biases in the AWI-ESM forcing and how the AWI-ESM compares to other PMIP4 models.

Minor Comments:

P2, Line 50: *The introduction notes that few models address permafrost conditions in past climates. The conclusion should explicitly state how this study advances that field and relates to previous work.*

We have revised and restructured the Conclusion section and included a statement on how our work advances the field of modelling past permafrost conditions.

Figure 3d: *Why is there no simulated permafrost in the southwestern part of Russia? A similar issue arises in Quebec, Canada—please clarify.*

In all simulations (pre-industrial, mid Holocene, Last Interglacial), no permafrost was simulated for the southwestern part of Russia. For the pre-industrial conditions this is consistent with modern known permafrost extent (e.g. Obu et al. 2019), and for the MH period it is consistent with paleo proxies (as discussed in section 3.1 of our article).

For eastern Canada, our simulations suggest a permafrost presence in the northern parts of Quebec and non-zero permafrost probabilities down to the southern limit of Hudson bay (MH and LIG: Fig. 3 b,e; PI: Fig. 4a). Maps of modern permafrost extent (e.g. Obu et al., 2019) show a similar permafrost distribution in northern Quebec.

However, we see that our model simulations tend to underestimate the presence of isolated patches or sporadic permafrost. This could be related to the fact that such sporadic occurrences of permafrost are often patches of permafrost which formed during colder conditions in the past and is now protected by organic-rich ecosystems like peatlands. It is thus a transient phenomenon which our equilibrium model simulations might not capture well. We discuss this in the last paragraph of section 4.1 which we have extended in the revised article.

Figure 4: *A consistent colorbar across all panels would improve clarity. What does the current colorbar represent?*

We decided to present the results in absolute terms for the original PI simulation (panel a), and as anomalies relative to the PI for the remaining simulations (panels b-i). Therefore, the first panel has a different colorbar (showing the absolute

permafrost probability as a value between 0 and 1), than all other panels (for which the colorbar shows the difference between the simulated permafrost probability and that of the PI simulation, i.e. a value between -1 and +1). The same applies to figures 6 and 7. In the revised manuscript, we extend the Figure captions to clarify this.

Figures: *Expand abbreviations like PI (Pre-Industrial), LIG (Last Interglacial), etc. Also, labels like "ΔTa" need more descriptive titles.*

We have revised the figure captions to more clearly explain the symbols or abbreviations used in the figures. We prefer not to use full-length labels in the figures.

P3, Line 60: *Clarify why focusing on both the current and last interglacial periods is important, particularly in relation to temperature amplitude.*

The PI, MH, and LIG climates are most suited for our analysis due to (i) the availability of proxy-based paleoclimate reconstructions, and (ii) the consistent definition of boundary conditions in the scope of PMIP4. We explain this in the revised article.

Cracking Quantification: *Explain in more detail how thermal contraction cracking is quantified in the model.*

Thermal contraction cracking is quantified based on the thermal conditions near the surface. As explained in section 2.2.3., we define the thermal contraction cracking activity as the number of days in a year, during which certain threshold values for the ground temperature in 1m depth, and the temperature gradient in the upper 1m of the subsurface are exceeded. The respective threshold values for shallow and deep thermal contraction cracking are based on the field observations of Matsuoka et al. (2018). The same approach was taken previously by Stadelmeier et al. (2021). We noticed a mistake in the description of the diagnostic which we fixed in the revised version.

P3, Line 80: *Are glacier masks, topography, and related inputs dynamic in the model, or are they static? How might this affect the results?*

For each time slice (PI, MH, LIG), the input parameters (including glacier mask, topography, etc.) are statically prescribed, following the PMIP4 protocols. We specified this in the revised version.
As each climate simulation is supposed to reflect a dynamic equilibrium climate state (i.e. there is inter-annual variability, but no long-term trend), the respective inputs are held constant. Letting ice sheets, topography etc. evolve dynamically would require

substantially extended spin-up periods, which is usually not done for time slice simulations and not consistent with the PMIP4 protocol.

P5, Line 108: *If dynamic hydrology were included in CryoGridLite, how might this change the study's findings?*

This is an interesting and valid question, which, however, would require further model development and simulations to be answered with confidence. In the present study, we limited ourselves to studying the effects of different surface temperatures and snowfall rates, which are, however, known to exert the dominant controls on the ground thermal regime in permafrost settings (REFS).

If a dynamic hydrology scheme was included, this would also include a surface energy balance calculation to be able to quantify evapotranspiration and the full water balance. By this, the simulations would likely be more sensitive to changes in precipitation between the different interglacial climate states. However, we assume that these would be of secondary importance relative to the changes in surface temperatures and snowfall.

We will include a paragraph discussing the dynamic hydrology and provide an outlook regarding its implementation in future work.

P6, Line 132: *The text mentions subsidence. Was uniform subsidence assumed across the region, or was spatial variability considered?*

We checked carefully, but could not find any mention of subsidence in our manuscript. In fact, we did not take any effects due to melting of excess ice (e.g. subsidence, thermokarst, thermo-erosion) explicitly into account in our modelling. However, we do discuss potential effects of the simulated thermal dynamics on the (excess) ground ice dynamics (cf. second and third paragraph of section 4.2.1).

P8, Line 192: *I suggest that important locations discussed in the results, such as the Nahanni Plateau, be marked directly on the maps for clarity.*

We appreciate this suggestion and have added respective labels in Fig. 3.

Figure 7: *Needs a more detailed and explanatory caption.*

We have revised the caption of Figs. 6 and 7, and included a reference to the section 2.2.3 where the diagnostic is defined.

Comparison Figures (multiple subplots): *Consider presenting differences between simulations for a clearer quantitative comparison. Matching areas could be quantified in percentage terms.*

> Indeed, the panels b-i in Figures 4,6, and 7 show the anomalies (i.e. the differences of the means) with respect to the PI simulation shown in panel a. By this, all simulation results can be compared to the PI state, but also the differences between other simulations can be compared relatively well.
>
> We do not see how to objectively indicate “matching areas” in these comparison, since for most diagnostics there will be some deviations between all simulations in most gridcells. Thus, it would be necessary to define an arbitrary threshold for which values would still be considered “matching”.
>
> Instead, we have decided to use a more discrete colormap for presenting the results. By this, the regions where deviations are minor can be identified more easily compared to the continuous colormaps.

Conclusions: *The conclusion should be restructured to directly reflect and address the key objectives and issues raised in the introduction, creating a more cohesive narrative.*

> We have revised parts of the introduction as well as the conclusions section such that it provides an overall more cohesive narrative.

References:

1. Groenke, B., Langer, M., Miesner, F., Westermann, S., Gallego, G., & Boike, J. (2024). Robust Reconstruction of Historical Climate Change From Permafrost Boreholes. Journal of Geophysical Research: Earth Surface, 129(7), e2024JF007734. https://doi.org/10.1029/2024JF007734