RC2:

This paper aims at developing a methodology to scale local, geophysics-derived estimates of ground ice content to a subcatchment scale. The study is part of a project that uses geophysical data to estimate ground-ice in an area of the Andes. This work is currently under review as the first part of this study.

Similar upscaling attempts have been shown to be successful, but were mostly applied to high-latitude environments, whereas this study is considering a high mountain array. Hence, the authors make use of geomorphological data and field observations.

As the authors demonstrate, estimating ground ice content of high altitude, headwater environments is important to assess groundwater resources further downstream, yet a quantitative estimation of this parameter is difficult. Here, the authors build on geophysical data, presented as Part I of this study, to estimate the ground ice content throughout a wider area. By using various input parameters to classify their sites, the authors are able to provide quantitative estimates of ground ice content. While the approach is very interesting, the classification, which forms an integral part of the study, seems poorly constrained, and mostly qualitative. The authors repeat much detail of the geophysical characterization (which is fair, given that this is the most important data set), there is very little detail on the actual classification. No maps are shown that show the other input parameters, such as slope angle, aspect, geomorphology, or the estimated soil parameters, including locations of soil probing, making it impossible to follow or understand how class parameters vary and how they were decided on. Similarly, it is not clear how the parameters that are critical for the ice content calculations (thickness, area, ice content) were upscaled, or determined, particularly for areas without geophysical data. It would be great to also see those as maps.

Thank you very much for pointing this out - we completely agree with this suggestion. We will add more maps showing the input parameters, i.e. slope and aspect, to make the classification process clearer. The geomorphology, or rather surface characteristics of Site D is already shown in figure 6 of the preprint. Except for the gelifluction lobes, the geomorphology of this site is very homogeneous, but for more complex study areas, the geomorphological classification would induce larger heterogeneities. We will further add the locations of soil probings and boreholes to the permafrost distribution maps.

The parameters that are critical for the ice content calculations (thickness, area and ice content) are determined using the geophysical data. Where no geophysical data is available, we relied on field observations. This will be better discussed in the revised Discussion section (by combining it with changes demanded by Reviewer 1 regarding a better discussion of the uncertainties shown in Figure 8)

These limitations of the current manuscript makes it difficult to understand what the benefit of the approach is. Comparing Figures 7 and 8, the shown difference between the geophysical based estimate and the empirical approach, could well fall within the uncertainties introduced by using different classifications. Given the strong reliance on field observations, it is also questionable whether similar approaches could be used more widely to estimate ground ice contents.

A similar comment was made by Reviewer 1 regarding the shown uncertainties in Figure 8. We agree that shown like that, the benefit of our approach does not become evident and we will try to improve our argumentation (i.e. that our field-based approach does *significantly improve the ice content quantification) in the revised version. It is also worth noting that field observations are an important aspect when characterising the ground ice content of a watershed (see discussion of Part I of this two-part contribution), i.e. we do not promote that the upscaling should be done using geophysical investigations only.*

As a direct reply to the reviewer comment above, we would first like to state that Figure 7 and Figure 8 do not show the same information. In Figure 7, we compare the water equivalents calculated using the upscaling methodology established in the paper per km² for both study sites. It shows that the calculated ground ice content of the site without rock glaciers (Site D) is larger than the ground ice content per km² of the rock glacier catchment. The uncertainties are only shown for Site D, because the classification for this catchment is much more ambiguous due to the lack of clearly definable landforms (such as rock glaciers, talus slopes, etc.). Figure 8, on the other hand, shows a comparison for the water equivalent values calculated for Site A with the approach from this study and the empirical approach presented by Brenning (2005). This is mainly to illustrate that the empirical approach, which is not based on any field data (geophysical data or any other investigations), most likely overestimates the water equivalent values stored in rock glaciers. Because of the vast geophysical data that we have from this study site on many of the rock glaciers, we are confident that our field-based estimates are more realistic. But we agree that the large (and rather theoretical) uncertainty bars do not express this well enough. We will revise the uncertainty discussion in the Discussion Section accordingly (see also our response to a similar comment by Reviewer 1).

Furthermore, we would like to highlight that a clear benefit of our approach is that it is not limited to rock glaciers alone but also includes other landforms or permafrost areas in general that may contain ice-rich permafrost, such as Site D. We show that the water equivalent stored in such areas may potentially be significant for further hydrological research. We will also highlight this better in our revised version.

Finally, we agree that we strongly rely on geophysical field observations, supported by in-situ ground truthing data, which may make the approach not directly transferable to any other study site, where no such data exist. However, we believe that field observations are essential to understand the ground ice content distribution of catchments exactly because they cannot be inferred from remote sensing data alone, not even for rock glaciers where at least the area can be delimited more or less clearly from satellite pictures. With this study, we want to emphasise this point and warn the scientific community to rely solely on large-scale remote sensing data when quantifying ground ice content in permafrost regions.

Next to those rather major comments, please find below some more minor comments:

Line 14-15: I don't think that an abstract should contain references, and I wonder whether the detail on the geophysics is actually needed here, as this paper focuses on the upscaling, not the geophysics.

We agree and will modify the abstract accordingly.

Line 89: In a previous sentence you mention that line locations were planned based on "safety within the mines". Does this mean that the chosen sites are active mining sites, and hence not in their natural state? That would make upscaling to natural systems impossible. According to Fig. 1, sub area 1 seems to be located within active mining, whereas others seem outside. I think some more detail is needed here on what the impact of mining on the chosen sites is to justify that mining has no impact on the results.

Site D is not an active mining site. Here, the impact of the mine is limited to the construction of roads (which facilitated access for us). Site A is an active mining site. However, for this study we only used the geophysical profile results from areas that are undisturbed and away from any disturbances, meaning that some of the profiles located at site A, Sub1, were not considered in this manuscript as they are located on mining waste rock material. We only considered profiles that are not impacted by the mining activities. We will add clarification on the impact of the mining activities on the upscaling process in the revised manuscript.

Line 126: “comparable near-surface substrate [...]” This is a critical assumption for the upscaling, yet the authors do not provide information on the geology and the variability of subsurface properties.

We agree that some general information on the geological setting would be helpful. We will include short geological descriptions of both study sites in the site description section. However, we did not use a detailed geological map as an input parameter and we propose to omit a geological map. However, we will add a reference to available geological maps so that interested readers may have a look at the geology.

Line 154: Potential incoming solar radiation: How and based on what did you calculate that?

Potential incoming solar radiation was calculated using ArcGis Pro’s “Area Solar Radiation (Spatial Analyst) tool, which uses a DEM as input. The tool also includes the latitude of the sites for calculations of solar declination and solar position in order to derive PISR. A corresponding sentence will be included in the revised manuscript.

Line 155: Equation for estimating permafrost occurrence: It would be good to show a figure that shows the data and model fit, and also details the parameters of the model.

We will look into this comment and decide whether to add a figure as a supplement.

Line 172-173: what do you mean by "high bedrock slopes"?

What we meant here was bedrock at the highest altitudes (> 5200 m a.s.l) of the study site. We will modify this sentence accordingly.

Line 186-187: On what data is this threshold based on?

This is mostly based on field observations and mapping, where in areas >30° (consulting a slope map based on a DEM), no sediment was observed. A corresponding sentence will be added in the revised manuscript.

Line 197-199: Although you describe the input parameters, there is no clear methodology described here on how you define the classes. This needs more detail and justification.

We agree that the justification of our methodology could be explained more clearly. Reviewer 1 also commented that additional explications to Figure 2 (upscaling approach) should be given. We will address these concerns by improving Figure 2, and by adding *new maps that show the spatial distribution of input parameters for the class definition (see also our response to one of the major comments) We will also further clarify in the revised manuscript how the classes are defined.*

Line 202: Given that soil properties will also impact on the ground temperature distributions, shouldn't the soil stratigraphy be an input to the classification?

The soil stratigraphy, as known from the geophysical surveys and the test pits, is used as qualitative input to the classification, because if two geophysical surveys showed the same stratigraphy this information is used to classify them into the same upscaling class. However, as soil stratigraphic information is usually very sparse (in our case the geophysical profiles, boreholes and test pits) it cannot be used as explicit input parameter continuously over larger areas.

Line 295 - 298: Given that the scope of your work is upscaling, why do you distinguish areas where geophysical data has been acquired and areas where this has not been done?

For rock glaciers located in the catchment for which we don't have any geophysical data, we upscale data from rock glaciers with similar altitude and aspect, where ERT and RST measurements are available for the ice content quantification. For rock glaciers where geophysical data exist, we can use the calculated ice contents directly. We will explain this better in the revised version.

Line 320 - 321: It would be great to see the estimated ground ice content as a map.

We agree and we will provide a figure showing the estimated ground ice content for each upscaling class of site D together with the soil stratigraphy conceptualizations for each class.

Discussion: The discussion on the geophysical results should be mentioned, but not in that much detail, as it should be part of Part 1 of this study. The uncertainty in the classification is of greater importance.

We will discuss the classification uncertainty in more detail. This was also a comment from reviewer 1. We will also go through the text again and remove parts where the geophysical results are described in too much detail.

Line 390: I don't think that this is necessarily an image classification problem. But you could use machine learning to exploit relationships between surface and subsurface parameters.

Thank you very much for this comment; that was also what we meant – we will rephrase the sentence accordingly

Figure 4: You prescribe ice-poor bedrock of D02 with an ice content of 4%, and bedrock of A15, which is overlain by ice-rich material with an ice content of 0%. How did you define that? Similar for Fig. 5, where the 4PM model indicates higher ice contents. Does the bedrock geology play a role in your definition of the ice content? If so, shouldn't this be an input to the classification?

The value for ice content in bedrock is based on field observations and the geophysical results. For site D, the bedrock is highly hydrothermally altered and fractured, which allows for larger pore spaces. We observed at several bedrock outcrops that the bedrock pore space was actually filled with ice. Therefore, we assume an ice content of 4 % for the bedrock at this site. Site A is located at much lower elevation and the geophysical results as well as the borehole data indicate that there is no ground ice present outside from rock glaciers. We conducted a geophysical profile located on bedrock (with a thin sediment veneer), which suggests no ground ice. Therefore, we assume an ice content for bedrock of 0%. So, the bedrock geology does play a role and we will explain this more clearly (and more generally) in the revised version.