Response to Reviewers

We thank the reviewers for their constructive comments and careful assessment of our manuscript, and we thank the Editor for their time and attention to our submission. Our responses to each comment are provided below, and the manuscript will be revised accordingly.

Reviewer #2

Major comment 1:

A major point of the results is that while the deployment of SAI in G6sulfur and G6solar are able to reduce the rate of decline of permafrost area and thaw depth compared to ssp585, the SAI deployment simulations rates of decline are still larger compared to ssp245, which is the considered to be the target of deployment (Figures 1 and 2). The authors test for significance using the two-sample Welch's t-test and state that the difference between the G6sulfur and G6solar simulations and ssp245 for the 2080-2099 period are statistically significantly different from ssp245 for permafrost area and thaw depth. The time series line (Figure 1a) and the spatial plots shown in Figure 2 in combination with the statistical significance testing is not a fully convincing argument that the results from the G6sulfur and G6solar simulations are significantly different from that in ssp245. To attempt to strengthen this argument, the authors should first provide more information about the details of the significance testing in the methods section (i.e., what is considered to be an independent sample (each model, each member?) and what the number of degrees of freedom are). Additionally, the authors should include the range of values predicted by individual ensemble members as uncertainty bars in plots such as Figure 1a (as is shown in Figure 4). The authors should also consider providing some measure of uncertainty related to the permafrost thaw depth difference plots shown in Figure 2a.

Response:

We thank the reviewer for this thoughtful comment. We agree that the original manuscript did not sufficiently describe the statistical framework used to assess differences between the SAI scenarios and SSP245, nor did it clearly distinguish between descriptive figures and formal statistical conclusions. In the revised manuscript, we will clarify the methodology, expand the description of uncertainty estimates, and include additional robustness analyses that address both interannual variability and inter-model uncertainty.

First, we would like to clarify the purpose of the figures. Figures 1a and 1d, as well as the spatial maps in Figure 2a, are intended to illustrate the temporal evolution and spatial patterns of permafrost extent and thaw depth. These panels are descriptive and are not intended to convey statistical significance. Formal statistical comparisons are presented in Figures 1b, 1c, and 2b.

For the original analysis, all available ensemble members for a given Earth System Model (ESM) and scenario were first averaged to produce a model-specific ensemble mean. The six model means were then averaged to obtain the multi-model mean, ensuring that each participating model contributed equally regardless of the number of ensemble members available. Statistical comparisons were based on annual values within the 20-year analysis periods.

In response to the reviewer's concern regarding the definition of independent samples, we re-evaluated the statistical significance using an autocorrelation-corrected framework. Because annual permafrost and thaw depth time series exhibit serial correlation, lag-1 autocorrelation was estimated from detrended annual anomalies following Zwiers and von Storch (1995), and effective sample sizes were calculated before applying Welch's t-test. Depending on scenario and period, effective sample sizes ranged from approximately 6 to 20 years, with corresponding effective Welch degrees of freedom ranging from approximately 10 to 31.

For permafrost extent, the autocorrelation correction does not alter the main conclusions. For example, the difference between G6sulfur and SSP245 during 2080–2099 remains statistically significant, with effective sample sizes of 8.9 and 5.7 years for the two scenarios, an effective Welch degree of freedom of 10.9, and a p-value of 0.003. Thus, the conclusion that G6sulfur produces significantly lower permafrost extent than SSP245 in the late twenty-first century remains robust after accounting for temporal dependence.

We further examined robustness across the six participating ESMs. Because the same models contribute simulations to all scenarios, model responses are naturally paired rather than independent. We therefore compared late-century (2080–2099) means on a model-by-model basis using paired analyses. This approach directly evaluates whether individual models respond systematically to the SAI forcing while accounting for differences in baseline permafrost extent among models.

For permafrost extent, the model-based analysis confirms that G6sulfur generally produces lower permafrost extent than SSP245, with five of the six participating models showing a reduction and a mean difference of approximately -1.15×10^6 km^2. This difference remains statistically significant in the paired framework. In contrast, G6solar also exhibits lower permafrost extent than SSP245, but the difference is smaller and does not reach the conventional 5% significance threshold ($p = 0.094$). These results indicate that the evidence for a late-century difference from SSP245 is stronger for G6sulfur than for G6solar.

We applied the same robustness framework to thaw depth. Using autocorrelation-adjusted annual time series, differences in domain-mean (= Northern High Latitudes region) thaw depth relative to SSP245 are generally modest and not statistically significant at the 5% level. For example, during 2080–2099 the total-domain thaw depth anomaly is +0.157 m for G6solar ($p = 0.44$) and +0.190 m for G6sulfur ($p = 0.20$). However, the regional response is substantially more heterogeneous than suggested by the domain mean. Significant differences emerge in the Western Hemisphere, and a pronounced east–west contrast is evident. The difference between eastern and western thaw depth anomalies yields a mean dipole magnitude of 0.769 m in the G6sulfur-SSP245 difference and remains highly significant after autocorrelation correction ($p < 0.001$).

To assess inter-model robustness, we additionally analysed model-specific late-century mean thaw depth. Domain-wide thaw depth differences relative to SSP245 are positive but not statistically robust across models (G6solar: +0.198 m; G6sulfur: +0.217 m). However, regional analyses reveal stronger and more coherent responses. For G6sulfur, the Eastern Hemisphere exhibits a mean anomaly of +0.480 m relative to SSP245, with agreement among five of the six participating models. Most notably, the east–west dipole emerges consistently across all six ESMs, with all six models agreeing on the sign of the dipole response in G6sulfur. Bootstrap confidence intervals and non-parametric tests indicate that this spatial pattern is substantially more robust than the domain-mean response. These results suggest that thaw depth changes under SAI are characterized primarily by regional redistribution rather than a uniform domain-wide shift.

Taken together, the additional analyses address the two main sources of uncertainty highlighted by the reviewer: interannual variability and inter-model structural uncertainty. The autocorrelation-corrected time-series analysis quantifies robustness against internal variability, while the paired model analysis evaluates consistency across Earth System Models. Across both frameworks, the results consistently show that the SAI scenarios substantially reduce permafrost loss relative to SSP585 throughout the late twenty-first century. Relative to SSP245, the evidence for a robust difference is strongest for G6sulfur and weaker for G6solar, consistent across methods. For thaw depth, domain-mean differences relative to SSP245 are not robust across analyses, and this lack of a uniform signal arises from a robust and spatially consistent east–west dipole pattern, in which opposing regional responses compensate in the domain mean.

In the revised manuscript, we will expand the Methods section and related supplementary materials to explicitly describe the treatment of ensemble members, the definition of independent samples, the autocorrelation correction, and the model-based paired analyses. Following the reviewer's suggestion, we will also re-evaluate the presentation of uncertainty in Figures 1 and 2 to better reflect model spread and ensemble variability where appropriate. For thaw depth, we include additional regional uncertainty diagnostics (model agreement and bootstrap-based confidence intervals) that support the interpretation of spatially heterogeneous responses.

Major comment 2:

The authors should consider reframing the introduction and include more details on the concepts surrounding permafrost including why permafrost acts as a carbon store and the transition of permafrost from a carbon store to a carbon source. This discussion should also explain why permafrost is a tipping element, and what implications permafrost collapse would have. Additionally, the authors should rewrite the introduction of the motivation to research the potential risks and benefits of methods of climate intervention, including SAI. While the critical role of Arctic permafrost and the risks posed by its thaw do provide motivation for research into the potential impacts of SAI, it is not the only motivator. The authors might point out that global mean temperature rise is projected to exceed 1.5°C in the next decade (IPCC, 2022), and then include a bit of discussion of some of the expected implications of exceeding this level of global temperature rise. The authors should consider removing discussion of smallscale field experiments that were cancelled, as well as private companies attempting to deploy SAI, as it is not relevant for discussion in this manuscript.

Response:
We thank the reviewer for this valuable feedback. We agree that the introduction would benefit from enhanced accessibility for a general audience and a broader framing and contextualization of the motivation for studying SAI. We will revise the introduction to:

1. Provide clearer explanations of why permafrost acts as a carbon store and the mechanisms underlying the transition from carbon store to carbon source.
2. Briefly discuss permafrost as a potential climate tipping element and the implications of crossing critical thresholds, while acknowledging that it is still being discussed if it constitutes a global tipping point (Nitzbon et al., 2024).
3. Reframe the SAI motivation by incorporating the IPCC projection that global temperature will exceed 1.5°C in the next decade, and discuss the broader climate risks that motivate research into climate intervention strategies.
4. Remove the discussion of cancelled field experiments and private company activities as suggested.

These revisions will make the manuscript more accessible while providing stronger scientific motivation and focus for the research.

Major comment 3:
There are a wide range of ensemble members used from each respective model (e.g., the number of members for CESM2-WACCM under ssp245 is 4-5, and the number of members for MPI-ESM1-2-LR is 50). Thus, just taking the mean across all members and models gives more weight to models that have more members. Further, for each simulation, the number of ensemble members available for each model varies. The authors should weight ensemble means so that each model (e.g., CESM2-WACCM, CNRM-ESM2-1, etc.) has an equal contribution to the ensemble mean so that each model contributes equally to the multimodel mean. Additionally, Table 1 shows that for some simulations, the number of ensemble members for some models varies (e.g., for CESM2-WACCM ssp245 simulations, 4-5 ensemble members are listed). I assume that this means that not all members of each simulation output each variable that is considered. Somewhere in the manuscript (e.g., Table 2, or in the supplement) what number of ensembles is available for each simulation in each model for each

variable should be listed explicitly. The weighting by number of ensemble members for each simulation for each model should then be adjusted accordingly.

Response:

We thank the reviewer for this important methodological clarification. We want to emphasize that our analysis does implement equal model weighting, exactly as the reviewer recommends.

As noted in our response to Major comment #1, our methodology follows standard CMIP/GeoMIP protocols: we first compute the ensemble mean for each model separately (averaging across all available ensemble members for that model), and then compute the multi-model mean by averaging across the six model means. This ensures that each model contributes equally to the final multi-model mean, regardless of the number of ensemble members (e.g., CESM2-WACCM with 4-5 members and MPI-ESM1-2-LR with 50 members each contribute equally).

We acknowledge that this two-step averaging process should be stated more explicitly in the Methods section. We will revise the text to clearly specify.

Regarding the notation "4-5" in Table 1: this indicates that not all ensemble members output all variables for that model-scenario combination (e.g., 4 members may provide variable X while 5 members provide variable Y). Since we apply equal model weighting at the model level (not the ensemble member level), this variation does not affect our results—each model contributes equally to the multi-model mean regardless of how many ensemble members it has for any given variable. The specific ensemble member counts for each variable are documented in the CMIP6/GeoMIP data archives and do not affect our results given the equal model-level weighting applied.

Major comment 4:

What is actually done in order to partition the variance contribution from each of the considered potential predictors is not explained with adequate detail. First, why are models constructed using only three predictors? What is the sensitivity of the results shown to the number of chosen predictors? LMG needs to be written out explicitly. What LMG is/does also needs to be defined and discussed. What does it mean to construct “all possible models comprising three predictors”? Why does the order of predictors in each model matter? How sensitive is this variance partitioning analysis to potential relationships (correlation) between predictors? In addition to adequate discussion of the details of the variance partitioning analysis, the interpretation of the results of this analysis should also be clarified. Some specific sections that need clarification are listed below:

- Lines 345-351: Why might downwelling longwave radiation, near-surface air temperature, and northward wind at 20 hPa explain the largest portion of the variance of the difference between the G6sulfur and ssp245 simulations?
- Lines 352-357: Why are the variables which describe the largest portion of the variance entirely different between the winter and summer, when both are to predict SPTD?
- Lines 386-393: This discussion hinges on the results shown in Figure 4. While there are subtle differences between the ensemble means of G6sulfur and ssp245, the range of values are overlapping in all cases, suggesting that none of the differences discussed in this section are significant. The authors might consider re-doing Figure 4, showing the G6sulfur-ssp245 for each individual model as its own line. Then, deviations from zero can be discussed.
- Figures 3c, 3d, 5c, 5d, and related subplots in Supplementary Figures 9 and 10 contain way too much information, most of which is unused. In section 3.3 specifically, the only correlations that are discussed are those between thaw depth (td) and the 18 predictors. The authors should revise these figures to only include the relevant correlations (the first column). There is no need to include the correlations of each predictor with the other predictors when those correlations are not included in any interpretation of the results.

Response:

We thank the reviewer for this detailed and constructive feedback on our variance partitioning methodology. We address each point systematically below.

PART A: METHODOLOGY

On the LMG method and its definition:

We apologize for the insufficient explanation of the LMG method in the current manuscript. We will substantially expand Section 2.3 and related supplementary material to provide a more detailed and accessible description of the variance partitioning methodology. Specifically, the

LMG method, named after Lindeman, Merenda and Gold (1980) and implemented in R using the relaimpo package (Grömping, 2006), is a variance decomposition technique that quantifies the relative importance of each predictor variable in explaining the variance of a response variable (in our case, summer permafrost thaw depth, SPTD) while accounting for correlations between predictors. This method has been used for earlier studies, including one on predictors of Arctic summer land surface energy budget (Oehri et al., 2022).

The method works as follows: for a linear regression model with k predictors, each predictor's incremental explained variance is computed across all possible orderings; from being entered first (its marginal R^2, ignoring all others) to last (its contribution beyond all others); and then averaged. For a confounded predictor (one correlated with others), the contribution shrinks the later it enters, since the shared variance has already been claimed; for an uncorrelated predictor, it stays roughly constant across orderings. This averaging of each predictor's marginal contribution over all orderings is the key feature of the LMG method that distinguishes it from simple correlation analysis or stepwise regression: it ensures that the contribution of each predictor is fairly assessed regardless of the order in which predictors enter the model, thereby mitigating the effects of multicollinearity among predictors.

On why models are constructed using three predictors:

The choice of three predictors per model balances the need to assess confounding and shared explanatory structure among predictors while maintaining interpretable model configurations. With 18 predictor variables, constructing all possible models with three predictors yields 680 unique predictor triplets, each tested across 6 possible predictor orderings, resulting in 4080 models in total. An additional advantage of this approach is that each predictor appears across many different triplets, yielding a distribution of explained-variance (LMG) values (Figure 3 and 5) rather than a single estimate. This allows each predictor's typical contribution to be summarized (e.g. using the median and interquartile range), while also quantifying how much its apparent importance varies depending on which other predictors are included. This variability is informative in itself, as it reflects the sensitivity of each predictor's contribution to model composition and thus its susceptibility to collinearity and confounding. In contrast, a single model including all 18 predictors would provide only one importance estimate per predictor and would not reveal this dependence on co-predictor structure.

On sensitivity to correlations between predictors:

The LMG method is specifically designed to handle correlated predictors. By averaging each predictor's marginal contribution across all possible orderings, the method distributes shared variance among correlated predictors in a fair and symmetric way. This is a fundamental advantage of the LMG approach over simpler methods such as stepwise regression or sequential R^2 decomposition, which are highly sensitive to predictor ordering and can produce misleading results when predictors are correlated. In addition, the use of 3-predictor models allows each predictor's contribution to be summarized across many model combinations (e.g., using the median and interquartile range of LMG values). This spread is itself informative, as it reflects how sensitive a predictor's explained variance is to the choice of co-predictors and therefore its susceptibility to collinearity and confounding effects. By contrast, a single relaimpo analysis including all 18 predictors would yield only one importance value per predictor and would not reveal this dependence on model composition. The correlation matrices shown in Figures 3c, 3d, 5c, and 5d provide additional information about the direction and strength of relationships between SPTD and each predictor, complementing the variance partitioning results.

Revised Methods text:

We will revise Section 2.3 to include the following expanded description:

We conduct a variance partitioning analysis using the LMG (Lindeman, Merenda and Gold, 1980) method, implemented in R version 4.1.3 using the relaimpo package (Grömping, 2006). This method quantifies the relative importance of each predictor variable in explaining the variance of summer permafrost thaw depth (SPTD) by averaging each predictor's marginal contribution to R^2 across all possible orderings of predictors entering the model. This averaging procedure ensures that the contribution of each predictor is assessed independently of predictor ordering, thereby accounting for correlations between predictors in a fair and symmetric manner. We construct all possible models comprising three predictors drawn from our set of 18 variables (a number high enough to assess confounding among predictors while low enough to avoid model overfitting, yielding 680 unique predictor triplets), test each across 6 possible predictor orderings (4080 models total), and report the average explained variance (%) for each predictor across all models.

PART B: RESULTS PRESENTATION AND INTERPRETATION

We thank the reviewer for these detailed comments on section 3.3 and the associated figures. We address each point systematically:

Regarding lines 345-351 and 352-357: The reviewer correctly notes that the physical reasoning behind the key predictors and their seasonal differences is not explained at the point of presentation. The physical mechanisms underlying these findings are discussed in detail in section 3.4, where we connect the identified drivers to large-scale atmospheric circulation changes induced by G6sulfur, particularly the NAO shift and associated changes in longwave radiation trapping and wind patterns. We will add explicit forward references at lines 345-351 and 352-357 directing readers to section 3.4, and will include brief explanatory sentences to clarify that winter drivers (longwave radiation, temperature) reflect NAO-induced warming and longwave trapping effects, while summer drivers (winds, cloud properties) reflect altered atmospheric circulation patterns.

Regarding lines 386-393 and Figure 4: We appreciate the reviewer's concern about the robustness of the described differences. We will revise lines 386-393 to use more cautious language that better reflects the descriptive and exploratory nature of Figure 4. Specifically, we will change 'Our results predict' to 'Our results suggest a tendency toward', add qualifiers such as 'generally' and 'tend toward', and acknowledge that these represent multi-model means where individual months and models may show varying degrees of agreement. Figure 4 serves a purely descriptive purpose, showing the seasonal cycle of meteorological parameters under each scenario, and we will ensure the revised language makes no implication of statistical significance. The statistical significance of the overall differences between G6sulfur and ssp245 is established in Figure 2b through formal statistical testing.

Regarding Figures 3c, 3d, 5c, 5d: We thank the reviewer for this suggestion. The full correlation matrices shown in Figures 3c, 3d, 5c, and 5d represent the standard output of the correlation analysis conducted in R, and we believe they provide valuable additional context for readers who wish to explore the inter-predictor relationships. While we acknowledge that the text discussion focuses primarily on the correlations between thaw depth (td) and the 18 predictors, the full matrices allow readers to assess potential multicollinearity among predictors, which is directly relevant to the interpretation of the variance partitioning results. We therefore propose to retain the full correlation matrices but will add a clarifying sentence in the figure captions explicitly directing readers' attention to the first column (td correlations) as the primary focus of the discussion.

Major comment 5:
The discussion connecting the NAO to differences in permafrost area and thaw depth between G6sulfur and ssp245 is not convincing as written.

Response:
We appreciate the reviewer's comment regarding the NAO discussion. We believe the connection between NAO shifts and permafrost thaw depth differences is well-supported by our analysis and existing literature. Specifically:

Our analysis demonstrates a statistically significant strengthening of the winter NAO index under G6sulfur compared to ssp245 (Figure 6e-f), with a NAO index increase of 1.02 (±1.10, P-value < 0.0001). Figure 6d shows that this NAO shift correlates strongly with key permafrost thaw drivers over the Eastern Hemisphere, particularly air temperature and cloud ice content. These correlations weaken with increasing longitude across the Eurasian continent, consistent with the spatial pattern of permafrost thaw depth changes observed in Figure 2a.

This NAO-driven mechanism is supported by recent literature. Jones et al. (2021, 2022) documented a positive NAO anomaly under G6sulfur in multiple models, and Banerjee et al. (2021) showed that intensified zonal flow associated with positive NAO phases drives high-latitude warming over Eurasia. Our results extend these findings by explicitly linking NAO shifts to changes in permafrost thaw drivers and spatial patterns.

We acknowledge that the complexity of the NAO-permafrost linkage could benefit from additional clarity in the manuscript. If the reviewer has specific concerns or suggestions for particular aspects of this discussion, we would welcome more detailed feedback to help us strengthen this section.

Minor comment 1: The Abstract should be edited to include mention of the G6solar simulations, which are not currently mentioned.

Response: We acknowledge that the Abstract currently focuses on the G6sulfur experiment and does not adequately highlight the G6solar simulations, which are an integral part of our analysis. We will revise the Abstract to explicitly mention both G6sulfur and G6solar experiments, making clear that our study compares both SAI deployment scenarios (aerosol injection vs. solar irradiance reduction) against the ssp245 and ssp585 pathways.

Minor comment 2: Please make sure that each citation has a publication date associated with it, there are currently a few that include no date.

Response: We thank the reviewer for this attention to detail. We have reviewed all citations in the manuscript. The three references without publication dates (SCoPEX, Make Sunsets, and The Geoengineering Model Intercomparison Project) are included in the reference list with retrieval dates as is standard practice for undated web resources. All other citations have appropriate publication dates.

Minor comment 3: In plots such as Figure 6a-6c, the authors should consider statistical significance testing by grid point, adjusted for the False Discovery Rate (Wilks, 2016), to show regions where the G6sulfur simulations differ significantly from ssp245, rather than field significance tests.

Response: We acknowledge that grid-point significance testing with FDR correction (Wilks, 2016) is a valuable approach for hypothesis testing in spatial climate data. However, our analysis in Figures 6a-6c is primarily descriptive and exploratory, aimed at understanding the spatial patterns and mechanisms by which G6sulfur alters atmospheric circulation compared to ssp245. The focus is on identifying regional differences in circulation patterns (sea level pressure, winds) that help explain the observed permafrost thaw depth variations, rather than formal hypothesis testing at individual grid points.

Minor comment 4: Lines 86-88: Add citation (e.g., Crutzen, 2006).

Response: We thank the reviewer for pointing out this omission. We acknowledge that Crutzen (2006) is the seminal work that first proposed stratospheric aerosol injection as a climate intervention strategy in the modern scientific literature. We will add this citation to lines 86-88.

Minor comment 5: Lines 84-85: Rephrase "strategies to manipulate the climate to reduce warming" to something like "strategies to prevent global temperatures from surpassing dangerous thresholds".

Response: We acknowledge that the phrase 'strategies to manipulate the climate' carries unintended negative connotations. We will revise lines 84-85 to use more appropriate language that frames these approaches as strategies to modify solar radiation to reduce climate warming, rather than as climate manipulation.

Minor comment 6: Lines 169-170: While Kravitz et al. (2015) state that the G6sulfur scenario was designed based off of radiative forcing reduction goal, Jones et al. (2022) state that this was eventually updated to a global mean near-surface air temperature goal. The manuscript should be updated to reflect this change.

Response: We thank the reviewer for this important clarification. While the original G6sulfur protocol (Kravitz et al., 2015) was framed in terms of radiative forcing reduction (4 W/m^2), the actual implementation evolved to use global mean near-surface air temperature targets to match SSP2-4.5 conditions, as documented in Jones et al. (2022). We will revise lines 169-170 and lines 591-594 to accurately reflect this temperature-based design goal and cite both Kravitz et al. (2015) and Jones et al. (2022) to show the evolution of the protocol.

Minor comment 7: Lines 172-174: cite O'Neill et al. (2017).

Response: We acknowledge that O'Neill et al. (2017) is the foundational paper that defined the Shared Socioeconomic Pathways (SSP) framework. We will add this citation to lines 172-174 to properly credit the SSP scenarios we use in our analysis.

Minor comment 8: Lines 176-182: This paragraph should be updated. The amount of aerosol or aerosol precursors injected in each model in G6sulfur is adjusted such that global mean near-surface air temperature is within 0.2°C of the global mean near-surface air temperature in ssp2-4.5, rather than a reduction in radiative forcing. It should also be noted that some of the models could not directly simulate the injection of aerosols into the stratosphere for the G6sulfur experiment, and so instead prescribe AOD distributions provided by GeoMIP (CNRM-ESM2-1) or by Niemeier and Schmidt (2017) and Niemeier et al. (2020) (MPI-ESM1-2-LR and MPI-ESM1-2-HR) (see Jones et al. (2022)). Lines 591-594 should also be updated accordingly.

Response: We thank the reviewer for this important methodological clarification. We acknowledge that our description of the G6sulfur implementation was incomplete and inconsistent. We will revise both lines 176-182 (Methods) and lines 591-594 (Conclusions) to clarify that: (1) aerosol injection amounts are adjusted in each model to match global mean near-surface air temperature targets (within 0.2°C of ssp245) rather than a fixed radiative forcing reduction, and (2) some models (CNRM-ESM2-1, MPI-ESM1-2-LR, and MPI-ESM1-2-HR) prescribe AOD distributions instead of directly simulating aerosol injection, following Niemeier and Schmidt (2017) and Niemeier et al. (2020). We will add the Niemeier et al. (2020) reference to our reference list and ensure consistency with Jones et al. (2022) throughout.

Minor comment 9: Lines 213-314: What is the complementary error function? Is it used to diagnose uncertainty? This explanation should be expanded.

Response: We thank the reviewer for this suggestion. We acknowledge that our description of Equation 1 could benefit from a brief clarification. We will add a short sentence noting that erfc is the complementary error function — a standard mathematical function that converts MAAT into a permafrost probability between 0 and 1 — following the established methodology of Gruber (2012).

Minor comment 10: Lines 216-218: Why are only grid cells with a land fraction greater than or equal to 35% and grounded ice sheet fraction less than or equal to 65% included? Has this been done before in similar studies? Include justification and citations here.

Response: We thank the reviewer for this important question. We acknowledge that these thresholds are not explicitly established in the existing literature but represent our own methodological choices. The 35% land fraction threshold was applied to exclude grid cells dominated by ocean or water bodies, allowing the analysis to focus on land-dominated grid cells where the permafrost probability relationship based on MAAT is expected to be applicable. The 65% grounded ice sheet fraction threshold was applied to exclude grid cells dominated by ice sheets (e.g., Greenland), where the permafrost probability relationship based on MAAT is not applicable. Since these filters primarily affect coastal regions, they have minimal impact on the overall permafrost area estimates. We will add this justification explicitly to lines 216-218 to make our methodological choices transparent.

Minor comment 11: Lines 222-223: Which models might have better (or worse) resolved soil temperature profiles?

Response: As shown in Table 1, soil temperature profile resolution varies considerably among the six models used in this study. Models with better-resolved soil profiles include CNRM-ESM2-1, IPSL-CM6A-LR, and CESM2-WACCM, which all feature 14 to 25 soil layers. The MPI-ESM1-2-LR, MPI-ESM1-2-HR, and UKESM1-0-LL models have the coarsest resolution with only 4 to 5 soil layers, which could introduce biases in soil temperature profile representation. We will clarify this in lines 222-223 by explicitly referencing Table 1 and noting which models have better or worse resolution.

Minor comment 12: Line 305: Why is thaw depth a critical element of permafrost stability?

Response: We acknowledge that our statement about thaw depth being "critical" needs better justification. We will expand line 305 to explain that thaw depth is a critical indicator of permafrost stability because: (1) it determines how much frozen soil organic carbon is exposed to decomposition, (2) it affects ecosystem functioning and infrastructure stability, and (3) it reflects the direct impact of surface warming on permafrost. We will cite appropriate references (such as Garnello et al., 2021) to support this explanation.

Minor comment 13: Line 463: What is the NAO index?

Response: We note that the NAO index is defined in Section 2.4 (lines 255-259), where we explain that the NAO is assessed through the pressure difference between Iceland and the Azores. However, to improve clarity for readers, we will add a brief parenthetical reminder at line 463 when the NAO index is first used in the Results section.

Minor comment 14: Lines 594-596: The sentence should be rephrased to say something like, "Although rates of decline of permafrost area and thaw depth are slowed in G6sulfur and

G6solar relative to ssp585, permafrost area and thaw depth declines more rapidly under SAI than in ssp245." This sentence should only remain in the manuscript if this result is valid after addressing the aforementioned methodological concerns.

Response: We appreciate the reviewer's suggestion. However, we believe our current framing better captures the complexity of the findings by explicitly highlighting the unintended atmospheric pattern changes induced by G6sulfur. While we acknowledge the positive effects relative to ssp585, we consider it important to emphasize that these changes come with regional trade-offs that limit the efficacy of this particular SAI strategy. We will consider adding explicit mention of G6solar for completeness.

References

Banerjee, A., Butler, A. H., Polvani, L. M., Robock, A., Simpson, I. R., & Sun, L. (2021). Robust winter warming over Eurasia under stratospheric sulfate geoengineering - The role of stratospheric dynamics. *Atmospheric Chemistry and Physics*, 21(9), 6985–6997. https://doi.org/10.5194/acp-21-6985-2021

Crutzen, P. J. (2006). Albedo Enhancement by Stratospheric Sulfur Injections: A Contribution to Resolve a Policy Dilemma? *Climatic Change*, 77(3), 211–220. https://doi.org/10.1007/s10584-006-9101-y

Garnello, A., Marchenko, S., Nicolsky, D., Romanovsky, V., Ledman, J., Celis, G., Schädel, C., Luo, Y., & Schuur, E. A. G. (2021). Projecting Permafrost Thaw of Sub-Arctic Tundra With a Thermodynamic Model Calibrated to Site Measurements. *Journal of Geophysical Research: Biogeosciences*, 126(6). https://doi.org/10.1029/2020JG006218

Grömping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. In *JSS Journal of Statistical Software* (Vol. 17). http://www.jstatsoft.org/

Gruber, S. (2012). Derivation and analysis of a high-resolution estimate of global permafrost zonation. *Cryosphere*, 6(1), 221–233. https://doi.org/10.5194/tc-6-221-2012

IPCC. (2022). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157940

Jones, A., Haywood, J. M., Jones, A. C., Tilmes, S., Kravitz, B., & Robock, A. (2021). North Atlantic Oscillation response in GeoMIP experiments G6solar and G6sulfur: Why detailed modelling is needed for understanding regional implications of solar radiation management. *Atmospheric Chemistry and Physics*, 21(2), 1287–1304. https://doi.org/10.5194/acp-21-1287-2021

Jones, A., Haywood, J. M., Scaife, A. A., Boucher, O., Henry, M., Kravitz, B., Lurton, T., Nabat, P., Niemeier, U., Séférian, R., Tilmes, S., & Visioni, D. (2022). The impact of stratospheric aerosol intervention on the North Atlantic and Quasi-Biennial Oscillations in the Geoengineering Model Intercomparison Project (GeoMIP) G6sulfur experiment. *Atmospheric Chemistry and Physics*, 22(5), 2999–3016. https://doi.org/10.5194/acp-22-2999-2022

Kravitz, B., Robock, A., Tilmes, S., Boucher, O., English, J. M., Irvine, P. J., et al. (2015). The Geoengineering Model Intercomparison Project Phase 6 (GeoMIP6): simulation

design and preliminary results. *Geoscientific Model Development*, 8(10), 3379–3392. https://doi.org/10.5194/gmd-8-3379-2015
Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis*. Glenview, Ill. : Scott, Foresman.
Niemeier, U., & Schmidt, H. (2017). Changing transport processes in the stratosphere by radiative heating of sulfate aerosols. *Atmospheric Chemistry and Physics*, 17(24), 14871–14886. https://doi.org/10.5194/acp-17-14871-2017
Niemeier, U., Richter, J. H., & Tilmes, S. (2020). Differing responses of the quasi-biennial oscillation to artificial SO_2 injections in two global models. *Atmospheric Chemistry and Physics*, *20*(14), 8975–8987. https://doi.org/10.5194/acp-20-8975-2020
Nitzbon, J., Schneider von Deimling, T., Aliyeva, M. *et al.* (2024). No respite from permafrost-thaw impacts in the absence of a global tipping point. *Nat. Clim. Chang.* 14, 573–585. https://doi.org/10.1038/s41558-024-02011-4
Oehri, J., Schaepman-Strub, G., Kim, J. S., Grysko, R., Kropp, H., Grünberg, I., Zemlianskii, V., Sonnentag, O., Euskirchen, E. S., Reji Chacko, M., Muscari, G., Blanken, P. D., Dean, J. F., di Sarra, A., Harding, R. J., Sobota, I., Kutzbach, L., Plekhanova, E., Riihelä, A., … Chambers, S. D. (2022). Vegetation type is an important predictor of the arctic summer land surface energy budget. *Nature Communications*, 13(1). https://doi.org/10.1038/s41467-022-34049-3
O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., et al. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180. https://doi.org/10.1016/j.gloenvcha.2015.01.004
Wilks, D. S. (2016). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bulletin of the American Meteorological Society*, 97(12), 2263–2273. https://doi.org/10.1175/BAMS-D-15-00267.1
Zwiers, F. W., & von Storch, H. (1995). Taking Serial Correlation into Account in Tests of the Mean. *Journal of Climate*, 8(2), 336–351. https://doi.org/10.1175/1520-0442(1995)008<0336:TSCIAI>2.0.CO;2