Review of manuscript by Huo and Bishop submitted to *Earth Surface Dynamics*

In this manuscript, Huo and Bishop have assembled a model to investigate the role of supraglacial ponds to glacier mass balance on Baltoro Glacier. They also use to model to investigate the links between ponds and debris gravitational remobilization, and the control of longitudinal gradient to supraglacial ponding. Simply assembling the model is an ambitious challenge, as the authors seeks to integrate representation of multiple complex processes that have relatively few observations. The authors rightly point out that this is an important and novel line of research, since supraglacial ponds are known to provide localized melt enhancement that is non-negligible at the glacier scale, yet studied in few instances, and their drainage can have consequences outside the glacier domain.

At present, the assembled model necessarily makes assumptions about pond water mixing, albedo, and (lack of) drainage. Unfortunately, I feel that the model choices at present invalidate the model's utility for the aims, which require additional process representation, testing, and evaluation using field or remote sensing observations (presently lacking entirely). For the second research question, the authors have a chance to make some novel conceptual contributions, but this would require extensive model testing (for parameter/resolution sensitivity, as well as distinct process assumptions, for example). In short, although the authors' effort at combining distinct process dynamics to understand supraglacial debris evolution is extremely worthwhile, neither the model nor the experimental design is presently adequate to address their research objectives.

Major comments
Scape/research questions. The authors have taken on a very complex and ambitious modelling task, with many poorly constrained processes. Realizing that, the key question is whether the process representation is appropriate to the research questions/aims. I am very impressed with the effort to assemble this model; nonetheless I strongly disagree with some of the modelling choices, and I do not think the manuscript presents any examination of plausible alternatives. As I see it, there are three distinct aims, but the model is not quite suited to answer any of them. Consequently, I would greatly recommend that the authors focus their revisions around research question 2 (relationship of ponds to debris mobility). See the details in the line-by-line comments, but:

- Effect of ponds for ablation on Baltoro. For this I would expect to see some validation or observational constraint of at least multitemporal modelled pond extent, and possibly water surface temperatures or albedo. I would have also expected to see some evaluation of the suitability of the DEM for this purpose, consideration of englacial drainage (which greatly changes the surface energy balance!) and sensitivity analyses related to, for example, albedo estimation, lapse rates, and modelling assumptions.

- Understand how gravitational redistribution controls supraglacial ponding. This objective is within reach for the study. At the moment, the study evaluates how a constructed parameter changes the formation of topographic depressions (and potential water accumulation in said depressions). However, there are alternative plausible formulations for Beta that would be worth investigating – what if Beta acts as a coefficient for static friction as well as for kinetic (thresholding as well as resistance to motion)? What if Beta is related to moisture availability? Also, much more sophisticated considerations are needed to make the links to, e.g. ice cliffs: are your results scale-independent? What about the sensitivity to hummock size? Can your model reproduce the conceptual changes or e.g. Bartlett, Rowan, and Ng (2020)? Also, how do ponds affect debris mobility (this is not a one-direction relationship)?

- Effect of longitudinal gradient. The fact that the model does not even attempt to represent englacial drainage greatly hampers the study’s ability to address the question. Water retention in depressions is only half of the equation, as pond drainage could also be related to longitudinal gradient for a variety of reasons. This is quite a simplistic experiment as a result.

Key problems with model process representation

- The lack of englacial drainage is a major problem (and challenge). Ponds rarely drain overland; most instead seem to be continuously (inefficiently) englacially connected, so their water level reflects a dynamic imbalance between inputs and outputs (e.g. Sakai et al, 2000; Miles et al, 2017, 2018 TC; Narama et al 2017). Maybe you could characterize the outlet mean discharge of each topographic sink using satellite imagery? Or can you use remote sensing observations to constrain pond extents at least? There are available ASTER and Landsat data that could be used for this, if they are not used independently for model validation (even better). However, I do not think it is appropriate for your ponds to continuously accumulate water over the season; this grossly overestimates the pond coverage and depth.

- Supraglacial ponds collect thick debris within their depression (e.g. Mertes et al, 2016), in fact for the reasons that your study aims to highlight, but this greatly reduces the pond-bottom melt rates (e.g. Miles et al, 2016, 2017 Frontiers).

- Supraglacial ponds are not well-mixed, but show diurnal temperature variations that are most pronounced at the water surface (e.g. Rohl et al, 2006,2008; Miles et al, 2016; Watson et al, 2017), where the temperature is also generally highest. Overturning often happens on a daily basis, but melt rates are generally highest at the waterline, where forced convection is also highest (also e.g. adjacent to exposed
ice...). Notably, the water temperature is closely related to the mass inputs and discharge (e.g. Watson et al, 2017; Miles et al, 2016)

- Supraglacial ponds on debris covered glaciers are generally turbid (e.g. Wessels et al, 2002; Kraaijenbrink et al, 2016; Watson et al, 2017). Please consider whether your albedo model is appropriate and justify it if you decide so.

- Appropriateness of input data. This depends entirely on your research question, but for the first objective (a realistic scenario for Baltoro):
  - Debris thickness. Note that the debris thickness data of Mihalcea et al (2008) contains ‘thin’ debris areas that are certainly ponds (ignored in that study) which exhibit colder surface temperatures than the surrounding debris. Ideally you would correct these debris thicknesses to that of the nearby terrain. Also, what about uncertainty?
  - A single ASTER pair is notoriously noisy and uncertain (15-20m! E.g. Willis et al, 2012). How much does this affect your results? It also would indicate the water surface elevation for pre-existing ponds; can/should you correct this? Is the DEM resolution adequate for the purpose of modelling overland drainage?
  - I am surprised that the authors do not use an initial pond coverage from the coincident ASTER scene to at least consider how well the modelled pond coverage fit observations, especially as they have produced such a coverage in their other publications.
  - I note that other recent studies by the authors (Huo et al, 2020, 2021a, 2021c) have considerable overlap with the present manuscript, including in terms of equations, process dynamics, and literature review. I think it would be very appropriate for the authors to clearly state in the introduction how this study builds on their own past studies (and, crucially, differs from them!).
  - As this paper examines quite complex dynamics, it is not surprising that there are several uncited studies that should also be examined to help design hypotheses or to choose the appropriate parameterizations. Please see mentions below for help in model formulation or experimental design.

Line-by-line comments

L23. This is true except for the case of a base-level water body, which can still be a ‘pond’ in initial stages; e.g. Benn et al, 2001; Thompson et al, 2012, etc.

L27. Comma after destructive should be a semi-colon. However, Reynolds (2000) and Richardson and Reynolds (2000) don’t look at supraglacial lake drainages so much. See Miles et al (2018, TC) for an example of a GLOF originating from a supraglacial pond. There is evidence of other outburst floods likely to involve supraglacial and englacial water storage (e.g. Rounce et al, 2017). See also Narama et al (2017) and Sakurai et al (2021) for some examples in the Tien Shan.

L36. Possibly see also Salerno et al, 2017.
We do have an estimate of the mass loss attributable to ponds in the Langtang Valley (e.g. Miles et al, 2018; GRL), but I agree that this is a single estimate for a single site.

This underestimation is calculated for both cliffs and ponds in Buri et al, (2021) but again only for one location.

It is not clear what is meant by debris flux in this context – the emergence of debris out of the glacier? The transport of debris down-glacier? The mobilization of debris across the surface? All of these factors could have a control on supraglacial pond incidence and dynamics, so please be clear which you are referring to (and why).

See King et al (2020) here.

Rather than slope, I would recommend longitudinal gradient in this context, as the surface slope is (locally) often very high.

Just a minor correction - the Rounce et al. (2018) study removed ponds and cliffs because their melt enhancement led to a lower effective debris thickness from their model inversion (because there was more melt). This does not mean that ponded areas have thinner debris, but that the model is not suited to this domain (as it was not designed to be). Similarly, the debris thickness map of Mihalcea et al (2008), which you use, shows thin debris in areas of supraglacial ponds, because the water temperature is much cooler than the debris (but above freezing). Perhaps more relevant here is Benn et al (2001) who identified the effect of ponds in removing marginal debris.

I would make the distinction that here you refer to numerical models, although I believe that your study can also better inform our collective conceptual model.

Comma here should be a period (full stop).

Individual ASTER DEMs are notoriously noisy, with uncertainties of up to 15-20m over stable terrain (e.g. Willis et al, 2012). Furthermore, they of course to not resolve the terrain where ponds are already present. How problematic do you think these aspects are for your results?

By ‘trend’ here (diurnal trend and seasonal trend) what exactly do you mean? Do you mean the diurnal and seasonal cycle?
L85. More recent studies have shown that a standard dry adiabatic lapse rate may not be the most suited to debris-covered glaciers (e.g. Steiner et al, 2016; Shaw et al, 2016). I would recommend considering including this parameter in your uncertainty analysis.

L89 and L142 (perhaps others). Rather than ‘radiation-driven’ I would recommend calling this an energy-balance model. ‘Radiation-driven’ sounds to me an Oerlemans or Pellicciotti-type melt model.

L95. Do I understand that you consider a pond to be present only if the water level is above the debris level by more than 0.5 m? Why not strictly when above the debris level?

L105. I don’t think the Taylor and Feltham (2004) albedo approach is particularly transferable to this situation; ponds on debris-covered glaciers have non-negligible suspended sediment, which actually serves in this case to raise the albedo, but also reduces the relationship of depth and albedo (because optical penetration depths are much lower). Considering the relationship plotted in Figure 3c, I don’t think this will particularly affect your results, but I would recommend exploring alternative assumptions.

L131-134. This assumption of turbulent, well-mixed water temperatures is rather severely flawed and may undermine the results of the study. Measurements of water column temperature (Rohl et al, 2006; Xin et al, 2012; Miles et al, 2016; Watson et al, 2017) and water-surface notch development (Rohl et al, 2008), which clearly show that the water surface temperature responds acutely to air temperatures (e.g. Miles et al, 2017; Frontiers), leading to enhanced waterline notch development especially when pond fetches are high enough (Sakai and Fujita, 2010). Furthermore, the subaqueous ablation (L136) seems to entirely ignore the thick debris layers often collected in ponded depressions (e.g. Mertes et al, 2016), which greatly suppress pond-bottom melt rates. Lastly, how do you close the energy (and mass) balance of the pond if you do not account for discharge?

L131. It’s not explicitly clear, but I assume that the surface temperature of the pond (= the mean temperature of the pond) is determined by solving Eq6 for deltaQ and then Eq7 for deltaTw?

L143. As explained above, I rather think that the lateral ablation is the primary pond melt and expansion process that should be accounted for. Perhaps calving can be neglected, but ponds tend to deepen initially, collect sufficient debris to suppress subaqueous subdebris ablation, then expand laterally.

L177. First definition of ‘debris flux’ as pertaining to the surface redistribution of debris. I think it’s great to test this implementation! One question is whether you expect Beta to be constant for a given glacier, or would it instead most likely vary with debris thickness/grain size/etc. To some degree Beta relates to the coefficients of friction; I’d
recommend making this link explicit?

Figure 2. Nice depiction. I’d suggest to add some annotations – ‘pore water enhances debris mobility’ or similar. I guess the thickness is constant in both panels (optical illusion that it is not?).

L181. What do you do about these supraglacial channels? Or does all water collect in any depression it runs into?

L200. This is the major obstacle for the model, as most meltwater ends up flowing into these supraglacial streams and draining englacially or subglacially (e.g. Benn et al, 2017; K Miles et al, 2019; Fyffe et al, 2019). Assuming that all meltwater flows over the surface or is collected is clearly not defensible given the observed situation of ponds typically deep within depressions and the numerous depressions without any ponded water (e.g. Miles et al, 2017; Frontiers). This greatly limits the study to testing the dynamics and interaction of processes (i.e. S1-S2 results are not meaningful).

L211. “…supraglacial ponds’ evolution” - ponds should be plural possessive.

L209-213. This is an appropriate and excellent set of tests for the model; I would recommend making this investigation the principal aim of the study, and expanding it with alternative model assumptions/formulations given the criticisms above.

L215-218. I think this is a great aim, but without accounting for the drainage aspect to the pond mass balance, you cannot preclude that drainage compensates for increase water accumulation with low longitudinal gradients.

Figure 3. It is worth noting that there are cloud-free Landsat 7 acquisitions over Baltoro in June, August, and September 2004, as well as the ASTER acquisition from August. In other words, you have the opportunity to evaluate how well your model reproduces ponding. Also, I do not find any description of how the debris albedo has been prescribed, whether a fixed value (in which case, what) or spatially distributed somehow, etc.

L241. Which simulation results are you comparing to the Gibson et al (2017) results? Your initial or final ponded water map (Figure 3a/b)?

L246. I recommend that you then focus on ‘only’ surface processes – how the surface types interact and influence one another rather than the glacier-scale results. Even in this
case, though, you may need to examine the effect of pond drainage on surface
morphology evolution – if a pond drains halfway through the ablation season, what does
the topography look like at the end of the season? Is there any sign that a pond was
present, or …?

L257. It is interesting that your results show enhanced melt also for the pond areas in the
Mihalcea et al (2008) debris thickness map (which often present as thin debris in the
lower debris-covered area). In this respect, the ‘real’ melt enhancement is considerably
higher for this domain, because your base value (the debris-only) actually includes
considerable melt enhancement attributable to ponds. Ideally, you would use the ASTER
data to map and remove ponds from the debris thickness map, then fill these gaps via
interpolation or similar.

L258. Again, the ponds do not respond solely to the radiative forcing, but the entire
meteorology – wind speed etc etc.

L266. This modelled nonlinearity is quite interesting and I assume it relates to the mean
water depth or temperature?

L250-275. Although I appreciate the effort to provide a real-world model implementation,
I don’t think the model’s formulation (e.g. some process representations, but especially
the lack of drainage or constraint by observed pond extent) or inputs (DEM, debris
thickness data) are robust to make glacier-scale interpretations; the discussion also does
not take these limitations into account.

Figure 5. This is a very cool experiment. Stylistically, I recommend reorienting panel C
such that the principal gradient is also directed down and left (as for A and B). Did you
test the effect of different amplitudes of imposed undulations? Ie. Is there a reason you
chose ~40m? It is also very interesting how low Beta is in all three simulations – what
happens when Beta is 0.5 or 1? I guess the redistribution is dramatically unrealistic as this
would imply very little friction, but this is also worthwhile to depict. What is the duration
for this simulation? Is there (also) a way to depict the mean melt rates in each scenario?
Where on Baltoro does the subset of 5C originate (perhaps depict somewhere in Fig3)?
Please also indicate that the black line indicates ice cliffs (idealized or interpreted or
observed as appropriate).

L300. I agree entirely that there could be feedbacks here, but I rather disagree with the
ponds encouraging further topographic lowering. Rather, they may do so initially, but as
they also collect debris, their enhanced lowering is short-lived; subsequently they expand
laterally.

L303-306. Not that repeated filling and drainage has been observed from the same ponds
even within a single year (Narama et al, 2017; Miles et al, 2017, Frontiers). However, I think you ought to test the effect of lateral pond expansion vs deepening.

L311. I don’t understand if these are ice cliffs (Figure 5C) known to exist on Baltoro in this position, or instead interpreted ice-cliff-like locations. In all cases the debris thickness is not realistic for an ice cliff (mm debris rather than 10cm). In part this is due to the resolution of the topography – 30m is exceedingly coarse to capture these dynamics (see e.g. Buri et al, 2016, JGR). Note also the effects of the ice cliffs themselves in reorganizing debris, which may happen within your grid cells or across them (e.g. Westoby et al, 2020). In short, I would recommend formulating a separate analysis to look at this specifically by investigating the cliff-pond interaction. Note that your Beta relates to friction, and therefore the moisture at the base of the debris, which has some strong implications for debris stability (e.g. Moore, 2017). Consequently, you might need to change Beta based on water content within the debris (another set of experiments that would be very interesting!).

L316. ‘overall glacier longitudinal gradient’

L318. See also Miles et al, 2017 (J Glaciology)

L315-333. I don’t particularly see the added value of this analysis. Firstly, glacier surfaces with 2 degree or 10 degree gradients may also have very different dynamics due to driving stress differences, which could lead to differences in strain and therefore internal drainage pathways. So your study has again only addressed the ability of differently-sloped surfaces to (possibly) retain water, not the (actual) retention of water. This is a fairly self-evident finding, but I am also surprised by how little of a different 10 degrees makes! What happens if you use the local debris thickness as a drainage criteria (i.e. ponds do not fill the real surface topography but the subdebris topography (Miles et al, 2017, Frontiers)? Considering that the DEM can only resolve the water level surface (not the depth of a depression filled with water at the time of acquisition) how much do you think the difference between the 2 degree and 10 degree water storage would change accounting for actual depressions?

L347. I agree that water can percolate through debris, but when it is within a pond, this would have to occur via convection. Usually, the energy exchange is weak between the pond base and surface, so pond-bottom water temperatures are very low, and there is fairly little subaqueous subdebris melt. Instead, the melt is focused at the pond margins, where debris is thinner or non-existent (e.g. Sakai et al, 2000, Benn et al, 2001, Miles et al, 2016, etc).

L386. I disagree with this entire paragraph.
This first feedback results from the assumed subaqueous bare ice and constant pond water temperature, along with the overland drainage requirement. Thus, in your model depressions grow deeper, but never drain, and the deepening process increases their bare ice surface area. None of these dynamics has been documented in reality. Rather, we see that ponds tend to accumulate debris, beneath which little melt is expected, have focused melt enhancement at their periphery, and generally drain englacially well below the apparent topographic constraint. This reduces the positive feedback tremendously – large ponds have a relatively small ice-contact area compared to their surface area (they become thin disks with bare ice at their perimeter), but warm considerably until they drain.

The second feedback is not addressed by the model at all, because it does not account for drainage whatsoever (the mechanism by which the englacial conduits are maintained). The authors seem to have misunderstood the results of Rounce et al (2018) regarding why the modelled debris thicknesses are lower in the vicinity of supraglacial ponds and ice cliffs – this is not due to a physical process but an inadequacy of the model. This content also seems to misunderstand the debris-emitted longwave concept, which is due to the strong longwave radiated from surrounding terrain to ice cliffs (the cliffs receive more longwave radiation, they are not warm at all, but essentially at the freezing point).

L400-415. It’s good to see this list. Do you think that any of these considerations undermine your results? I’m sad to see that no sensitivity experiments were undertaken to determine the appropriate simulations for your S1 setup. More crucially, where is the validation or observational constraint? How do we trust the glacier-scale results without either of those?

References


