

The Cryosphere Discuss., referee comment RC2  
<https://doi.org/10.5194/tc-2022-4-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on tc-2022-4

Anonymous Referee #2

---

Referee comment on "Post Little Ice Age rock wall permafrost evolution in Norway" by Justyna Czekirda et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2022-4-RC2>, 2022

---

The work presents results from 2D numerical modelling of heat conduction from several transects crossing steep rock walls in Norway. The aim of the study is not explicitly stated (it needs to be), but 1.92-93 and the choice of sites suggest the aim is to investigate bedrock temperatures behind steep rockwalls and the adjacent landscape, including at or near the sites of known slope instabilities. At least one transect crosses an active instability (Gámanjunni).

Reliably modelling bedrock temperatures down to depths of relevance for the evolution of slope instabilities where permafrost may exist is very important and within the scope of TC. However, it seems that the model used has never been validated against borehole measurements even in simpler topographic settings and thermal systems. No borehole data for validation is presented here either. By its formulation, the model cannot account for effects that are especially important in landscapes susceptible to slope instability, where more pervasive and widening fractures are to be expected. Snow is also treated in a very simple way, and several parts of the transects are not oriented favourably to the topography for a 2D model. The resulting uncertainties are impossible to quantify through sensitivity experiments without real borehole data.

Boreholes data suitable for comparison do exist, and are claimed to agree 'quite well' with the model. But they are not shown. If the current level of detail in the presentation and discussion of the model results is to be preserved, the validation against at least this set of boreholes needs to be included. The good agreement would build some confidence in the model performance. Otherwise I suggest to significantly shorten the ms. and only focus on discussing those feature of the modelled bedrock temperatures that are broadly consistent across different sites or parts of different transects featuring similar conditions.

I am particularly concerned because one of the cited papers (Kristensen et al. 2021) already applied the same model, again without borehole measurements, to an unstable slope that it described as highly fractured and with seasonally varying permeability. As

such, a site completely unsuitable for a model only accounting for heat conduction. We still know very little about the links between temperatures and slope stability, and false leads from unproven models can cause unnecessary confusion. More details on the limitations of the model and why some of the broadly accepted assumptions would not apply here are below in my comments to the text.

The ms. is not suitable for publication in its present form but see below for comments and suggestions that may make it acceptable after major revisions.

Throughout the text: improve the use of articles, e.g. l. 58 'the Northern Norway' -> 'Northern Norway', l. 100 'Permafrost limit' -> 'The permafrost limit', and in general improve the use of the English language and remove repetitions.

l. 15 'selected': explain (here or in §1 or §2 if there it takes more than a few words) selected based on what

#### §1

l. 67 'attributing variations in mountain permafrost occurrence owing to' -> 'attributing mountain permafrost occurrence to'

l. 80 'selected sites', 'other sites' summarize how they chose those sites, as the choice can introduce biases in the derived works.

l. 87 the model uses forcing calibrated with observations but the model is not 'observation-constrained' because no subsurface measurements are used to constrain the 2D model results.

l. 91-93 this hints to the aims of the work but they need to be more clearly and explicitly stated.

#### §2:

This section needs to state clearly how the selected sites were chosen, how each transect was drawn, why none of the sites with boreholes were modelled, or if they were, why no comparison is shown between model and measurements.

Most of the place names mentioned in this section are not shown in fig. 1 nor any of the map so they aren't useful to any reader who isn't already familiar with these regions. Places that really need to be mentioned usually also deserve to be shown on a map.

Consider reducing unessential info on geography, tectonostratigraphic units and mineralogy, and descriptions of slope instabilities at other sites than the modelled profiles unless they share some fundamental similarity with the latter.

### §3.1:

If CryoGrid 2D has been compared to actual borehole measurements, cite the relevant publications, summarize the observed performance of the model, and indicate any difference in the model configuration used here. If not, mention that the model has not yet been validated. The point here is that heat conduction and numerical solvers are well understood tools, but any new implementation needs to be proven correct. Additionally, the impacts of not accounting for convective and advective heat transport, fractures (both filled or open, with or without circulation of air or water), as well as the chosen ways of accounting for surface material and snow all reduce the accuracy of the model by an unknown margin. This is especially important in landscapes prone to slope instability and close to freezing temperatures, such as many of those discussed here, where more pervasive and widening fractures are to be expected and the any infiltrating surface waters can have a large impact on stability.

### §3.2:

State how the exact trace of the profiles was drawn and how its geometric relation with the topography may influence the results. Fig. 1 seems to show that many of the chosen profiles have long sections running at a slant angle from the slope, e.g. the first half of Hogrenningsnibba and most of the Kvernhusfjellet profiles. The Mannen profile between 1000 and 1500 m even follows the top of a ridge. While some compromise is unavoidable when dealing with real topography, these profiles are very far from the requirement of a 2D model that lateral heat fluxes through the plane of the model are zero.

### §3.3.1

I. 239-240 unclear whether the trend is with elevation or with time.

I. 243 'air inversions' -> 'temperature inversions'

I. 246 'coverage' -> 'resolution' (?) Also state what the resolution is.

I. 247 'each dataset' -> 'each profile' (?)

I. 252 (step 2) doesn't this bring back the problem that the nearby valley bottom meteorological station or seNorge grid cell overestimate cold periods?

I. 254-255 (step 3) this lapse rate will be affected by inversions, is this intentional?

### §3.3.2

The method used to make up for the lack of snow cover observations is reasonable, but it is quite simplified and uses many arbitrarily set values (Table 3) so again the model performance will be affected in an unknown way without borehole measurements.

### §3.4

I. 302 'correlation'?

I. 318-320 Explain what is an uncertainty run, what a control run (never mentioned anywhere else in the text) and what are the uncertainty simulations mentioned in the text but not here. Are 'run', 'simulation' and 'scenario' synonyms throughout the text?

### §4.1

I.326-328 'observations-constrained modelling...': it's more clear for the reader to call this 'calibration of GST forcing input using the measured SOs', similar to how it is already worded later on at I. 696. Once calibrated, of course the mean error is 0.00 for all 20 loggers, so the captions of figures S1 to S20 calling it 'validation' are misleading since the comparison and statistics are done against the same data used for calibration.

### §§4.2-4.3-4.4 (I. 329-557)

These sections are much too long compared to what the reader gains from them. It's not necessary to describe in words everything that is visible in the figures, unless some feature contradicts other evidence or solves an open questions about some important point along a transect. Commenting each scenario and the details of each site is quite uninteresting because the model hasn't been validated using borehole data neither at these sites nor elsewhere (otherwise provide reference). So the model outputs remain speculative and cannot be used to draw such detailed conclusions as described here about any one specific site.

I suggest these three sections are shortened and rewritten so that instead of discussing in unwarranted detail the state of each site under each scenario, they discuss how similar profile features (e.g. shape of the slope, presence of blockfields) and model scenario across different sites may lead (or not) to similar model outputs. This would show whether the model behaves consistently under similar conditions. Some of this is currently in §5.3. If instead your primary focus is on describing the conditions at each site, consider consolidating all info on each site in one subsection, discuss primarily the 'main' scenario and mention the other scenarios when they clearly improve the understanding of the 'main' scenario.

### §5.1.1

Four major limitations are not mentioned:

- the correctness of the CryoGrid 2D code has not been validated against observations either in this study or in the literature. Magnin et al. 2017 notes that validation against boreholes measurements is rare due to lack of data and because established heat conduction models are assumed to be reliable under simple thermal systems. But CryoGrid 2D is not yet an established code. Myhra et al., 2017 didn't have subsurface measurements, Kristensen et al. 2021 applied it to an unstable slope without any validation dataset, and in a setting that is clearly not a simple thermal system based on the extreme heterogeneity of the unstable volume shown in their fig. 5 and on their finding that permeability varies with temperature.
- most of the sites chosen in this study are close to unstable slopes where fractures can be expected to be particularly frequent, pervasive and quite possibly open or filled with ice or water, i.e. not at all a simple thermal system that can be modelled with confidence by assuming heat conduction dominates below just a few metres.
- the chosen profiles, especially (but not only) at Hogrenningsnibba, Kvernhusfjellet and Mannen have long sections where the geometric relationship to the slope (e.g. crossing it at a slant angle, or running along a ridge) is such that significant lateral heat flux through must be expected, violating a basic assumption of 2D models. Clearly this type of profiles and topography are not what Myhra et al., 2017 referred to when arguing for the suitability of 2D models in the Norwegian mountains because of their flat plateaus and long valleys.

#### §5.1.2

The use SOs calculated at each logger from GST measurements is mentioned in many places as a key strength of the study but where the SOs from different loggers are used is not well explained except for table 4. Each transect runs across a range of elevation, slope and aspect widely deviating from those of the closest logger. It seems that most transect 'main runs' use different loggers for their easternmost and westernmost slopes, is this so? Then some are run with alternative loggers too. Calling these 'sensitivity scenarios' is confusing because they are not investigating the sensitivity to some parameter of the model configuration, but rather creating alternative forcing scenarios based on less realistic calculated SOs. Magnin et al. (2017) remarks that accurate GST are a prerequisite for a scheme without surface energy balance and without snow. Given the mentioned frequent occurrence of temperature inversions, if GST is only calibrated at one logger along a transect (or one per mountain side) and then used everywhere along that transect or mountain side, accurate GST can't be assumed and the reference to the good performance below 6-8 m of the Magnin et al. (2007) model is unwarranted. See also my note to figg. S1-S20

#### §§5.1.3, 5.1.4

Because of the very effects described in these sections, snow is another element that makes it impossible to rely on the performance of the model without validation against borehole measurements. Especially when most of the nF values in Table 3 are rather arbitrary round numbers.

The claim that snow effects are in some measure also accounted for through the use of mean monthly SOs computed from GST measurements is significantly weakened if only one logger per scenario is used along each entire transect, or side of the mountain, as discussed in my comment to §5.1.2.

## §5.2

Is this section's primary focus on finding whether GST is a reliable indicator of permafrost by comparing where  $GST < 0$  vs. where the 2D model predicts permafrost? Or on comparing the 2D model predictions vs. other studies that used SAT or GST predicted permafrost? Or is the aim is to suggest what the most likely conditions at each site are?

I recommend focusing on a brief and clear comparison between the permafrost/no permafrost prediction of the model match earlier published studies.

l. 733-735 Given that the major weakness of the ms. is that the 2D model is unproven, mentioning the existence of data from some boreholes in Jotunheimen that are suitable for comparison without showing such comparison is disconcerting. Claiming they 'agree quite well' is unwarranted. The comparison with borehole data needs to be shown and documented, especially as the boreholes are said to span different snow conditions. If this is not possible, the existence of the data needs to be mentioned for future reference but without making unsupported claims of good model agreement.

## §5.3

This is by far the most valuable part of all §5 and in fact the section of the manuscript that I feel provides most insight, even considering the limitations of the underlying model. It may be expanded a bit after other sections are shortened.

## §6

These conclusions need to be qualified by stating that the model used has not been validated at depth and that it cannot account for effects that are especially important in landscapes prone to slope instability, such as many of those discussed here, where more pervasive and possibly open fractures is to be expected. Also mention (maybe in the Introduction) that a heavily fractured rock mass is thermally different and can be expected to have deeper reaching impacts on temperature compared to the local and shallow effect of a single fracture some metres below the rock face, like the one described in Magnin et al., 2005.

l. 803 'leaded' -> 'led'

l. 812-814 this contradicts the claims in the earlier sections that the model results below 6 m are insensitive to the details of how snow is treated.

l. 821 this confirm that the model isn't necessarily accurate accurate below a few metres if only one SOs calculated from one GST measurement is used over an entire transect or mountain side for each scenario.

l. 831-831 is any evidence for this discussed in the ms.?

fig.1: add some (labeled) elevation contour lines and any place name mentioned in the text. If other important place names have to be mentioned in the text but are outside these areas, add a map showing them. Or remove them from the text.

fig.3: the meaning of each letter is in Table 1, not 2

fig. 4-5-6: The caption mentions both sensitivity and uncertainty runs but the differences between these two types of runs is not explained anywhere in the ms. Are they all in Table 5 where they are called sensitivity scenarios? Which ones are uncertainty runs? Please add to the main text an explanation of the meaning of the 'sensitivity test maximum range in the modelled maximum GT' and how it may relate to the actual uncertainty in the model. It seems quite optimistically low close to surface and surprisingly high at large depths. The maps need to be of the same style and they need to show clearly where the sites are in the broader orography and distance from the sea. Roads etc. are not essential.

fig. 10, 12: individual lines are very hard to see

figures S1 to S20: these do not show validation of GST (here called RST which is not consistent with the rest of the ms.), they simply show that the calibration was correctly done, resulting in mean error of exactly 0.00 C at the site of each one of the 20 loggers.