Reply to RC1

We thank the referee for their review and constructive comments. Original review comments are shown in **black,** while our replies are provided in **green**.

Review "Scale-dependency in modeling nivo-glacial hydrological systems: the case of the Arolla basin, Switzerland" by Argentin et al.

This study presents a modelling exercise for the small Arolla basin in Switzerland (26 km2), with the aim of finding out how model parameters vary for nested and neighboring catchments and if parameters are thus spatially transferable for a semi-distributed/semi-lumped hydrological model. The calibration of the hydrological model is done for three different melt models, two different objective functions (NSE and KGE), and with or without considering debris cover on the glaciers. The many results show that parameters are transferable, and that similar parameters can thus be used to simulate streamflow along a river network (up- and downstream). For the temperature index melt model of Hock, that includes potential solar radiation, parameters are more similar across the basins when compared to the other two melt models that use less spatial information.

Thanks for the nice summary of our work.

Overall, I found this a very interesting study with many different aspects that were looked at, a well written methodology section and a good presentation of the results and the discussion. However, I have a few points that require some further attention, which mainly relate to the framing of the study and clarity of the results.

1. Introduction – a) when reading the introduction the first time, I was a bit confused by the reasoning of why parameter transferability may be difficult in glacierized catchments and the later following explanations of the different melt models. At first, I thought that the study would focus on the different storage and routing parameters, but it turned out that the study focuses rather on the melt modelling (catchment-wide melt contribution). Although this becomes clear towards the end of the introduction, I think it should be mentioned earlier on, with an explanation of why this focus was chosen. b) What is also missing in the introduction is a discussion of studies that use glacier mass balance or snow related data to calibrate models. In such cases, melt parameters are calibrated independent of catchment size and the problems outlined here are not (less) applicable, especially if these measurements are available at large scale (remote sensing data). This is an important consideration for the framing of this study. c) And last, I think that the comparison between calibrating the model with NSE or KGE as objective function should be already mentioned as background material in the introduction. It would be good if it becomes clear how that fits too into the story of modelling glacierized catchments.

a) We agree with the reviewer and will change the introduction to clarify that we focus on the melt contribution of the different areas modeled and why. We chose this focus because storage in high alpine glacierized catchments is often shallow and plays a more minor role compared to the melt contribution.

b) In the introduction, we will add a discussion of studies using glacial mass balance and/or snow data to calibrate models. We will discuss how glacier mass balance and snow data can indeed help constrain the calibration of the hydrological model, yet it remains necessary to study the transferability of parameters between catchments to predict discharge at ungauged locations situated within the catchment.

c) We will introduce the comparison between NSE and KGE earlier in the article to clarify that our goal is primarily to assess the sensitivity to the objective function and if our conclusions still hold with another commonly used objective function.

1. Although section 4.4 is very interesting, I found the formulation and presentation a bit weak at the moment. It would help if there was already a discussion on NSE and KGE and its different sensitivities in the introduction. Furthermore, the section mixes discussion with results, although a discussion section is following. And third, I do not understand why the neighboring catchments were not included too in the analyses of Figure 12?

1) This comment ties in with the previous one, and we will discuss the sensitivities of the NSE and KGE (based on existing knowledge) in the introduction. 2) Indeed, we will move the second paragraph, which is more of a discussion paragraph, to the Discussion section. 3) We will include the neighboring catchments in the analyses of Figure 12.

2. Section 4.6 is supposed to give some insights into the physical meaningfulness of the transferred parameters, but does not discuss any of them. After reading this section I was not sure what message to take from it, besides that catchments with higher glacier cover show higher discharges later in the season. However, this is not related to the different weather patterns suggested in the lines before. Maybe this section could be tied in with the section where figure 12 is discussed?

We agree and will tie this section to the section 4.4.

3. Overparameterization – one of the conclusions of the study is that including debris cover in the calculations likely leads to overparameterization and therefore a less good fit when transferring parameters from one catchment to the other. While this may well be the case, I found the reasoning not always strong, as for example some of the melt models use more parameters (ATI) than including debris cover in the TI or HTI model. Without testing the removal of more parameters (even simpler models) and seeing when the model reaches some state of overparameterization, these conclusions are quite speculative. Maybe the authors could better argue why they choose that overparameterization might be occurring, rather than the debris melt being too spatially variable due to variable thickness?

We agree and we will modify the text to open to additional explanations on why debris cover did not improve the simulations.

4. In the last part of the discussion (5.3) I found the discussion on catchment sizes and transferability of parameters hard to follow (L412-L426). It seems to start with explaining that there is no converging of parameters for all melt models in small catchments, and that there is no improvement for the HTI model compared to the TI model, to then conclude that the HTI model parameters for catchments smaller than 7 km2 are good transferable. Please check the reasoning here and clarify. More on a general note, I was missing some discussion on the implications of the results presented in this study. For example, do the results hold for this specific set-up, or for this catchment size (26 km2 is still rather small), or can we expect similar transferability effects in larger catchments? And would the results also apply to parameters other than melt related parameters?

We will clarify this part of the discussion and will add a discussion paragraph on the implications of this study on transferability applied to larger catchments and other, non-melt related parameters.

Detailed comments

Thank you for the detailed comments, which we will consider during the manuscript's revision. Below we answer those comments that go beyond simple corrections:

- L8 “streamflow patterns” – what is meant here with patterns?
  We meant it in the sense “streamflow regime”, but we will remove the word “patterns” to make the sentence clearer.
- L29 – L34 – maybe it would help to also add that the category of “lumped” models here, do not include a “real” spatial location, and that is the reason why they are calibrated at the outlet (as the functioning of the whole system).
  In my opinion, since most “lumped” models still take into account the characteristics of the area they are trying to model (glaciated area, debris cover area) they are still anchored in a “real” spatial location. We will however highlight that they reflect the functioning of the whole system, as requested.
- L62 what is a “parameter bias”?
  We mean it in the sense of “parameter error”. We will clarify.
- L66, possibly use another symbol for degree day factor, as “a” is often also used for albedo.
  We do not think this is a problem in this paper.
- L74 what is meant with “correlation of aspects?”
  This refers to the spatial decorrelation, i.e. the distance beyond which the spatial variance does not increase anymore. We will clarify this point.
- L133-135 possibly add that the discharge data were normalized with the same maximum values? Otherwise biases between simulations and observations will not be visible.
  Yes, we will specify this “normalized by the same highest observed discharge values.”
- L161 “along with baseflow from melt and rainfall” – what is meant here? The outflow from the groundwater/slow reservoir?
  We mean the baseflow originating from snowmelt or rainfall on ice-free ground and ice melt. Contrary to the baseflow originating from snowmelt, icemelt and rainfall on glacier-covered ground. We will add a reference to the Figure A1 that explains this better.
- L163 – “fed into (to – remove) two lumped parallel reservoirs” – can you briefly describe what these reservoirs represent?
  Yes, thank you, we will add that.
- L193 “night time” – at what temporal scale is Ipot calculated?
  Ipot is computed at the 15 min scale, we will add this to the manuscript.
- L202 “downscaling” – how was this done?
  We will reformulate the sentence. The downscaling is done by computing the weights representing the contribution of each data cell to each HRU based on their spatial coverage. These weights are then used to derive the mean values for each HRU, for each daily time step.
- L211 “but an inequality is added” – can this be described in a more elaborative way? How much is it, and to which parameters is it applied? Are estimates of debris thickness known to assure it has a lower melting effect instead of an increasing effect?
  We will describe this part in more detail.

- L250 “more robust yearly signatures” – this sounds like stable flow, while what is meant here is that it is more variable flow, right?
  You are right, this is wrong wording, we will reformulate.
- In Section 4.5, what is the difference between reason i and reason iii?
  We mean in i) the use of a parameter that overfits the model and might not be transferable to new data, while in iii) we refer to the impossibility to find a unique set of parameter explaining the results. These reasons are tied, and we will try to clarify this part.
- L371-372 “Longer in-stream flow paths lead hereby to a stronger dampening effect of hillslope- and glacier scale runoff variability” – Could the mechanisms behind this be explained here? Does it relate to more sub-surface storage and other streamflow components to compensate the glacier runoff variability?
  Thanks for this comment, we were referring here simply to the smoothing effect of streamflow waves traveling downstream (geomorphological dispersion; Rinaldo et al., 1991).

Figures:

Thank you for the detailed comments on the Figures, which we will consider. Below, we answer those comments that require an answer:

- Figure 5 – what does the distribution of parameters represent here? Are these all parameters of the 10000 simulations?
  Yes, these are the all the parameters from all the simulations.
- Figure 12 the “relative difference” units are not very clear. Maybe “relative difference” can be explained in the caption, as well as “relative performance change”, is it a fraction of the original one (i.e. calibration run /BI characteristic?). And how was the over/underestimation assessed?
  We agree that the unit is not clear, we will thus rewrite the explanation as a formula as suggested. The over/underestimation was assessed visually, we will specify it.
- L446 “when focusing on attenuation of discharge trend offsets” – what is referred to here? This comes new in the conclusion and hasn’t been explained before, at least not using these wordings.

  True, we will change the wording to match the places where we mentioned this.

References:

Rinaldo, A., Marani, M., and Rigon, R.: Geomorphological dispersion, Water Resources Research, 27, 513-525, 1991.