Response to Anonymous Referee #1

We would like to thank the reviewer 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 the 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*.

Major comments:

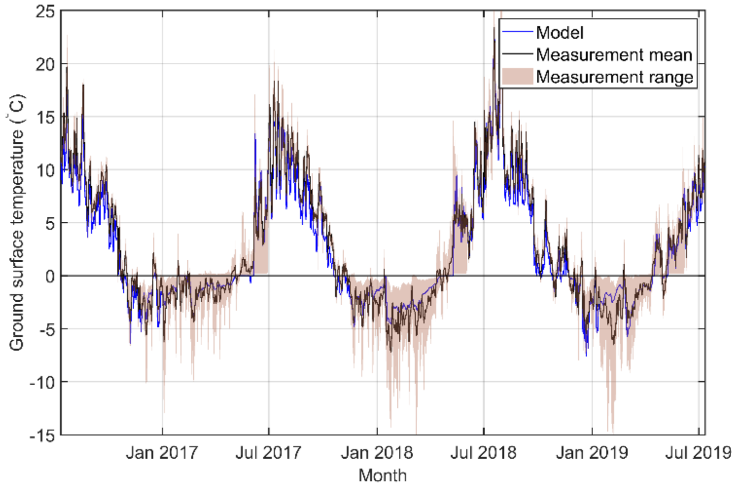
1 The idea of the model experiment – testing the impact of drainage on the ground thermal regime – is very interesting and the approach seems sound and straightforward. However, the model validation and the links to the two sites in Norway are weak. For one of the sites there are observed temperatures available to compare with the model results but these data are not used to validate the model results in a robust way. Instead, validation is only carried out visually and no objective statistics are used to evaluate the model results. For the other site, there are no data on ground temperatures but only observed ground surface temperatures. The validation and comparison to observations from the Norwegian sites thereby adds next to nothing to the modeling study. My suggestion would be to either remove these sites from the manuscript, to add some statistical analysis based on the available ground temperature data, and/or to find a rock glacier site with available ground temperature data which could be used for a real validation of model results.

The two Norwegian field sites are of high importance to the study. They are used for validation with the available data sets and the comparison of observed and modelled data shows that the model can capture the main characteristics of the seasonal evolution of measured temperatures. We agree with the reviewer that temperature data from boreholes would be beneficial. However, as these do not exist for Norwegian rock glaciers, Ivarsfjorden is a well suited field site for this study, as it is well documented by a previous study (Lilleøren et al., 2022) and near-surface ground temperatures are recorded by several temperature loggers. Furthermore, the presence of the rock glacier landform itself, as well as geophysics investigations on ground ice presence, allows to constrain the permafrost evolution at least in a qualitative fashion (Lilleøren et al., 2022), and his can also be compared to the simulation results. We therefore decided not to remove the site from the study, but improve validation. To do so, we have followed the suggestion to include a statistical analysis of the model performance at both sites. Secondly, the model validation for the rock glacier site (Ivarsfjorden) has been redone in a much improved and more robust manner. Please find below the changes in our manuscript.

We added a more detailed description of the temperature loggers at Ivarsfjorden to explain the setup for model validation:

Line 287-291 : At Ivarsfjorden, we considered 11 temperature loggers within the rock glacier outline (Fig. 1d in Lilleøren et al., 2022), of which all except for one are placed on the relict surface of the rock glacier (Fig. 2a in Lilleøren et al., 2022). On the relict surface, deposition of finer sediment in between blocks is more likely than on the active surface, due to the lack of movement. Here, the blocks with sediment stratigraphy is considered appropriate and used as starting point for the calibration.

Temperatures resulting from the best-fitting model setup are now presented in a similar way as Juvvasshøe, showing the model can reproduce daily ground surface temperatures in a satisfactory manner.



Line 349-352 : Figure 4: Daily modelled and measured ground surface temperatures in Ivarsfjorden from July 2016 to July 2019. The shaded area indicates the minimum to maximum range of measured daily values from 11 loggers (based on Lilleøren et al., 2022), while the black line represents the mean value of all loggers.

At both of the sites, the bias and RMSE are calculated for daily averages.

Line 291-293 : At both sites, the root-mean-square-error (RMSE) and bias are calculated in order to provide an objective measure of the model fit. At Juvvasshøe this was accomplished for daily values at 0.4 m and 2 m depth, while at Ivarsfjorden the mean daily ground surface temperature of the loggers within the rock glacier outline is used.

Line 328-330 : This configuration used drained conditions, although differences with undrained conditions are minimal for this stratigraphy. For daily temperatures at 0.4 m depth, the RMSE and bias are 2.1 °C and -0.6 °C, respectively, while they are 1.2 °C and -0.7 °C at 2 m depth.

Line 340-341 : The best-fitting model configuration was found to be the blocks with sediment stratigraphy and a snowfall factor of 1.0, resulting in an RMSE of 1.3 °C and a bias of -0.4 °C.

2 The methods section would benefit from restructuring and extensive editing, because it is rather confusing as currently written.

We agree that the method chapter was not clearly enough structured and required restructuring and clarification. In the revised manuscript, we have completely restructured the Methods section for clarity.

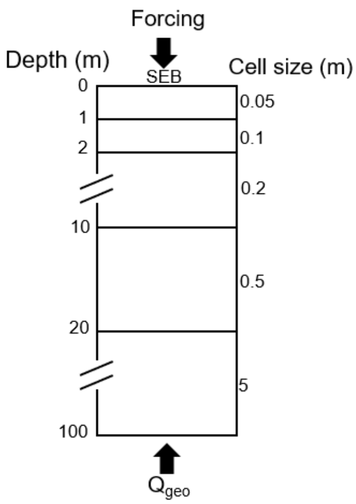
Section “3.1 The CryoGrid community model” contains a general description of the used model, not covering any site specifics or parameter values. We removed extensive descriptions of CryoGrid capabilities which are not used in our model setup. This section is followed by “3.2 Downscaling of model forcing”, a description of the model forcing data, which gives the reader information about the time period and time resolution of the model forcing. Section “3.3 Model setup” now covers the model setup and parameters, which were previously scattered throughout the chapter. As suggested by the reviewer, we created separate subchapters for the Validation runs (3.3.1), Equilibrium runs (3.3.2) and Transient runs (3.3.3). Specific improvements / clarifications in the Methods will be highlighted at the corresponding review comment.

2.1 Model description. It would be helpful to start the model description with a presentation of the mesh used: is this a vertical column (1D) mesh? What is the thickness and cell size of the domain?

Some of this information is available in other parts of the methods section, but starting off by clarifying such basic facts would help the reader to envision the model setup.

In the process of improving and clarifying the methods section, we followed the reviewer's recommendation and included a schematic in the beginning of the methods section, that shows the model domain with its depth, cell sizes and boundary conditions. In this manner, the reader is directly informed about how the model domain looks.

Line 154-155: We use a one-dimensional model column with a domain depth of 100 m (as in e.g. Westermann et al., 2016; Schmidt et al., 2021) and grid cell sizes increasing with depth (Fig. 2).



Line 165-168: Figure 2: Schematic of the model grid, indicating cell sizes at different depths and upper and lower boundary conditions. As upper boundary condition, the surface energy balance (SEB) forced by near-surface meteorological data is used. The lower boundary condition is provided by a constant geothermal heat flux at 100 m depth.

2.2 Snow model. A list of processes included in the model is provided, but there is no explanation of how these are represented in the model. If these are explained in previous publications, please clearly refer to those for the specific processes listed. For example, "the physical effect of wind drift on the snowpack" is included, according to the text, but later it is stated that the redistribution of snow by wind is phenomenologically represented using the snowfall factor. So, what effects of wind on the snowpack are actually included? Does the snowfall factor allow for variable distribution of snow on the domain, or is there also a redistribution of snow over time?

We agree with the reviewer that better explanations and referencing is needed regarding the implemented snow processes in the model. We therefore refer to Vionnet et al. (2012) for defining equations of the Crocus snow model, and to Zweigel et al. (2021) for implementation in CryoGrid). Furthermore, we improved the explanation of the snow model in the manuscript, especially clarifying the snow redistribution.

Line 189-197:The snow model used in this study was introduced by Zweigel et al. (2021) and is based on the Crocus snow scheme (Vionnet et al. 2012) which accounts for snow microphysics and is designed to reproduce a realistic snow pack structure (see Vionnet et al. 2012 for defining equations; Zweigel et al. 2021 for implementation in CryoGrid). Snowfall is added with density and microphysical properties derived from model forcing data, in particular air temperature and wind speed. The snow density evolves due to compaction by the overburden pressure of overlying snow layers, as well as wind compaction and refreezing of melt- and rainwater (Vionnet et al., 2012). The amount of snowfall from the forcing data can be adjusted by a so-called snowfall factor, sf, with which the snowfall rate from the model forcing is multiplied. With this, the effects of wind-induced snow redistribution on ground temperatures can be *represented at least phenomenologically (Martin et al., 2019), using sf < 1 for areas with net snow ablation and sf > 1 for areas with net deposition.*

2.3 The description of the different model runs (validation, equilibrium, and transient) is rather confusing and split up in different sub-sections of the methods description. I would suggest to clarify the purpose of the model runs and explain them in a separate sub-section. The validation runs are runs validated based on visual fit, which is not a very robust metric for model validation. It seems like the purpose of validation runs was to determine (calibrate?) the proper sediment stratigraphy for each site. But how were other parameter values determined? Table 1 presents some physical properties, but not e.g. Kh. Snowfall factor is for some reason tested in the equilibrium runs (why?). In general, the purpose of testing the snowfall factor sensitivity is rather unclear to me. The outcome of this test is not mentioned in the abstract, but the current text does not suggest that this is pure model calibration. L230 suggests that snowfall factors were also tested in validation runs, but this is not mentioned previously.

As mentioned in the response to major comment 2, we followed the suggestion to split the different model runs in sub-sections and clarified their purpose. The manuscript includes now the sections "3.3.1 Validation runs", "3.3.2 Equilibrium runs" and "3.3.3 Transient runs" and the purpose of the model runs has been clarified. Please refer to the corresponding sections of the revised manuscript for all changes. Specific comments are addressed below.

<u>Model validation</u>

We clarified the goal of the model validation:

*Line 271-272**:** Validation runs are set up to show that the model can reproduce key characteristics of the thermal regime at the two sites in a satisfactory manner (based on available observations).*

We agree with the reviewer that a visual fit is not sufficient, therefore we included a statistical analysis of the model performance and significantly improved validation for the rock glacier site (for changes in the manuscript see response to major comment 1).

We now describe that snowfall factors are tested in the model validation:

Line 282-283: Manual adjustment of the ground stratigraphy (porosity and thus mineral content) and snowfall factor are performed until a good fit with daily measurements is achieved.

<u>Equilibrium runs</u>

It is important to test different snowfall factors, as the thermal regime is highly sensitive towards snow conditions and the snow cover shows a high spatial variability in Norwegian mountains. Therefore, we investigated the sensitivity of ground temperatures by running the equilibrium runs with a range of snowfall factors, to estimate the threshold snow amount for permafrost existence in different model scenarios. We added further explanations to the revised manuscript by including previous studies and highlighting the purpose of the model runs.

Line 295-301: The goal of equilibrium runs is to investigate the sensitivity of the ground thermal regime towards ground properties and drainage conditions, using both the undrained and drained setup for the three idealized stratigraphies (Table 1), which result in six scenarios. As the heavily wind-affected snow cover is a key source of spatial variability in ground temperatures in the Norwegian mountains (Gisnås et al., 2014; Signåa et al., 2016), the model is run for a range of snowfall factors between 0.0 and 1.5 (Table 2) for each scenario. This analysis allows us to identify the magnitude of the thermal anomaly that the subsurface drainage induces at various amounts of snow, as estimate the threshold snow amount for permafrost existence in the six scenarios.

The outcome of this sensitivity test is now also included in the Abstract and Conclusions

Line 24-25: The thermal anomaly increases with larger amounts of snowfall, showing that well drained blocky deposits are less sensitive to snow insulation than other soils.

Line 589-591: The largest anomalies occur in simulations with a thick winter snow cover as ground temperatures in well drained blocky deposits are less sensitive to insulation by snow than other soils.

Model setup

We have provided further details on the choice of the hydraulic conductivity and the parameter d^{lat} and also present further sensitivity studies in a Supplement:

Line 245-255: To investigate the effect of subsurface drainage ground temperatures and ground ice conditions, we distinguish undrained and drained scenarios by using two different values of d^{lat} (Eq. 1) for in the idealized stratigraphies. A d^{lat} value of 10^4 m is used for undrained cases, which emulates conditions at a flat surface, resulting in a to a good approximation one-dimensional water balance, where only surface water is removed. For the drained cases, a d^{lat} value of 1 m is used, which results in well-drained conditions which are typical in sloping terrain. For the saturated hydraulic conductivity K_H, a fixed value of 10^{-5} m s^{-1} is used for all stratigraphies, although the true hydraulic conductivities almost certainly differ between stratigraphies. However, the key parameter controlling lateral water fluxes in Eq. 1 is in reality the "drainage timescale" K_H/d^{lat} [s^{-1}], which is varied by four orders of magnitude between K_H/d^{lat} = 10^{-5} s^{-1} (d^{lat} = 1 m, well-drained conditions) and K_H/d^{lat} = 10^{-9} s^{-1} (d^{lat} = 10^{-4} m undrained conditions). As the study setup is designed to analyse these two "confining cases", it is sufficient to only vary d^{lat} and leave K_H constant for simplicity. Further sensitivity tests for d^{lat} and K_H are provided in the Supplement.

2.4 The forcing data used for spinup and simulation runs are also not presented in a clear manner. The spinup and initialization procedure and data vary between the sites and the runs and the motivation for the choice of spinup periods is not always clear. Perhaps a table presenting the basic details of each run would be more helpful. Also not clear in the current text: Is forcing data downloaded and downscaled for the two sites (e.i., two separate series of data)? What is the time resolution (daily? monthly?) of the data used for simulations and for validation of simulation results?

We restructured the section "3.2 Downscaling of forcing data" and made distinct changes to the explanations, so that the procedure is now presented in a clearer manner. Specific comments are addressed below.

We present a table that gives an overview about the simulation types and includes information about the model forcing.

Line 265-268: Table 2: Overview of basic model settings for the different simulations types. A spin-up of subsurface temperatures is achieved by repeated simulations for the spin-up period (until a stable temperature profile is reached), before the actual model run for the simulation period is conducted. "Idealized" stratigraphy and drainage refers to three subsurface stratigraphies (Table 1) combined with two types of drainage conditions. See Sect. 3.3.1 and Sect. 3.3.3 for details.

Simulation type	Site	Spin-up period	Simulation period	Stratigraphy and drainage	Snowfall factor
Validation	Juvvasshøe	1951-2010	2010-2019	Best-fit	0.25
	Ivarsfjorden	1951-2016	2016-2019	Best-fit	1
Equilibrium	Juvvasshøe	2000-2010	2000-2010	Idealized	0.0, 0.25, 0.5, 0.75, 1.0, 1.5
	Ivarsfjorden	1962-1971	1962-1971	Idealized	0.0, 0.25, 0.5, 0.75, 1.0, 1.5
Transient	Juvvasshøe	1962-1971	1951-2019	Idealized	0.25
	Ivarsfjorden	1962-1971	1951-2019	Idealized	1

A separate ERA5 dataset is downloaded from Copernicus for both of the sites, using the nearest grid cells to the corresponding location. For both sites, the corresponding dataset is then downscaled with TopoScale. The time resolution of the downloaded ERA5 forcing data is three hours. This has been clarified in the text.

Line 204-207: ERA5 output is provided as interpolated point values on a regular latitude-longitude grid at a resolution of 0.25° at an hourly frequency, both at the surface level, corresponding to Earth's surface as represented in the reanalysis, and at 37 pressure levels in the atmosphere from 1000 to 1 hPa. We considered data for the reanalysis period from 1951 to 2019 at three-hourly resolution.

Minor comments:

1 L155-164: Seepage face is at atmospheric pressure, but does that mean that the seepage face is located at Zwt? Could that sentence just be removed and eq. 1 explains all? Where are Kh values from? Eq. 1 mirrors Darcy's law but values seem made up, and dz is based on the grid z. Explain the rationale behind this model and why it was chosen. (alternatively, if Kh and dlat are unknown, these two parameters could have been replaced by one single parameter.)

We agree that the rationale and the description of the seepage face was not clear. The entire paragraph was reformulated for clarity

Line 173-185: We use a one-dimensional model setup, but simulate lateral drainage of water by introducing a seepage face, i.e. a lateral boundary conditions for water fluxes representing flow between saturated grid cells of the model domain and a stream channel (or the atmosphere) to which the water can freely flow out from the subsurface (e.g. Scudeler et al., 2017). Using the elevation of the water table, z_{wt} (computed as the elevation of the uppermost saturated grid cell), lateral water fluxes F_i^{lat} are derived for all saturated unfrozen grid cells i below the water table (i.e. at elevations $z_{\mathrm{i}} < z_{\mathrm{wt}}$) as

$$F_i^{lat} = -\mathrm{K}_H \frac{z_{\mathrm{wt}} - z_{\mathrm{i}}}{d^{lat}},$$

where K_H is the saturated hydraulic conductivity, d^{lat} is the lateral distance to the seepage face and the flux is determined by the difference between the hydrostatic potential (proportional to z_{wt}) of the water column and the gravitational potential of free water at the elevation of each cell (proportional to z_i). Note that Eq. (1) is an approximation for small changes of the water table and small outflow fluxes for which the potential in the saturated zone can be approximated by the hydrostatic potential. The parameter d^{lat} is used to control the strength of the drainage, with small distances resulting in a well-drained column, while high values lead to suppressed drainage.

The concerns of the reviewer regarding the K_H values are addressed in major comment 2.3

2 L282: But table 1 does not include any tested case with a porosity of 0.2!! Either the methods section does not completely describe what was done in the validation runs or there is a typo here?

As part of the improved methods chapter, we stated that the ground stratigraphy (porosity) is manually adjusted to find a good fit between measurements and simulated ground temperatures. The best fit was then used for model validation. We made the following changes to the text.

Line 282-283: Manual adjustment of the ground stratigraphy (porosity and thus mineral content) and snowfall factor are performed until a good fit with daily measurements is achieved.

3 L296-312: The model results suggest that the differences in model setups have very little impact on MAGST and that overall there seems to be a slight cold bias in the model. What then does this validation add to our understanding, if the purpose of the study is to investigate the impact of drainage on ground temperatures and ground ice?

The comparison to near-surface ground temperatures is clear evidence that the model can capture key processes, in particular the insulating effect of the snow cover, in a satisfactory way. We argue that is a prerequisite for being able to perform a meaningful sensitivity analysis for the effect of subsurface stratigraphies and drainage conditions. For the Ivarsfjorden site, the simulated ground surface temperatures indeed do not show a large spread between the different scenarios, but the good agreement with measurements shows that we capture the key processes relevant for the near-surface temperatures (i.e. snow cover, surface energy balance, etc.) Secondly, we show that (with the correct near-surface temperatures) it is possible to explain permafrost presence until today in the *blocks only, drained* scenario which fits e.g. to observations of cold air exiting the rock glacier (Lilleøren et al., 2022). Without the comparison to near-surface temperatures, one could simply select a low value for the snowfall factor which would likely also explain permafrost presence, but such simulations would not provide agreement for near-surface ground temperatures. Sect. 5.1 now contains an in-depth discussion on this issue:

At Ivarsfjorden, simulations with full snowfall yielded a similar performance for the ground surface temperature, approximately reproducing the mean of measurements at 11 sites. A statistical evaluation at both sites indicated a cold bias of the model of approximately -0.5 °C which we considered acceptable, considering the spatial variability of the ground thermal regime at both sites (see Gisnås et al. 2014 for Juvasshøe). At Ivarsfjorden, the transient simulations are in broad agreement with observations at the rock glacier which indicate that permafrost has been present in the recent past (Lilleøren et al., 2022). Permafrost conditions are simulated for all stratigraphies during model spin using the cold period 1962-1971 for which temperatures are closest to Little Ice Age conditions when the rock glacier was likely active.

Line 447-465: At Ivarsfjorden, simulations with full snowfall yielded a similar performance for the ground surface temperature, approximately reproducing the mean of measurements at 11 sites. A statistical evaluation at both sites indicated a cold bias of the model of approximately -0.5 °C which we considered acceptable, considering the spatial variability of the ground thermal regime at both sites (see Gisnås et al. 2014 for Juvasshøe). At Ivarsfjorden, the transient simulations are in broad agreement with observations at the rock glacier which indicate that permafrost has been present in the recent past (Lilleøren et al., 2022). Permafrost conditions are simulated for all stratigraphies during model spin using the cold period 1962-1971 for which temperatures are closest to Little Ice Age conditions when the rock glacier was likely active.

Within the model setup, in particular the exact ground stratigraphy and other poorly constrained parameters, such as the albedo, give rise to uncertainties. While the real porosity of the ground is unknown, sensitivity tests show a maximum of 0.4 °C differences in simulated ground temperatures between the highest and lowest porosity values tested (Supplement). Only at Juvvasshøe, the stratigraphy has been described from the borehole (Isaksen et al., 2003), while no thorough evaluation of the subsurface stratigraphy is available for Ivarsfjorden. Lilleøren et al. (2022) described the site as a complex creeping system with inhomogeneous subsurface properties. Most of the rock glacier surface is described as 'relict' (Lilleøren et al., 2022) with sand and gravel in between blocks. For these 'relict' areas, the simulations for the blocks with sediment stratigraphy, in which near-surface permafrost fully or partially degrades, could indeed represent the thermal state adequately. This is supported by the validation run with the blocks with sediment stratigraphy which yielded a good fit with ground surface temperature measurements at sites largely located on this 'relict' surface (Lilleøren et al., 2022). Two areas are described as 'fresh' which could indicate lateral movements due to the presence of ground ice. These contain larger blocks and could thus be better described by the blocks only stratigraphy for which permafrost and ground ice still persist at the end of the simulations.

Please also refer to Major comments 1 and 2.

4 Section 4.1: This section would greatly benefit from some objective measure of model fit instead of just some conclusions from visual inspection of plotted temperature curves.

We agree with the reviewer and included a statistical analysis based on the available temperature data at both of the sites (see major comment 1).

5 Section 4.2: It is not surprising that snow insulates the ground from winter cooling, generally leading to higher ground temperatures. More relevant than a comparison of snowfall factors, would be to check if observed snow depths and ground temperatures are recreated with the forcing data used here. When reading this section, I get curious about how much snow is generally observed at these two sites and does the forcing data capture the range of observed snow, or is a snowfall factor scaling needed to "correct" the forcing data

It is true that the insulating effect of snow on ground temperatures is well known. However, as the snow conditions show a high spatial variability in Norwegian mountains and measurements have shown that permafrost presence is often associated with a low snow cover, it is important to investigate the sensitivity of ground temperatures towards snowfall factors. Consequently, we consider the testing of snowfall factors to be relevant for this study. See as well the response to major comment 2.3.

The observed snow depth cannot be recreated directly with the forcing data, as wind redistribution plays an important role, especially at Juvvasshøe. Therefore, a snowfall factor was applied and fitted for model validation. This is described in the results of the original manuscript:

Line 321-324: The snowfall factor for this model setup is 0.25, i.e. incoming snowfall is reduced 75% in order to capture the effect of snow ablation due to wind drift. This resulted in mean annual maximum snow depths of 34 cm, in broad agreement with observations from the site (Iskasen et al., 2003) and earlier modeling studies at the site (Westermann et al., 2013).

We understand that a constant snowfall factor cannot represent the temporal evolution of the snow cover. This is discussed in section "5.1 Limitations of the model setup".

Line 492-498: In this study, we have performed a sensitivity study with respect to the amount of snow (by modifying the snowfall factor, Sect. 4.2), but simply scaling snowfall cannot represent the true time evolution of the snow cover due to wind redistribution (e.g. Liston & Sturm, 1998; Martin et al., 2019), possibly resulting in differences between observed and simulated temperatures. Nevertheless, it seems unlikely that the exact time dynamics of snow ablation and/or deposition events strongly affects the dependence of the thermal anomaly in the blocks only, drained scenario on overall winter snow depths. We therefore conclude that the significant negative thermal anomaly for the blocks only, drained scenario is likely robust in the light of the model uncertainty.