
Review 3

RC3: '[Comment on egusphere-2024-3072](#)', Anonymous Referee #3, 05 Mar 2025

This study describes the implementation of several modifications to an existing Large Hydrological model and evaluates the effects of implementing those changes stepwise. Specifically, it evaluates the implementation of a snow transport scheme, an altered snow model parametrization, improvements related to calibration of that snow model, a new glacier model, and altered soil parameters. The authors also evaluate model differences related to meteorological forcing uncertainty.

General comments:

Overall this is a well-organized paper that clearly tracks the series of implemented changes. The authors justify their choices of evaluation data and techniques, and provide sufficient evidence for most conclusions. However given the complexity of the study, there are a few places where further clarification, justification, or tempered conclusions are needed.

Response: Thank you very much for the time and effort you put into reviewing our manuscript and your detailed comments and feedback. We have responded to your questions, suggestions and concerns below.

The only substantial piece of additional justification pertains to Figure 7 and your conclusions of performance in non-regulated versus regulated catchments. While the correlation of performance with water gap sign is quite clear for the second and third row, I'm not totally convinced how well the value of the water gap fraction works in identifying natural vs regulated catchments (to my eye both improvements and deterioration of performance are pretty evenly split between those with and without reservoirs marked as + and o). This is important because you later equate locations with performance improvements to natural catchments and locations with deterioration to regulated catchments (Table 3 and around line 565 and 606). Can you provide any additional justification for this association over your domain? Since the Salwey study was over Great Britain it may not transfer well to more inland mountainous locations. How do you know that another variable isn't controlling the relationship between WB and performance improvement? For example, maybe improvements in the model occur at locations that correspond to thin soil in the real world and where the hydrological response is flashier (since you effectively biased your model to better performance at such locations). Such locations might have limited capacity to store water longer term which would correlate with WB and could appear as a signal in non-regulated catchments.

Response: Thank you for making us aware that additional justification is needed. Based on this comment and comments by the other reviewers, we have implemented several changes.

1. The WB signature is indeed a complicated metric and is influenced by many other processes. We have now rewritten the text to better indicate that we do not equate the WB signature to reservoir regulation and to better highlight that reservoirs are just one process that can affect this signature. We also explicitly derive conclusions for reservoirs only from a combination of the WB metric and data on reservoirs. We have made several changes to the text:

“One way to separate regulated from natural catchments is the water balance signature (WB),

which describes the deviation from a closed water balance assuming no long-term storage effects (Salwey et al., 2023)."

"Positive values of WB indicate that the catchment discharge is higher than expected. Assuming that the meteorological components could be well-estimated, such positive values could suggest that a catchment gains more water than what comes in through precipitation. Negative values of WB indicate that a catchment loses more water than just the potential evapotranspiration."

*"However, other factors such as errors in the meteorological forcing or additional water input from glaciers **due to imbalance** can also lead to strong water balance deviations, which can ~~also~~ affect model performance with respect to discharge. **We thus use WB as a general metric to study the effect of such deviations in the water balance on model performance. In addition, we also use WB as an indication for water transfers, hydropower production and other water balance deviations in combination with catchment-based information on reservoirs.**"*

2. We strengthened the evidence for a connection between WB and the reservoirs. We added a new Figure i to the Supporting Information that shows that WB is related to a high degree of reservoir influence (which we here define as the total reservoir volume per catchment divided by the annual average discharge). We also redefine when we consider a catchment to be regulated, namely that we define this based on this degree of regulation. We only consider catchments where the degree of regulation exceeds 0.1 days to be regulated, which removes catchments with reservoirs but hardly any regulation.

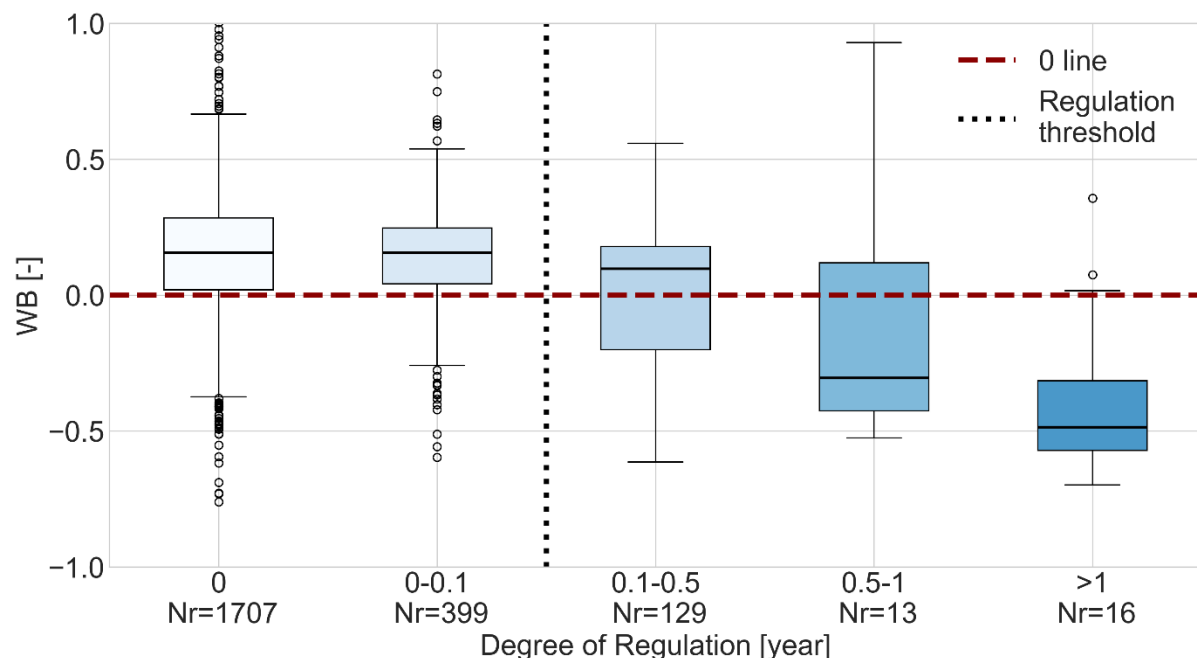


Figure i Relationship between the degree of reservoir regulation and the WB signature. The degree of reservoir regulation is here defined as the total reservoir volume over the catchment divided by the mean annual discharge.

Other Specific comments (generally minor):

Line 37: “regional to local-scale” can be interpreted differently. Please be more explicit.

Response: Thank you for pointing this out. We clarified this in the text.

regional to local-scale information:

*“... limits the usefulness of their output for policymakers, who are often interested in **more detailed information for their specific domain.**”*

Figure 1: The range of colors on the map doesn't look like it fully matches those on the color bar (which doesn't seem to have the darker greens and bluer greens). Is there some sort of transparent overlay of other colors on the map?

Response: Thank you for pointing this out. That is correct, since we wanted to highlight the most important river basins on the map. We have updated the figure to show the topography without the shading and include subplots that show the subbasins, as well as the countries and discharge stations.

Line 160: Since the precipitation isn't downscaled to the resolution of your model, it might be clearer to specify “(3) CERRA-CHELSA, a mixed dataset with temperature from the Copernicus European Regional ReAnalysis (CERRA) further downscaled using the CHELSA algorithm and precipitation data directly from CERRA-Land.”

Response: Thank you for your suggestion. We have implemented the suggested specification in the text.

Line 180-221: This section needs some clean-up to help clarify details. You mix in both data required for evaluation and ancillary data required to run the model (e.g. RGI) or produce metrics (e.g. snowfall fraction, PET). It's hard to sort out what is what, especially at a first read. In some cases, datasets mentioned don't appear in the Table (e.g. RGI, GLAMOS) and in others they appear in the table but aren't discussed here (Farinotti glacier volumes). You might try splitting up the discussion of strictly evaluation data versus ancillary data needed to run the model or compute metrics.

Response: Thank you for making us aware that this section needs additional structuring and a check for consistency. We follow your suggestion and make a clearer split between input, evaluation and ancillary data.

We restructured the text and made this split more apparent. The section is now called “Datasets”, with subsections “Model input”, “Evaluation and calibration data” and “Ancillary data”. We reordered the information accordingly.

Furthermore, we added the RGI and GLAMOS to the Table and added a description of the Farinotti dataset to the text in this section as well:

“As we implement a new glacier module, we had to define the locations of the glaciers and their initial thickness. This information was derived from the consensus estimates from Farinotti et al., 2019, which are representative for the year 2003 for most glaciers.”

Lines 237-251: What values were chosen for m_m and m_p ? Do these not alter the calibration? - at line 320 you state that you only calibrate the two DFF values.

Response: Thank you for pointing out the need for further clarification. We further specify that these two parameters were taken from the study by Magnusson et al., 2014. They derived the m_p value from measurements and they fixed the m_m value to reduce the number of

parameters for calibration.

These parameters can be found together with the other parameters in the Supporting Information in Table S1.

In section Model development:

*“The parameters used in the equations **and any fixed values** are listed in Table S1 in the **Supporting Information.**”*

In section Snow module:

*“... where m_m is a parameter controlling the transition between melt and no melt ($^{\circ}\text{C}$) **and was kept constant by Magnusson et al., 2014.**”*

*“...and the parameter m_p determines the range where snow and rainfall co-occur ($^{\circ}\text{C}$) **and was derived from snowfall observations by Magnusson et al., 2014.**”*

In section Calibration:

*“In contrast, model evaluation is performed using the full model run. The considered calibration period is 2000–2009 (see Section 2.4). **Note that any other parameters not mentioned here remain fixed and their values can be found in Table S1 in the Supporting Information.** The updated snow module required the calibration of 2 parameters, namely DDFmax and DDFmin.”*

Magnusson, J., Gustafsson, D., Hüsler, F., & Jonas, T. (2014). Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water resources research*, 50(10), 7816–7835.

Lines 255–270: Please reword to clarify the interactions between snow and glaciers regarding both lateral transport, accumulation from snowfall, and melt. Including an arrow for lateral transport in Figure 3 may help. In particular I think it would help to more explicitly describe how the model treats the three possible cases: transport “onto” glaciers (i.e. non-glacier to glacier; I think this is what you have implemented and focus on), transport “off” glaciers (i.e. glacier to non-glacier; maybe this is what you restrict from occurring?), but also clarify whether lateral transport still occurs from non-glacier to non-glacier cell. On glaciers, is the laterally transferred snow considered as a separate source from the snow accumulation from snowfall? For example, is laterally transferred snow converted to glacier ice based on its mass but the snow accumulated from snowfall sits “on top” of the glacier and must melt off? (“the glacier only melts when it is not covered by snow”). Or are the two sources of snow put into the same reservoir which can only convert to glacier ice on sept 1? (in which case I guess there’s no glacier melt that season)

Response: Thank you for highlighting the need to better clarify the different cases of transport. We have implemented the given suggestion in the following way:

*“However, the snow **that is transported** is part of the glacier accumulation in many locations, which is why we here apply the lateral transport scheme only outside of glaciers. ~~and define the accumulated snow on glaciers as glacier accumulation.~~ When we introduce the glacier module, we thus restrict lateral snow transport to non-glacierized areas only. ~~Note that~~ This means that snow can **only** be transported **from a.) a non-glacierized cell to a non-glacierized cell and b.) from a non-glacierized cell to a glacierized cell.** There is no snow transport from a glacierized cell to either a glacierized cell or a non-glacierized cell. When snow is transported onto a glacierized cell, it is added to the snow cover and can thus later become part of the glacier accumulation (Kuhn et al., 2003; Freudiger et al., 2017).”*

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., & Weiler, M. (2017). Snow redistribution for the hydrological modeling of alpine catchments. *Wiley Interdisciplinary Reviews: Water*, 4(5), e1232.

Kuhn, M. (2003). Redistribution of snow and glacier mass balance from a hydrometeorological model. *Journal of Hydrology*, 282(1-4), 95-103.

Line 268: "The glacier ice reservoir only decreases when..."

Response: Thank you for your suggestion, which we have implemented this in the text:

"The glacier ice reservoir only decreases by melting when it is not covered by snow, following a simple temperature-index scheme ..."

Line 288-295: This is also confusing and needs clarification. Is the Huss et al relationship applied to the distribution of elevations from combining all the rasterized glacier cells across the domain? Or do you group individual rasterized glacier cells as belonging to specific real-world glaciers based on where they are located? And then you use the distribution of model elevations associated with those real-life groupings? Otherwise wouldn't each rasterized glacier cell have an elevation change in direct correspondence to its mass balance change?

Response: Thank you for making us aware of the need for further clarification of the delta h parameterization. We have clarified that we do indeed apply this on a glacier by glacier basis.

"Before running the model, we assign each individual glacier cell to the glacier it belongs to based on Farinotti et al., 2019. Then, we create offline maps of distributed glacier thickness, where each individual glacier loses a specific fraction of its mass (e.g. in steps of 1 percent mass loss) following Seibert et al. (2018a)."

"Mass changes do not necessarily occur in steps of 1 percent, leading to leftover mass or mass loss: for example, if the total mass loss of a glacier is 2.3 percent of the initial volume, we are left with 0.3 percent of leftover mass loss for this glacier. To address this, we distribute such leftover mass or mass loss evenly over its glacier area."

Line 303-308: I would describe this as a sensitivity experiment. Do both the thickness of the upper soil layer and the total soil thickness vary spatially in the model? When you state that your sensitivity test is to halve the upper layer thickness it sounds like it varies over the region, but then when you state the maximum upper layer thickness it sounds like it is spatially uniform.

Response: Thank you for your suggestion. While the different model experiments we performed to find a suitable change can be seen as a sensitivity analysis, we do not consider the entire change to the soil thickness to be a sensitivity experiment: our aim is not to test how sensitive the model is to changes in the model architecture, but we apply a targeted and deliberate change to the soil thickness to make it more realistic. In our opinion, this is more analogous to manual calibration than to a sensitivity experiment. We have clarified that the upper layer thickness is indeed uniform over the domain.

"Based on a sensitivity analysis, we decided to reduce the size of the top soil layer by making it half as thick everywhere, while maintaining the total soil thickness constant. Since the maximum depth of the upper soil layer in PCR-GLOBWB 2.0 is constant over the domain and by default set to 30 cm (Bierkens and Van Beek, 2009), halving the thickness still corresponds to a thickness of 15 cm, which is in line with the range of thicknesses that other LHMs are able to resolve (Telteu et al., 2021)."

Lines 369: afterwards this is referred to as the “water gap”, so please put this in parenthesis somewhere here.

Response: Thank you for making us aware of this. In response to the comments of another reviewer, we have decided to use the terminology provided by the original paper of Salwey et al., 2023, namely water balance signature (WB). The “water gap” would then be referred to as “negative value of WB”. We have implemented this in the text.

“One way to separate regulated from natural catchments is the water balance signature (WB), which describes the deviation from a closed water balance assuming no long-term storage effects (Salwey et al., 2023).”

Line 414-417: Check references to figures and forcings. I think there is a mistake here where either Figure 4H should read 4I or one of the references to the forcings should read CHELSA instead of CERRA-CHELSA.

Response: Thank you for pointing out this inconsistency. We have changed the references to the Figures.

Lines ~425/Figure 5: It might be helpful to explicitly state that the large amount of SWE present during the summer in the benchmark (at high elevations) and ERA5 data (at middle and high elevations) is due to the presence of snow towers and that the inclusion of snow transport (present in the runs labelled “transport”, “uncalibrated”, and “full run”) removes this unphysical effect. Also, please discuss the differences in SWE magnitude between the LHM model versions and the CERRA-Land analysis – does it also have snow build up in some cells, but resets to near-zero every year?

Response: Thank you for pointing out the need for further clarification. We have highlighted in the text now that snow transport is the reason why these snow towers are removed and that ERA5-Land and CERRA-Land also suffer from snow build up.

“..., where unrealistic snow towers were a major issue (compare the Transport run with the CERRA-CHELSA benchmark run in Figure 5 C and F). Here, the snow transport scheme ensures that the snow is redistributed to lower elevations, where it subsequently melts away. ERA5-Land and CERRA-Land also show very high SWE values suggesting that these models also suffer from unrealistic snow build-up.”

Figure 5: This is a really small point but your vertical axis starts a zero in plots a and d but below zero in plots b,c,e,f.

Response: Thank you for the detailed look at our figures. We have changed it so the axis starts at 0 everywhere.

Line 431: There is no calibrated snow run labelled in the figure. Are you using the Full run as a proxy for it?

Response: Thank you for pointing out this inconsistency. Indeed, we use the Full model run as a proxy for Y as it includes the same snow set-up. We have changed this here in the text and have further highlighted this in the Figure caption.

“...and the SNOWGRID product for Austria. Note that for simplicity we only show the Full model run instead of the Calibrated snow run, as these have the same snow module configuration.”

Line 442-443: I don't think I agree with this conclusion. Based on Fig 7a there doesn't seem to be much correlation with water gap sign and I don't see a pattern of KGESS associated with either +

or o in plots b,c. I do see a correlation of increased (decreased) KGE skill at locations with higher (lower) snowfall fraction with perhaps a weak dependence on glacier fraction. (The connection between performance and water gap sign for the snow/glacier modules and soil change are much more apparent).

Response: Thank you for highlighting this. We acknowledge that the KGESS mostly improves for catchments with high snowfall fractions. Our conclusions regarding the WB signature or reservoirs are mostly valid for such snow-dominated catchments. In the text, we have tried to better explain what we meant, limited the generalizability of the conclusions and do acknowledge that the relationship is weaker than for the other variables.

*“Figure 7A, B, and C illustrate that the **changes to the snow module** mainly **improve** model performance for discharge in **catchments with high snowfall fractions**, whereas in **catchments with low snowfall fractions the changes are negligible or slightly negative**. **Still, some catchments with high snowfall fractions experience decreases in performance for discharge: these decreases mostly happens in the presence of reservoirs and/or negative values of WB.**”*

Figure 6: This is a useful figure and tracks the progression of alterations nicely. Based on this, I'd suggest moving the KGE plot below plots a-f so that the full figure can take up more width on the page.

Response: We have changed the Figure accordingly.

Figure 7: I suggest labelling snow fraction as “snowfall fraction”.

Response: Thank you for the suggestion, which we have implemented in the Figures and in the text.

Figure 8: I don't find plot 8i helpful/insightful as currently presented and discussed. It would be fine to remove, retain the stated numbers in the text regarding the equilibrium experiment and just leave the model-obs comparisons as shown in plots 8g,h.

Response: Thank you for indicating the need to improve this Figure. Based on this comment and comments of other reviewers, we have made significant changes to Figure 8 to improve the legibility of the figure. We have removed subplot 8i from the Figure and given more space to 8g and h.

Figure 8: The glacier outlines really obscure the elevation change results. Is it possible to improve on these maps? Perhaps it's possible to use grey in all four maps to represent non-glaciated regions and to use a color palette for the elevation change that goes through white at zero instead of grey?

Response: We have made significant changes to Figure 8 and removed the glacier outlines of smaller glaciers to improve the figure legibility.

Figure 8: I suggest removing the sentence “Finally, the response of glaciers to continuous forcing with the mean mass balance from 1990–2018 (see Section 2.4).” from the caption.

Response: Thank you for the suggestion, which we have implemented in the caption. We did retain the time period indication and added it later in the caption.

*“(H) Comparison of e-folding response time **of glaciers forced with their mean mass balance from 1990-2018** to modelled estimates from Zekollari et al., (2020).”*

Line 461: You subsequently define rainfall-dominated catchments to be $P_s/P < 30\%$ and improvements in Fig 7e are all for snowfall fractions higher than this value.

Response: Thank you pointing out this inconsistency. We have changed this in the text.

“The positive effect of glaciers is much less visible in catchments with a small snow fraction; although some individual rainfall-dominated catchments also show a slight improvement in discharge simulations as a result of adding a glacier module (see Figure 7E).”

Line 491: Are the KGESS color bars the same in Fig 11 and 4? If so the changes in skill between non-calibrated and calibrated periods appear smaller on average than the effect of different forcing data (although the spread in values they cover at the extremes is close)

Response: Thank you for pointing this out. We have better quantified this and it is indeed the case that the difference in model performance when using different forcing datasets is larger than the change between the calibration and evaluation period. Note however that most of the model is not calibrated.

However, as we have moved this Figure to the Supporting Information along with most of the associated text based on comments from other reviewers, we have not highlighted this in the text.

Figure 11: It's hard to distinguish the blue and black colors used, particularly in the legend. Try a more differentiated color choice.

Response: Thank you for pointing this out. We have changed the colors in the Figure (which was moved to the Supporting Information).

Line 505: Your conclusion that “The representation of soil moisture ... needs to be improved in LHMs”: This might be true but I would argue your results also suggest there is a need to compare observed estimates of soil moisture and simulated values in a more representative manner. (I don't think you need to do this yourself in this paper.)

Response: Thank you for pointing out the need for a more representative comparison. We added this call to the text.

*“Please note that the reference product used for computing model errors can have its own biases (Dorigo et al., 2015, 2017) **and has a much coarser spatial resolution than our model, so error estimates might not be entirely representative. Our results therefore suggest that there is a need for more representative ways to compare soil moisture simulations and observations. Still, the representation of soil moisture and fast discharge responses needs to be further improved if LHMs are supposed to be applicable at smaller spatial scales.**”*

Figure 12: I suggest removing the label ‘B.’ from the first plot as it looks like a location you will refer to afterwards. Instead specify in the caption: Swiss canton of Grisons (inset shown in plot A)

Response: Thank you for your suggestion, which we have implemented accordingly in the figure.

Line 618: It might be worth specifying that these improvements apply even without calibration.

Response: We have added the following statement to the text:

*“Our analysis shows that our higher resolution model setup outperforms coarser reanalysis products such as CERRA-Land and ERA5-Land, **even without calibration** (Figure 5), but does not quite reach performance of national-scale reanalysis products such as the OSHD product.”*

Lines 630: I think this claim would require additional testing. You only test about 0.5 degrees away from your calibration period temperature (local increase) but for a climate change study you'd probably want to model a global mean temperature increase of another 2 degrees (and more locally). This would push your model quite a bit further than you've tested.

Response: Thank you for pointing out the need for further testing to support this claim. We acknowledge that the temperature changes over the study period are limited and agree that the statement here is too strong.

Furthermore, , we have moved the Figure to the Supporting Information, as it is not the main focus of this paper and the removal could simplify the structure of the text.

Textual changes:

“A short evaluation, which explicitly addressed model transferability, shows that model performance for discharge and SWE remains mostly consistent over the warmer and colder evaluation periods compared to the reference period, **although the temperature changes over the study period are limited** (Figure S1 in the **Supporting Information**).”

***Old:** “We therefore have some confidence in the transferability of our model to warmer climate conditions. Still, the general caveats the model remain applicable ...”*

***New:** “All of this suggests a reduced sensitivity of our model to increases in temperature. Still, the general caveats the model remain applicable ...”*

Line 656: I don't think it's fair to consider the CERRA-CHELSA forcing dynamically downscaled for the given model setup since the precip (likely the most important control) is taken as is from CERRA-Land at 5.5km and this is substantially coarser than your model at ~1km. Whereas the other products were downscaled (statistically) to 30". The way it is currently worded it sounds like the expectation was dynamical downscaling should yield improvements over statistical downscaling but I don't think you tested this hypothesis fairly (at similar resolutions) in your setup. I think you can still conclude that the choice of precip forcing did not make as large a difference on the resulting discharge accuracy as one might naively expect given the higher correlation and lower bias of the CERRA-CHELSA precip with observations). I think the conclusions as worded at lines 598-601 are more consistent with your experiments.

Response: Thank you for raising this point. While it is correct that the precipitation product is not further downscaled from CERRA-Land to the model resolution, we argue that the products should be placed in the context of the model dynamics of the underlying products.

Whereas products like ERA5 (or W5E5) underlying the STANDARD or CHELSA products have resolutions of 31 km or more, the dynamics in CERRA and CERRA-Land are run at a resolution of 5.5 km (Ridal et al., 2024), much closer to the resolution of our model. We thus hypothesized that CERRA should be able to represent small scale dynamics at a much higher resolution than the other products, even if these other products are statistically downscaled to the model resolution.

That being said, we agree that the wording can be improved: we remove the phrase dynamically-downscaled without context as we think this might indeed lead to the assumption that the data were downscaled to the model resolution and we try to name the underlying resolutions more explicitly.

In the introduction:

*“and (H3) using **dynamical-downscaled** forcing products that include a representation of smaller-scale atmospheric dynamics compared to other forcing products.”*

In the conclusion:

*“Discharge simulations forced by a **reanalysis product using high-resolution atmospheric dynamics at 5.5 km** (CERRA-CHELSA) did not consistently outperform runs with the other forcing products **which use coarser atmospheric dynamics at 31 km** (STANDARD and CHELSA; Hypothesis 3 is not supported).”*

Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., ... & Wang, Z. Q. (2024). Cerra, the Copernicus European Regional ReAnalysis system. Quarterly Journal of the Royal Meteorological Society, 150(763), 3385-3411.

663-664: The final sentence is really vaguely worded. Omit or add some more specificity on the types of questions you think the model is ideally suited for.

Response: Thank you for point this out. We rephrased this sentence and highlight possible applications.

*“Finally, we presented a new model setup with an improved representation of hydrological processes relevant in alpine regions, which is well suited to study regional and larger-scale streamflow and snow patterns in and around mountain regions. This new setup can be used to **help quantify water resources and to study how these are impacted by human water use or climate in the Alpine region or around the world.**”*

Line 701: 50 meters? Specify that E_{\max} and E_{\min} are measured in meters.

Response: Thank you for pointing out that the missing units, we have added these to the text:

*“Each glacier has a minimum surface elevation E_{\min} and a maximum surface elevation E_{\max} (**both in meters above sea level**). This topographic range of the glacier surface can be split into N elevation zones (in our case into 20 steps, with E_{\min} and E_{\max} rounded to the nearest multiple of 50 m).”*

Line 710: What is lower-case m ? the units of total glacier mass loss specified in units of meters? To me, writing (m) reads like “is a function of the variable m ”. I suggest rewriting to avoid this with something along the lines of: “This is done by means of a scaling factor f_S . This scaling factor is the ratio between the total mass loss over the glacier ΔM (units of m . water equivalent) and the integrated normalized change in surface elevation, ...) The units of all the subsequent variables should be clear from specifying those of ΔM .”

Response: Thank your for your suggestion. We have implemented it in the suggested way in the text. However, we still maintained the specification that h_i is the ice thickness also in units of m , water equivalent to facilitate orientation of the reader.

*“This is done by means of a scaling factor f_S . This scaling factor is the ratio between the total mass loss over the glacier ΔM (**units of m water equivalent**) and the integrated normalized change in surface elevation, scaled by the surface area of the elevation zone A_i (as a fraction of the total glacier area).”*

*“...and h_i is the ice thickness (**units of m water equivalent**) in each cell in that specific elevation zone.”*

Technical edits:

Line 136: downscaled

Response: Based on one of your earlier comments, we have decided to remove this word from the text.

Line 554: “than that of our LHM” or “than the resolution simulated here”

Response: We have changed this in the text.