----------------------- **Review 2** --

RC2: 'Comment on egusphere-2024-3072', Kristian Förster, 04 Mar 2025

In their manuscript "Hyper-resolution large-scale hydrological modelling benefits from improved process representation in mountain regions" Joren Janzing et al. describe a set of model extensions to PCR-GLOBWB 2.0 in order to make it more suitable for kilometre-scale modelling in mountainous regions. Together with model improvements in representing snow, glaciers, and soil, different forcing datasets are also tested. In essence, they found that the new model is capable of representing relevant processes in mountain hydrology quite well, even though the original model has been developed for large scale analyses with coarser spatial resolution. The manuscript fits well into the scope of Hydrology and Earth System Sciences. I found that the analyses carried out in the manuscript clearly show the added value of the most recent advances even though I think that some points require some more explanations and / or discussions. Apart from that I believe that the manuscript is an important contribution and I am looking forward to reading the final revised manuscript. Please find my comments below.

Response: Thank you for assessing our work and for your careful and constructive feedback. We address your suggestions and concerns below.

General comments

1. Structure of experiments / evaluations. The order of single steps is different throughout the manuscript. Abstract: forcing, snow, glaciers, soil; Introduction: snow and glaciers, soil, forcing. Hypotheses: snow and glaciers, "reviewing standard parameterizations" (I think that H2 should be rephrased in order to better reflect what was actually done. Do you mean reviewing and adjusting?), forcing. Methods and later results: Forcing, snow and glaciers, and soil. Conclusions: Snow and glaciers, soil, forcing. I was wondering if it would help the readers to agree on a consistent order of steps throughout the manuscript?

   Response: Thank you for highlighting the need for more consistency in the order of the experiments. We have followed your suggestions and improved consistency by generally following the order "Forcing; Snow; Glaciers; Soil" - this aligns with the order of our experiments. We have changed this in the conclusions and the hypotheses, meaning that the numbers of the hypotheses have changed.

   We make one exception and deviate from this rule in the introduction. The reason for this is that we consider the snow and glacier component to be more important for our storyline than the meteorological forcing.
   We have implemented the following changes in the text to clarify why the specific order was chosen (note that we have also changed the hypothesis regarding the standard parameterization in accordance with your suggestion).

   In section Introduction:
   *"**Nevertheless, any potential gains in model performance due to improved process representation or parameterization are constrained by the quality of the model input. It is thus important that such conditions are met first.** Hydrological modelling*

is sensitive to the meteorological forcing dataset used as input…"

"While large-scale hydrological simulations at higher spatial resolution have become feasible thanks to increasingly available computational resources, it is yet unclear by how much hydrological simulations can improve when combining such high-resolution models ***with the latest generation of meteorological forcing datasets and improved process representation****.*
Therefore, we here aim to explore the effect of ***(1)*** *using different meteorological datasets,* ***(2)*** *improving snow and glacier representations, and* ***(3)*** *changing runoff partitioning in the soil* ***on discharge simulations in PCR-GLOBWB 2.0****."*

"We hypothesize that hyper-resolution LHM performance for discharge in mountain regions will increase by ***(H1)*** *using forcing products that include a representation of smaller-scale atmospheric dynamics compared to other forcing products,* ***(H2)*** *improving the representation of mountain hydrological processes, such as snow and ice melt; and* ***(H3)*** *reviewing* ***and adjusting*** *standard parameterizations."*

In section Conclusion:
"- Meteorological forcing is uncertain over Alpine regions and different forcing datasets lead to major differences in model performance for discharge. These differences vary spatially between different forcing datasets.
- 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 1*** *is not supported).*
*- An improved representation of snow and glaciers improves SWE and discharge simulations in high Alpine catchments with natural flow conditions (****Hypothesis 2*** *is supported).*
*- The introduction of better runoff partitioning in the soil leads to an improvement of discharge simulations in smaller rain-dominated catchments by increasing their flashy response to rainfall peaks. However, these catchments still show overall weak model performance for discharge, suggesting that there is more to gain from improved routing in the soil (****Hypothesis 3*** *is supported)."*

1. Water balance signature in L371 (and also Figure 7): I found the adoption of the Water balance signature not well introduced into the evaluation even though it seems to be a reasonable indicator. Firstly, I would recommend to better explain what positive and negative values mean. From what I could get from the equation (and later by referring to Salwey et al., 2023), it seems to me that a positive value suggests a gain in water, while negative values would indicate a loss of water. Is that true? Please consider adding a short description that could help the readers to better follow your interpretation later. Secondly, I was wondering why you call this water balance signature later in the results section "water gap"? This even more confusing, since positive values seem to be related to additional water rather than a gap. I would suggest to refer to the original term suggested by Salwey et al., 2023.

Response: Thank you for the suggestions. We have changed our terminology to "Water balance signature" to match the terminology used in Salwey et al.,2023 and later in the text mostly refer to it as WB. Furthermore, we have added additional clarification of how this metric should be interpreted.

Textual changes:

*"**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 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."***

1. Glaciers in Figure 8: I agree that your modelling approach might be more suitable to regions of glaciers rather than individual ones. However, I really went through hard times reading Figure 8. Panels a) and b) are hard to read. Would be worth to increase the zoom level and to make glacier outlines smaller? Where is Mer de Glace (L447) on the map? I would suggest updating Figure 8 a) to d) in order to improve readability. The larger glaciers seem to have a positive mass balance bias, see panels e) and f), while in panel g) the model seem to have quite good average skill for glaciers with an area of more than 8 square kilometers. Why do we see a dramatic depletion in skill when looking at smaller glaciers? Why is 8 square kilometers chosen here? Is it related to the threshold in the Delta h method, proposed by Huss et al. (2010)? I was wondering if a regional mass balance, similar to e) and f), would better support your findings? Moreover, what remains unclear is whether the mismatch in glacier response time is related to differences in model (structure) or related to different forcings, when compared to Zekollari et al. (2020)?

   Response: Thank you for making us aware that this. Figure is difficult to understand. We have improved Figure 8 by implementing your helpful suggestions, such as further zooming in on the considered glaciers, showing thin outlines and highlighting some specific glaciers.

   Regarding the additional points you listed:
   - Mass balance bias of larger glaciers: we propose to highlight the Aletsch glacier and Glacier d'Argentiere in plot G and H, to indicate that there is indeed some bias, as you point out. Note however that the mean mass balances are only shown over the Evaluation period II.

   - Depletion of skill for smaller glaciers and for response times compared to Zekollari et al., 2020: Indeed, there is a significant drop in performance for smaller glaciers. We touched upon the possible reasons for this in the Discussion section.

"Still, we note that individual and especially smaller glaciers can show significant biases in the mass balance (see Figure 8G) and in the response times (see Figure 8H). The generally shorter response times than in previous experiments might be partly explained by the Δh-parameterization, which ignores potential delays in glacier response to mass changes (Seibert et al., 2018a). Further biases in both mass balance estimates and response times are likely related to the relatively coarse spatial resolution of our model compared to the size of individual glaciers, which makes it more difficult to accurately describe patterns in melt or snow accumulation for small glaciers consisting of only a few grid cells. Melt representation of individual glaciers could be improved by resolving glaciers at higher spatial resolution than the one of the LHM, for example by including elevation zones (Seibert et al., 2018a). Resolving glaciers at higher spatial resolution could also lead to even more realistic glacier retreat, since the Δh-parameterization was originally designed for higher spatial resolutions (Huss et al., 2010) and is likely less accurate at the coarser 30-arcsec model grid."

- Choice for 8 km2: We had chosen this threshold somewhat arbitrarily to illustrate the effect of glacier size on performance, as we noted a clear relationship between the two. We have now removed this sharp glacier area division and instead show a more continuous relationship between model performance and glacier area to make it less arbitrary.

- Regional mass balance: Thank you for this interesting suggestion. We have considered adding a subplot of a time series of the regional mass balance, similar to 8e and f, but in the end decided against this as this approach would include many additional assumptions. The resulting time series would heavily depend on our selection of glaciers -- of which only a few have (long term) yearly measurements and which might not be equally distributed in space and time.
Instead, we recreate subplot 8g based on geodetic measurements from Hugonnet et al., 2021, which have full spatial coverage and thus allow us to have regional estimates of the mass balance but only over a 10-year time period.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., ... & Kääb, A. (2021). Accelerated global glacier mass loss in the early twenty-first century. *Nature*, *592*(7856), 726-731.

1. Glacier initialization in 1990. How do initialize glaciers (area, volume) in your model setup when you deviate from the calibration period (which coincides with the data in Farinotti et al., 2019 for the early 2000)?

   Response: We indeed use the consensus estimate from Farinotti et al., 2019, which are representative for the early 2000s, to initialize our glaciers in 1990. We agree that this may lead to inconsistencies in glacier outlines around 2000. Therefore, we changed the implementation and keep the glacier volume and outline constant until 2000, although we still calculate ice melt and accumulation. This new approach is largely similar to previous large-scale hydrological modelling studies (e.g. Hanus et al., 2024).
   We acknowledge that even this updated procedure is still not optimal. However, to our

knowledge, there are no glacier volumes/ice thickness available representing conditions before 1990 and are available globally or continentally (needed to easily apply the model to other regions).

We have added some further clarification of the new approach in the text:

In section Datasets:
"***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. To our knowledge, there are no estimates of glacier volumes available that go further back in time and that cover all glaciers on the global, continental or Alpine scale. This glacier volume dataset has previously also been used to initialize glaciers for the same study period by Hanus et al., 2024.***"

In section Glacier module
"*To account for such temporal changes, we implement the empirical Δh-parameterization scheme from Huss et al., 2010, as implemented in HBV (Seibert et al., 2018).* ***As the glacier volumes are representative for around the year 2003, we only apply this scheme after the year 2000, keeping the glacier volume and area constant between 1990-2000 while still calculating the ice melt and accumulation."***

Farinotti, D., Huss, M., Fürst, J.J., Landmann, J., Machguth, H., Maussion, F., Pandit, A., 2019. A consensus estimate for the ice thickness distribution of all glaciers on Earth. Nature Geoscience. https://doi.org/10.1038/s41561-019-0300-3

Hanus, S., Schuster, L., Burek, P., Maussion, F., Wada, Y., & Viviroli, D. (2024). Coupling a large-scale glacier and hydrological model (OGGM v1. 5.3 and CWatM V1. 08)–Towards an improved representation of mountain water resources in global assessments. *Geoscientific Model Development*, *17*(13), 5123-5144.

Specific comments:

L15: What do you mean by unidirectional?
Response: We clarified that we mean that changes can be both positive and negative:

"Our evaluation of the effect of these different adjustments on model performance for discharge shows that while the meteorological forcing has a major effect on discharge simulations, ***it results in a mixed pattern of performance gains and losses*** *over the domain."*

L28: "thanks to heatwave". In my opinion you should highlight that this was only possible through a very negative mass balance.
Response: We have addressed this point by highlighting in the text that the mass balances were very negative.
"…whereas Alpine glaciers provided surrounding rivers with surplus melt-water during the 2003 Central European Drought ***due to very negative mass balances*** *(Van Tiel et al., 2023)."*

L60pp.: I think that there are lots of other references that support your statement here.
Response: Thank you for pointing out the need to include additional literature for this statement. We have added several references to support this statement, and have also specified that with "a constant DDF" we refer to a "time-constant DDF".

"However, LHMs often use very simplistic temperature-index schemes e.g. by using only **a time-constant** DDF or by omitting calibration **(e.g. Gosling and Arnell, 2011; Sutanudjaja et al., 2018; Müller-Schmied et al., 2021; Stacke and Hagemann, 2021)**."

Gosling, S. N., & Arnell, N. W. (2011). Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. Hydrological Processes, 25(7), 1129-1145.
Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., ... & Döll, P. (2021). The global water resources and use model WaterGAP v2. 2d: Model description and evaluation. Geoscientific Model Development, 14(2), 1037-1079.
Stacke, T., & Hagemann, S. (2021). HydroPy (v1. 0): A new global hydrology model written in Python. Geoscientific Model Development, 14(12), 7795-7816.
Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., ... & Bierkens, M. F. (2018). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. Geoscientific Model Development, 11(6), 2429-2453.

L74pp.: Please explain where exactly theses percentage values refer to.
Response: We have further clarified to which gauging stations these percentages refer to:

"Glaciers can be an important additional water source during their melt season (e.g. the average glacier storage change contribution to total runoff near the river mouth for the rivers Rhone ***at Beaucaire*** *in August: 25%, Rhine* ***at Lobith*** *in August: 7%, Danube* ***at Ceatal Izmail*** *in September: 4%, Huss, 2011)."*

L77pp.: Introducing one-way coupling would also suggest that two-way-coupling exists, too (see Pesci et al., 2023).
Response: We have removed the specification "one-way" to acknowledge this, as the general set-up (using glacier output into the model) is still the same.

"b.) using the output from an external glacier model as the input to the hydrological model (i.e. ~~***one-way***~~ *coupling the models; "coupled models")."*

L113. I felt some statement is missing here before introducing meteorological forcing.
Response: Thank you for your suggestion. We have added a sentence to smoothen the transition from the previous paragraph to the one on meteorological forcing:

*"****Nevertheless, any potential gains in model performance due to improved process representation or parameterization are constrained by the quality of the model input. It is thus important that such conditions are met first.*** *Hydrological modelling is* ***indeed*** *sensitive to..."*

L135: Please consider rephrasing H2 (as suggested before).
Response: Following your earlier comment, we have changed H2 in the way you suggested (it is H3 now):
*"...**(H3)** reviewing **and adjusting** standard parameterizations;..."*

L176: I am not sure whether "to guarantee" is correct in this context.
Response: Thank you for highlighting this. We have changed the word "guarantee" to "generate more":

"In contrast, near-surface air temperature can be further downscaled to ***generate more*** *accurate spatial melt patterns at 30 arcsec resolution."*

L239: DDF in general depend on land use. For the large-scale model it's clear that working with an "average" DDF makes sense. However, I remember from own research that DDF shows a quite different behavior in forests. Given the increased resolution and the relevance of forests in the water balance in mountainous areas, I think it is at least worth to discuss this later.
Response: We have expanded our discussion on spatial patterns in DDFs by explicitly discussing a potential role for land cover and vegetation.

In section Discussion:
"Another reason for this slight decrease in discharge performance at lower elevations could be related to our choice of regionally-averaged DDFs, since in reality these DDFs show smaller-scale variability in space (e.g. with aspect, albedo, elevation, ***landcover, vegetation,****…)* ***(Kuusisto, 1980; Rango and Martinec, 1995;*** *Hock, 2003; Ismail et al., 2023****)****"*

"More elaborate snow module formulations, such as parameterizations that include aspect (e.g. Immerzeel et al., 2012) or radiation (e.g. Hock, 1999), could increase our ability to capture more detailed spatial melt patterns. ***Such approaches can become feasible now that higher spatial resolutions can resolve slopes and represent vegetation cover and land use in a more detailed way.*** *However, they come at the cost of increased model complexity and a larger number of input variables."*

Kuusisto, E. (1980). On the values and variability of degree-day melting factor in Finland. Hydrology Research, 11(5), 235-242.
Rango, A., & Martinec, J. (1995). Revisiting the degree-day method for snowmelt computations 1. JAWRA Journal of the American Water Resources Association, 31(4), 657-669.

L261p: I do not understand that snow is transported to glaciers, given that transport is only considered to non-glaciated cells. Could you please explain this please?
Response: Thank you for highlighting the need for further clarification. Based on this comment and comments from the other reviewers, we have added further details:

"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 glacier accumulation (Kuhn, 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.

L271: Do you mean lower albedo?
Response: Thank you for making us aware of this inconsistency. Indeed, we meant lower albedo and have changed this in the text.

L294: If glaciers can only grow to its original extent, how do you realize to initialize the glaciers with a larger extent in the past?
Response: As pointed out above, this is a severe limitation caused by the lack of glacier volumes available further back in time. We thus keep the glacier volume constant until 2000, when we assume the volumes are the same as in the observed datasets.
We refer to the same added text as mentioned above:

"This information was derived from the consensus estimates from Farinotti et al., 2019, which are representative for the year 2003 for most glaciers. To our knowledge, there are no estimates of glacier volumes available that go further back in time and that cover all glaciers on the global, continental or Alpine scale. This glacier volume dataset has previously also been used to initialize glaciers for the same study period by Hanus et al., 2024."

L311: I am sorry if I missed this important detail but for me it's not clear if you calibrate a single parameter (e.g., DDFmin) for the entire domain or catchment-wise. From the table in supplement, it seems to me that it's a single value only which is valid for the entire domain after calibration.
Response: Thank you for raising this point. Indeed, we calibrate these parameters for the domain as a whole. We added some additional clarification.

*"We calibrate on the specific process (e.g. on SWE) instead of discharge to avoid compensating with parameter calibration for deficiencies in other processes and to increase the stability of these parameters under varying temperatures (Sleziak et al., 2020). We try to find one parameter set that is regionally valid****, i.e. constant over the entire domain****."*

L378: Why is snow calibrated only in Switzerland (I hope that I understood it correctly)? Here you focus on both Switzerland and Austria?
Response: Thank you for raising this point. We will further clarify that we do this because we are interested in testing the transferability of the model to other regions, as we do not have comparable snow products for other parts of the Alps such as Italy or France.

In section Calibration:
"... and because Switzerland covers diverse climatic regions. ***Note that we do not calibrate over Austria, even though we have similar data available there as well, to be able to test the transferability of the snow scheme to other regions.*** *To explicitly account for elevation-dependent melt patterns,..."*

In section Evaluation:

"In addition to discharge, we evaluate model performance for SWE by calculating spatial averages over different elevation bands (0-1000 m, 1000-2000 m, 2000-3000 m or the entire country) over Switzerland and ***also*** *over Austria****, as reference for domains where the model is not calibrated.****"*

L383: How do you evaluate glacier geometry evolution? Later you explain that this was only done visually.
Response: Thank you for raising this point. We have highlighted that we evaluate the glacier geometry evolution both visually to check the plausibility of spatial patterns and over longer time intervals compared to Zekollari et al., 2020**.**

*"To evaluate glacier **geometry evolution**, we both visually compare spatial patterns of glacier surface elevation changes against observations (GECM; see Table 2) **and quantitatively evaluate their long-term changes**. It is difficult to evaluate the long-term response of simulated glaciers to climate forcing given the relatively short study period."*

Zekollari, H., Huss, M., Farinotti, D., 2020. On the Imbalance and Response Time of Glaciers in the European Alps. Geophysical Research Letters 47, e2019GL085578. https://doi.org/10.1029/2019GL085578

L442p: Her, I think it would be interesting to provide a few more details what we see, e.g., the values around zero for small snow fractions. By the way: Is snow fraction the percentage of precipitation that is solid or how is it defined?
Response: Based on your comment and comments from other reviewers, we have provided some further details about the results and also indicated in the text that snow fraction indeed refers to the percentage of precipitation that is solid. We have further clarified this by using the term "snowfall fraction" and adding an explanation to the text.

*"Figure 7A, B, and C illustrate that the **changes in 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."***

L491pp: When comparing the drop in median for the different periods, it would be relevant to mention them in the text. Given the very good representation of SWE in the warmer period, is there any observation regarding the glaciers?

Response: Thank you for your comment. Based on comments from Reviewer 1, we have decided to place less focus on this section and moved the Figure to the Appendix. We add a subplot on comparing glacier performance between the different periods. We also add a quantification of the changes to the text in the Supporting Information.

Figure 12: Adding KGE or some other score would be helpful. Why does the full model compute a very high peak for Massa. Given that it is a 10 yrs. average, it seems to be a quite high event.
Response: We have added a KGE score for the benchmark and the final run to the figure.

We have further investigated the peak for the Massa river and it appears to be related to the choice that we move all the snow on the same day (i.e. beginning of the hydrological year), which led to increased melt around that day. We have rerun the runs with the glacier module, now remaining closer to the original method by Huss et al., 2010 and remove the snow continuously, which also removes this peak in discharge.

Huss, M., Jouvet, G., Farinotti, D., Bauder, A., 2010. Future high-mountain hydrology: a new parameterization of glacier retreat. Hydrology and Earth System Sciences 14, 815–829.

https://doi.org/10.5194/hess-14-815-2010

Technical comments:

Figure 4: Unfortunately, the legend hides important details of the distribution. Would it be possible to make its background transparent? See also Figure 5, 6, 11

Response: Thank you for the suggestion. We have moved the legend outside of the distributions plot for all suggested figures.

Figure 10: KGE refers to the upper row A-C, while KGESS to the bottom row (D-F)?

Response: That is indeed correct. We have further highlighted this in the Figure caption. *"Model performance of the Full model run for discharge (top row**, showing KGE**) and the total model performance changes for discharge compared to the CERRA-CHELSA benchmark run (bottom row**, showing KGESS**) in the evaluation catchments in relation to different catchment characteristics."*

References

Farinotti, D., Huss, M., Fürst, J.J., Landmann, J., Machguth, H., Maussion, F., Pandit, A., 2019. A consensus estimate for the ice thickness distribution of all glaciers on Earth. Nature Geoscience. https://doi.org/10.1038/s41561-019-0300-3

Huss, M., Jouvet, G., Farinotti, D., Bauder, A., 2010. Future high-mountain hydrology: a new parameterization of glacier retreat. Hydrology and Earth System Sciences 14, 815–829. https://doi.org/10.5194/hess-14-815-2010

Pesci, M.H., Overberg, P.S., Bosshard, T., Förster, K., 2023. From global glacier modeling to catchment hydrology: bridging the gap with the WaSiM-OGGM coupling scheme. Frontiers in Water 5, 1296344. https://doi.org/10.3389/frwa.2023.1296344

Salwey, S., Coxon, G., Pianosi, F., Singer, M.B., Hutton, C., 2023. National-Scale Detection of Reservoir Impacts Through Hydrological Signatures. Water Resources Research 59, e2022WR033893. https://doi.org/10.1029/2022WR033893

Zekollari, H., Huss, M., Farinotti, D., 2020. On the Imbalance and Response Time of Glaciers in the European Alps. Geophysical Research Letters 47, e2019GL085578. https://doi.org/10.1029/2019GL085578