Response to referee Ian Shirley

We would like to thank Ian Shirley for the detailed and useful comments, which have helped to improve the quality and readability of our manuscript. In the following, we provide a reply to the points discussed by the reviewer as well as changes in the manuscript.

The comments of the reviewer are written in **bold**, the extracts of the manuscript in *italics* with changes highlighted in blue (added text) and red (deleted text). Line numbers are referring to the revised manuscript.

This paper introduces a slope stability index within the recently developed geomechanical module of the CryoGrid community model to evaluate slope instability associated with thawing ice-rich permafrost terrain.

The geomechanical scheme is based on established geotechnical formulations and is appropriately implemented. Its coupling to CryoGrid, which mechanistically simulates permafrost thermal and hydrological dynamics, represents an important advance for physically based prediction of thaw-driven slope instability under climate change. The work is timely given the expected increase in permafrost slope failures and their implications for infrastructure, landscape evolution, and carbon-climate feedbacks.

A particular strength of the paper is the attempt to bridge process-based permafrost modeling with regional-scale observations of slope failure occurrence using the proposed Thawing Slope Stability Index (TSSI). At Banks Island, Canada, the modeled slope stability index compares favorably with long-term records of retrogressive thaw slump (RTS) initiation. I do think, however, that the paper would benefit from a more explicit and nuanced discussion of this comparison, given that the model represents active layer detachment failures (ALDF) rather than the observed RTS failures.

The paper is very well written and clearly organized, and will make an important contribution to the field. Please see some more detailed comments below.

We thank Ian Shirley for his detailed feedback, which helped to improve the manuscript. We address the points raised above, particularly the comparison to the RTS dataset, in the detailed comments below.

1. Comparison between modeled ALDF and observed RTS initiation

The main conceptual issue I have with this manuscript concerns the comparison between the modeled instability and the observational dataset used for validation. The implemented mechanics represent ALDF-style instability, since

thaw consolidation and elevated pore water pressures drive reduced shear strength at the permafrost table. The observational dataset, on the other hand, consists of RTS initiation, which involves additional thermo-erosional and geomorphic drivers. I therefore think the manuscript should state much more explicitly throughout (including in the abstract/discussion) that the model is simulating ALDF-style instability rather than RTS initiation itself.

I do think the comparison to RTS observations is meaningful, particularly given the strong association between widespread RTS initiation and simulated ALDF, but the manuscript should spend more time discussing why this association exists. For example, ALDF may directly contribute to RTS initiation through exposure of massive ice and subsequent thermoerosional degradation of the exposed headwall. This idea is somewhat implicit in the manuscript, but is never explicitly discussed. At the same time, the manuscript overstates the degree to which RTS initiation can be attributed specifically to ALDF, e.g. "with warm summers causing the majority of RTS in the fieldsite, their initiation is likely connected to the scar zones of ALDF, rather than slope-undercuttings or blockfalls " (L185). I do not think this conclusion is yet fully justified mechanistically. Warm summers could plausibly enhance RTS initiation in different ways, including via thaw-driven ALDF, but also via warming accelerated fluvial or coastal thermoerosion. Indeed, the authors observe that RTS are strongly concentrated along rivers, lakes, and coastlines (L340), which suggests that undercutting mechanisms may still play an important role. Some of the agreement between model and observations may therefore reflect the fact that the meteorological conditions that increase susceptibility to ALDF also increase susceptibility to RTS in these environments, even where the precise triggering mechanisms differ.

Overall, I think the manuscript would be strengthened by framing the model more explicitly as a simulation of ALDF-style slope instability that may contribute to the observed RTS initiation through exposure of massive ice, while also recognizing that ALDF and RTS initiation may simply be promoted by similar meteorological conditions. This would help avoid implying that RTS initiation itself is being directly simulated.

We clarified throughout the entire manuscript (abstract, introduction, methods, discussion, conclusions) that the model represents the failure processes of ALDF, and explained the reasoning for using the RTS database for validation. To do so, we added a new section in the methods (2.5 Model validation). In this section, we refer to the paper presenting the RTS database, in which the authors conclude that the main trigger for the observed RTS are ALDF, while fluvial, lacustrine and coastal processes act more as pre-conditioners. Here an excerpt from Lewkowicz and Way (2019):

"RTS were initiated most frequently next to rivers (45%), and less often on slopes (27%), around lakes (23%) and at the coast (5%) [...] Notwithstanding relative change by locus of initiation, each of the four types of RTS increased by at least 20 times in absolute numbers. The finding that all types of RTS increased in the major initiation years (Fig. 2a) is unexpected. There is no obvious reason why processes as diverse as fluvial, lacustrine and coastal erosion or long-term talik development adjacent to lakes, all previously linked to the initial exposure of ground ice, should initiate large numbers of RTS in the same year. Our results, therefore, suggest an alternate mechanism. In a warm summer, as with active layer detachment formation, thaw consolidation at the base of the active layer or in the thawing transient layer leads to high porewater pressures, a reduction in effective shear strength, and slope failure where the factor of safety falls below unity. This exposes ice-rich permafrost protected beneath previously undisturbed slopes, or underlying the floors of previously stabilised RTS and triggers new RTS. Fluvial, coastal and lacustrine erosion remain important over the long-term, acting as pre-conditioners to RTS initiation by de-buttressing slopes through lateral erosion, vertical incision or thaw settlement. The major initiator of RTS on Banks Island, however, is a deepening thaw layer caused by particularly warm summers such as 1998. A progressive rise in mean summer air temperature due to climate change, punctuated by positive deviations from that moving average, is therefore expected to trigger new RTS (Fig. 2d)." (Lewkowicz and Way, 2019)

Line 7 ff.: *The model was tested using a multi-decadal database of RTS initiation for an area of 2300* km^2 *on Banks Island, Canada. We used RTS as a proxy for ALDF, since the majority of RTS in the field area were triggered in warm summers by exposure of massive ice following ALDF.*

Line 69 ff.: *In this study, we calculate the factor of safety within the framework of a land surface model and evaluate the potential for slope failure by introducing the Thawing Slope Stability Index (TSSI). This approach represents a physically-based, climate-dependent assessment of the susceptibility to ~~slope failure~~ ALDF in ice-rich permafrost.*

Line 148 ff.: *The existing model functionalities are extended to represent the processes associated with slope failures in ice-rich permafrost, specifically active layer detachment failures (ALDF).*

Line 261 ff.: **2.5 Model validation**

The model represents the processes resulting in active layer detachment failures (ALDF), however, no inventory of ALDF was available to support a statistical analysis of their occurrence. Therefore, we used a long-term inventory of newly initiated retrogressive thaw slumps (RTS) for validation (Lewkowicz and Way, 2019). The observed peaks in RTS initiation were associated with exceptionally warm summers and a deepening of the active layer. Lewkowicz and Way (2019) point out that the main trigger of RTS in the field area was ALDF exposing massive ground ice, while fluvial, lacustrine and coastal erosion may have acted as pre-conditioners to RTS initiation but cannot explain the high numbers of newly initiated RTS within the same year. Consequently, the RTS database can be used as a proxy for ALDF.

Line 323 ff.: *The modelling produced a range of values across the study area with positive median TSSI of 5.2-18.4 for 1998, 2010, 2011, and 2012 (Table 3 and Fig. 6a), indicating a high likelihood of ALDF~~and RTS initiation~~.*

Line 385 ff.: *This study advances the field in four main ways. First, it demonstrates that the geomechanical model that was incorporated in the CryoGrid community model by Aga et al. (2023) can be used within an infinite slope analysis to predict* *potential slope instability, which may lead to the initiation of ALDF ~~or RTS~~ and potentially RTS when triggered by exposure of massive ice through ALDF.*

Line 403 ff.: *Third, the validity of the TSSI as an indicator of slope instability was demonstrated by comparing it to a long-term inventory of newly observed RTS (Lewkowicz and Way, 2019). The ~~majority is linked to warm summer, so that their initiation is likely connected to~~ occurrence of new RTS peaked in warm summers, since the main trigger of RTS in this field area were thaw-driven ~~slope instabilities such as ALDF , rather than slope-undercuttings or blockfalls, so that the database was well suited for evaluating the model results.~~ ALDF exposing massive ice (Lewkowicz and Way, 2019). The database is therefore a suitable proxy for ALDF occurrence (Sect. 2.5).*

Line 479 ff.: *As a result of the limitations and uncertainties, the model captures the susceptibility to slope failure in ice-rich permafrost, specifically ALDF, at a landscape scale but it does not predict the locations of individual events.*

Line 493 ff.: *The presented model could be a valuable tool for further investigations of the triggers for ALDF ~~, in addition to RTS, in~~ in ice-rich permafrost environments.*

Line 507 ff.: *In this paper we describe the incorporation of a physically-based infinite slope analysis within the CryoGrid community model to develop a multi-decadal time series of the factor of safety for ice-rich permafrost slopes, representing the processes leading to active layer detachment failures (ALDF).*

Line 517 ff.: *The TSSI was tested in a study area on Banks Island where hundreds of newly initiated retrogressive thaw slumps (RTS) occurred between 1985 and 2015. ~~These~~ In this study, we use RTS initiation as a proxy for ALDF, as the exposure of massive ground ice through ALDF was reported as the main trigger for RTS initiation (Lewkowicz and Way, 2019). The newly initiated RTS developed in years with positive TSSI and were concentrated spatially in areas of the landscape with the highest TSSI values.*

2. Aspect controls on TSSI

In section 3.3.2, you discuss the somewhat surprising result that variation in aspect has relatively little influence on simulated TSSI. This is an interesting and potentially important result, given the strong topographic control on incoming shortwave radiation that is often assumed to influence thaw-driven slope instability. However, I do not think the implications of this result are fully explored, and the discussion around L365-370 somewhat overstates the role of aspect relative to the presented results. I think this discussion should be revised to better reflect the weak modeled aspect sensitivity shown in Fig. 6.

Since this weak aspect dependence is somewhat unexpected, I also think it would be useful to include at least a simple comparison of observed RTS occurrence versus aspect class. If observed RTS occurrence likewise shows weak aspect dependence, this would substantially strengthen the interpretation that landscape-scale variability in thaw-driven slope instability at this site is controlled more strongly by factors such as slope angle and ground ice conditions than by aspect-driven differences in radiation loading. On the other hand, if the observations exhibit a stronger aspect dependence than the model, this could point to an important model limitation.

The weak effect of aspect can be explained by the gently undulating terrain on Banks Island and by the effect of slope angle masking the aspect-related signal. As suggested by the reviewer, we included a comparison to the observational data in the results. We also included a separate analysis for slope angles < 4° in Fig. 7, to restrict the masking by steep slope angles. Furthermore, we extended the discussion.

Line 353 ff.: *The relationship between TSSI and aspect is weak (Fig. 7~~b~~a & c). High susceptibility occurs across all cardinal directions, and TSSI varies substantially even among similar aspects (e.g., east-facing slopes). This variability is ~~mainly due to slope angle~~ primarily controlled by slope angle (through the occurrence of steep slopes in all cardinal directions), which exerts a stronger influence than aspect and thus masks ~~aspect effects. Restricting the~~ any aspect-related signal. The weak dependency on aspect is also evident in the observed RTS: 25% are north-facing, 41% east-facing, 18% south-facing and 16% west-facing (each defined as the cardinal direction ± 45°). This distribution reflects well the peaks in TSSI shown in Fig. 7c. To reduce the masking effect of steep slope angles, Fig. 7d restricts the analysis to gentle slopes (<~~5°)reveals a slight increase in TSSI for south-east facing slopes~~ 4°), which reveals that the highest values for TSSI are found for east- and south-facing slopes, aligning with the expected impact of increased solar irradiation.*

Line 395 ff.: *Representing the topography by clustering proved essential because ~~both~~ slope angle and to a limited extent also aspect influence the TSSI. The slope angle appears in the equations to calculate the factor of safety and thus determines the TSSI. Aspect affects the TSSI indirectly by controlling the ground temperatures and active layer thickness, especially through differences in incoming short-wave radiation. ~~Although both parameters varied only slightly among the clusters, these minor differences led to notable variations in TSSI, underscoring their importance even for~~ The weak dependence on aspect indicates that variability in thaw-driven slope instability on Banks Island is only little influenced by differences in short-wave radiation, consistent with the similar weak signal in the observational data. In landscapes with steeper terrain and stronger topographic shading, however, the influence of aspect is likely to be more pronounced. But also the relatively small effect in the gently undulating terrain of Banks Island underscores the importance of slope angle and aspect when assessing susceptibility to slope failure.*

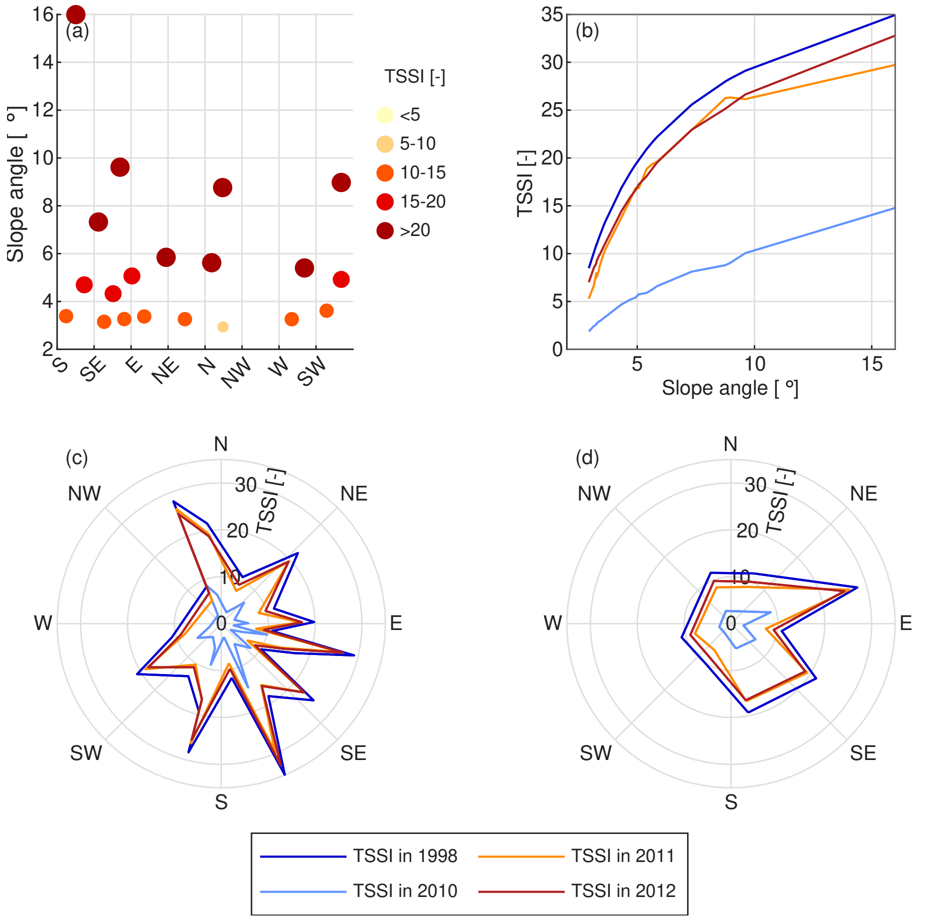

Figure 7. Relationship between the simulated TSSI ~~for years with median positive values~~ and (a) slope angle and aspect in the year 1998, (b) slope angle, (c) aspect ~~. Soil properties were assumed to be the same in all~~ and (d) aspect for clusters with slope angles <= 4°. Note: Years with a mean TSSI of zero are not shown.

Minor comments

Table 1 is missing entries for S and c', even though they are mentioned in the caption.

We added the saturation S in the table. c' was not added and removed from the caption, as it is not assigned by the user but calculated in every time step based on the ice content.

Table 1. The soil stratigraphy for the upper 9 m of the model domain after initialization, consisting of an organic layer at the top and the glacial till below. θ_m: volumetric fraction of mineral content; θ_o: volumetric fraction of organic content; θ_{wi}: total volumetric fraction of water and ice content; S: saturation; e_0: initial void ratio before compaction; σ_0: residual stress; C_c: compression index; $\phi^{'}$: effective friction angle in unfrozen conditions; ~~$c^{'}$: effective cohesion;~~ k_w: permeability; α: alpha coefficient; n: n coefficient.

Depth [m]	θ_m [-]	θ_o [-]	θ_{wi} [-]	S [-]	e_0 [-]	σ_0 [Pa]	C_c [-]	$\phi^{'}$ [°]	k_w [m^2]	α [m^{-1}]	n [-]
0 - 0.05	0.25	0.25	0.5	0.5	1.5	100	1	15	10^{-12}	4.06	2.00
0.05 - 1.00	0.54	0	0.46	1	1	1000	0.5	20	10^{-13}	1.49	1.25
1.00 - 9.00	0.18	0	0.82	1	1	1000	0.5	20	10^{-13}	1.49	1.25

“do now show any TSSI above zero” (L333) should likely read “do not show any TSSI above zero.”

We corrected the typo.

Line 362 ff.: *Years with stable conditions, such as for example the year 2000, do ~~now~~ not show any TSSI above zero and appear completely transparent.*