----------------------- **Review 1** ---

RC1: 'Comment on egusphere-2024-3072', Anonymous Referee #1, 17 Feb 2025

Main comment

The authors take a global hydrological model (PCR-GLOBWB 2.0) and apply it over a large Alpine region. They propose several model adjustments to improve the model in order to better capture snow cover and discharge dynamics. I find it a very interesting study on a relevant topic and I see great potential in their work, however, I do not think that the analysis and manuscript at its current state fulfills the journal's requirements and some clarifications and revisions are required.

Response: Thank you for your careful assessment of our manuscript, we very much appreciate your detailed and constructive feedback. Below we will address your suggested points and clarifications one by one.

In my opinion, the strength of the manuscript (i.e. the usage of a large selection of forcing and evaluation data sets and the investigation of several processes) is its weakness at the same time, as it is getting more and more difficult to keep a concise and understandable workflow.

Response: Thank you for pointing out the need for an understandable workflow. We agree that our manuscript would benefit from a clearer structure and increased conciseness. We have addressed this by removing superfluous steps and increasing the consistency and readability of our text:

First, we moved the model transferability analysis, including Figure 11, to the Supporting Information, because it was not the main focus of the paper.
Second, we have restructured the "Forcing and datasets" section. We have now called the full section "Datasets", which we subdivided into subsections related to "Model input", "Evaluation and calibration data" or "Ancillary data". These sections will help the reader to understand why we used each dataset.
Third, we made sure to follow the same order of the steps in each section as in the experiments, that is "Forcing; Snow; Glaciers; Soil". The increased consistency will help readers to better follow the workflow as described in the manuscript. Accordingly, we have changed the order and numbers of the hypotheses and the conclusions. The one exception is the introduction, as we think that the main relevance is in the snow and glacier components and that these should therefore be mentioned first. We highlight this in the 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."

Furthermore, I lack the overall justification of applying a global hydrological model to the Alpine regional scale, as there are better suited models available for this task. Why not using models that have the required process implemented already?

Response: Thank you for pointing out that the model choice requires further justification. Global hydrological models are run over larger spatial domains, which also include mountain regions. However, these regions are often not well represented in such large-scale models.

First, we want to test such a global-scale model over a mountain region which is usually implicitly modelled in larger (global) scale applications. We decided to use the Alpine region as an example of such a mountain region and use the domain to evaluate the effects and limitations of different model adjustments, which can in principle be applied at larger scales.

Second, we want to highlight that other large-scale modelling approaches also have difficulty representing mountain regions. One of the reasons for this is that the smaller Alpine catchments with large elevation differences generally require higher resolution models with more advanced structures (Gurtz et al., 2003), so that detailed (sub-)national scale modelling approaches remain extremely popular in Alpine countries like Switzerland (Horton et al., 2022). Unfortunately, such spatial fragmentation precludes Alpine-wide studies, which is why improved kilometer-scale resolution larger-scale modelling approaches – as the one presented in this manuscript - can facilitate Alpine-scale spatial comparisons.

We have added the following statement to the introduction:

"We test the model over the larger Alpine region, as an example of a mountain region that is normally implicitly simulated by LHMs. However, a similar set-up can in principle be applied at larger scales. While more detailed hydrological modelling approaches are available for the Alps at the national or catchment scale, the Alps are very rarely studied as a whole at a similarly high spatial resolution, which inhibits comparisons between different regions."

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A., & Vitvar, T. (2003). A comparative study in modelling runoff and its components in two mountainous catchments. *Hydrological Processes*, *17*(2), 297-311.

Horton, P., Schaefli, B., & Kauzlaric, M. (2022). Why do we have so many different hydrological models? A review based on the case of Switzerland. *Wiley Interdisciplinary Reviews: Water*, *9*(1), e1574.

Specific comments

Title: Why is it hyper-resolution? If if would be a global application, then a 1km resolution run is termed hyper-resolution, I think. However, I do not understand why a regional application over the Alps in 1km should be termed hyper-resolution. Please clarify. Please be aware that in the snow-hydrological community, snow simulations in 1km are considered very course and not adequate to capture snow processes.

Response: Thank you for highlighting the need for further clarification. Indeed, within the global or continental scale modelling community, a 1km resolution is considered to be hyper-resolution. As mentioned earlier, we framed our study in the context of such large-scale hydrological models, with the Alpine domain serving as an example region, where high spatial resolution could provide added skill as it enables capturing the complex topography. A similar hyper-resolution model set-up can be easily expanded towards the European or global domain.

We are aware that a 1km resolution can be considered very coarse from the perspective of the snow hydrological community. However, high-resolution models applied in this community are often developed for local rather than larger scale applications. Since a trade-off exists between spatial domain and resolution, snow hydrological products at larger scales are also often relatively coarse. For example, Kraaijenbrink et al., 2021 modelled snow and snowmelt over the Himalaya at 5.7 km resolution and Mortimer et al., 2020 mention that the typical resolution of large to global scale SWE products is 25 to 100 km. To clarify the focus of our study, i.e. large-scale applications, we therefore refer to "hyper-resolution large-scale" hydrological modelling in the title to clarify that the focus is not on local-scale "hyper-resolution" modelling performed over a limited domain.

We have now further clarified this point in the introduction (same as the previous point):

*"We focus on the larger Alpine region, as an example of a mountain region that is normally implicitly simulated by LHMs. However, **a similar set-up can in principle be applied at larger scales.** While more detailed hydrological modelling approaches are available for the Alps at the national or catchment scale, **the Alps are very rarely studied as a whole at a similarly high spatial resolution**, which inhibits comparisons between different regions."*

Kraaijenbrink, P. D., Stigter, E. E., Yao, T., & Immerzeel, W. W. (2021). Climate change decisive for Asia's snow meltwater supply. *Nature Climate Change*, *11*(7), 591-597.
Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., & Tedesco, M. (2020). Evaluation of long-term Northern Hemisphere snow water equivalent products. *The Cryosphere*, *14*(5), 1579-1594.

Line 1: I was wondering. Are there actually major rivers that do not originate in mountain regions?

Response: Thank you for raising this question. There are indeed major rivers not originating in mountain regions. To provide just one example, the Volga river in Russia originates at an elevation of just 228 m (Litvinov et al., 2009). This river does encounter some tributaries further downstream that source from the Ural mountains, but these are also not very high (< 1900 m).

Litvinov, A. S., Mineeva, N. M., Papchenkov, V. G., Korneva, L. G., Lazareva, V., Shcherbina, G. K., & Shurganova, G. V. (2009). Volga river basin. Rivers of Europe, 2, 1.

Line 8: Please add in the abstract what is hyper-resolution for you? Please provide numbers.

Response: Thank you for pointing this out. We have now specified that the model runs at a 30 arc-sec resolution (approx. 1km).

Line 140: I do not understand 'calibrating SWE agains a reagional SWE'. Usually one calibrates parameters such as the DDF using measured snow. Please clarify.

Response: Thank you for your comment. As we are working over a large domain with grid cells of 1 by 1km, we decided to calibrate against a regional SWE reanalysis product that assimilates underlying observations and has both a high resolution and high quality. Alternatively, we could have calibrated against point measurements or satellite data. However, point measurements at stations are not representative at the spatial scale of the modelling effort and satellite observations of snow cover fraction might not contain information on the quantity of water

available in mountain regions. We clarified the statement in the introduction:
"...by calibrating SWE against a ***detailed*** *regional SWE reanalysis product* ***with assimilated observations****..."*

Fig. 1: For me this map was a bit misleading, as I was expecting you to simulate the runoff of the large rivers, but as I understood, you do not look at discharge from those rivers. This map shows another area than what you actually analyze and simulate. Please consider to adapt and show the exact modelling domain.

Response: Thank you for raising this point. The way our model is set up, we do simulate the entire domain depicted in Figure 1 and we have grid cells representative of the whole domain. Our analysis is indeed mostly focused on the comparison at different sets of observational stations, with some stations located on the rivers depicted. However, soil moisture is evaluated over the entire domain in Figure 9 and we also depict model results for the full domain in Figure 12.

We have updated Fig. 1 now to show both the model domain and added subplots with the locations of the stations where the model is evaluated.

Fig. 2: As far as I understand, the implementation of the snow transport was not part of this study and hence should be part of the benchmark model in my opinion. The presentation and evaluation of the snow transport scheme was done in another study, right? The removal of this step of complexity in the analysis also could make your total analytical set-up more concise.

Response: Thank you for your suggestion. The snow transport implementation was indeed introduced in Van Jaarsveld et al., 2025. However, that paper evaluated snow performance against snow cover fraction only. Here, we explicitly evaluate the added value of this specific step using SWE data. This enables us to quantify its effect relative to the effect of the other development steps and highlights the benefit of the snow transport step in terms of removing the snow tower (i.e. continuous build up of higher and higher SWE values over time) problem. We think that it is important to show how snow transport influences the presence of snow towers, as many models do struggle with such unrealistic features. Therefore, we retained this step as a separate analysis step.

Van Jaarsveld, B., Wanders, N., Sutanudjaja, E. H., Hoch, J., Droppers, B., Janzing, J., ... & Bierkens, M. F. (2025). A first attempt to model global hydrology at hyper-resolution. *Earth System Dynamics*, *16*(1), 29-54.

Line 259: What is the SWE threshold used in this study? Please state and explain how it was derived.

Response: Thank you for highlighting the need for clarification. We used a threshold of 0.625 m as in Van Jaarsveld et al., 2025, which is based on Frey and Holzmann, 2015. We have specified this in the text:

"This scheme transports part of the snow downhill based on the surface slope whenever the snowcover exceeds ***a SWE content of 0.625m****.* ***This threshold is based on values for the snow holding capacity for forest cover and snow density in Frey and Holzmann , 2015, using a similar approach as in CWatM (Burek et al., 2020).****"*

Frey, S., & Holzmann, H. (2015). A conceptual, distributed snow redistribution model. *Hydrology*

and Earth System Sciences, *19*(11), 4517-4530.
Van Jaarsveld, B., Wanders, N., Sutanudjaja, E. H., Hoch, J., Droppers, B., Janzing, J., ... & Bierkens, M. F. (2025). A first attempt to model global hydrology at hyper-resolution. *Earth System Dynamics*, *16*(1), 29-54.

Line 310: You add the snow routing and calibrate it offline. What were the other model parameters of the model calibrated on? If they were calibrated on discharge, do not all parameters have to be re-calibrated again?

Response: Thank you for your comment. The original setup of PCR-GLOBWB has not been calibrated (Sutanudjaja et al., 2018) and the other parameters were derived using information on land cover and other types of spatial information. For example, the soil parameters are derived from existing landcover maps (Hoch et al., 2023). We clarified this in the text as follows:

"PCR-GLOBWB 2.0 ***has not been*** *calibrated* ***and generally uses parameters derived from external datasets (Sutanudjaja et al., 2018). Still,*** *we do calibrate the degree-day factors of snow and ice to increase regional applicability and to ensure realistic glacier geometry evolution, since this evolution can be sensitive to biases in glacier mass balance."*

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., Van Beek, R. L., & Bierkens, M. F. (2023). Hyper-resolution PCR-GLOBWB: opportunities and challenges from refining model spatial resolution to 1 km over the European continent. *Hydrology and Earth System Sciences*, *27*(6), 1383-1401.

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.

Line 313: Please be aware that the SWE data products you use also only are model output and the also these models are (for different reasons) often incorrect.

Response: Thank you for pointing this out. We acknowledge that the SWE reanalysis product is still a modelled product. As mentioned before, we used this data product because it matched the purpose of our analysis best and because sufficiently detailed observational products over the Alps do not exist. We worked with this dataset, because it has high quality, assimilates underlying observations, and has regional spatial coverage.

We justified this choice in the text as follows:

"...we used a snow water equivalent reanalysis product over Switzerland (Mott et al., 2023). We chose this dataset because of its ***extensive data assimilation,*** *its spatial continuity, its high quality which is unmet by products covering larger spatial domains (e.g. ERA5-Land or CERRA-Land), and because Switzerland covers diverse climatic regions.* ***However, as every other SWE dataset, this product also comes with some uncertainties."***

Line 342: What different climatic regimes are meant here?
Response: We will further clarify this and specify that we tested for the transferability of the model to time periods with a higher or a lower average temperature:
"...we use both evaluation periods to assess the transferability of the new model set up to ***time periods with different mean temperatures*** *within the framework..."*

As mentioned under the general comments, we will move this part of the analysis to the appendix and only briefly mention it in the text in order to reduce complexity.

Line 370: I am a bit skeptical about the applicability of the WB measure in the Alps to assess the influence of reservoirs. As you mention, it is also strongly impacted, e.g., by the the meteorological data used. In my opinion, your results (e.g. Fig. 7) showing the general deviations of the WB from zero are more an indication of the big uncertainties in precipitation and evapotranspoiration. Hence, the WB is a poor measure for reservoir influence and I am not sure what is then the validity of this measure to stay in the manuscript. To me the calculation of the WB does not provide new insights and only adds an unnecessary level of complexity to the study. Please think again what is the added value of calculating and showing the WB so prominently.

Response: Thank you for raising this point and asking us to reflect on the added value of this metric. The WB signature is indeed influenced by other processes than regulation such as uncertainties related to meteorological forcing.

Still, we argue that there is value in using this metric in our manuscript:
1. The WB signature is related to the presence of strong regulation as shown in Figure i in this response. The hidden variable that confuses the picture might be the "degree of regulation" or the "type of regulation". There are many catchments with reservoirs that are not so heavily regulated, which can obscure some of the results. However, Figure i, which we added to the Supporting Information, clearly demonstrates that the degree of regulation (here defined as "the total reservoir volume over the catchment divided by the mean annual discharge"; this results in the amount of time the reservoirs can store the catchment discharge) is generally related to negative values of the WB signature and on average to stronger deviations from 0.
However, this finding clearly does not hold for all catchments, as not all catchments experience strong water abstractions. Salwey et al., (2023) – who introduced the metric – acknowledge that there is no one-size-fits-all indicator of reservoir regulation on streamflow. Therefore, we think that the presence of reservoirs and the WB signature can together provide complementary information on the effect of regulation on model performance.

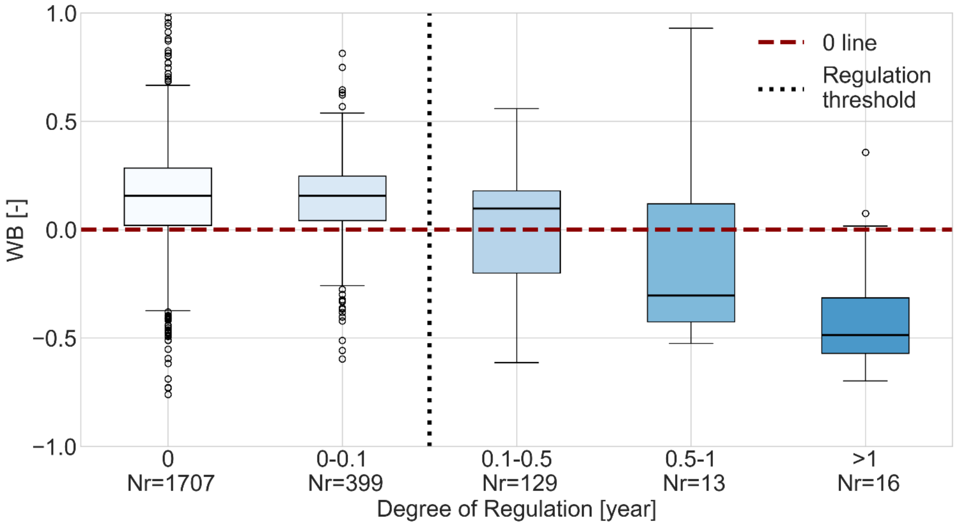

Figure i Relationship between the degree of reservoir regulation and the WB signature. The degree of reservoir regulation is here defined as the total reservoir volume over the catchment divided by the mean annual discharge.

2. There is a clear dependence of model performance changes on this WB metric (see Figure 7 and 10 in the manuscript). Performance mostly increases when this metric is positive and decreases when this metric is negative. Regardless of the relationship to reservoir regulation, this adds value by itself.
The metric shows that also other influences -- such as biases in meteorological forcing – can importantly influence model performance and that our model can often not compensate in catchments where meteorological forcing is biased. Furthermore, the metric can identify catchments with a significant meltwater input related to glacier mass loss.

We do acknowledge that the choice of this metric has to be better justified in the manuscript and propose the following modifications and clarifications.

1. We redefine the term "regulated catchment" in the revised version of the manuscript. Instead of just showing if a reservoir is present, we now define a threshold based on the degree of regulation using Figure i. Catchments are classified as regulated when its degree of regulation exceeds 0.1 year (threshold based on the first change of the median in Figure i): this change will avoid classifying catchments with little regulation as regulated.

2. We improve the separation of conclusions derived from the analyses using the WB signature only and conclusions for regulated catchments derived from analyses jointly considering the WB signature and information on reservoirs. We also added quantitative qualifications for regulated and unregulated catchments in the text. We further highlight in the text that WB anomalies can for example be related to biases in meteorological forcing and glacier mass loss.

Textual changes:

*"**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.**"*

*"A major control limiting model performance for discharge in Alpine catchments **are strong deviations from a closed water balance. Such deviations are sometimes related to** hydropower plants or water transfers that are not accurately represented in the model, which should be a priority of future model development."*

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*(5), e2022WR033893.

Line 402: In my opinion, the comparison of the meteorological input data sets with regard to the discharge performance (Fig. 4 G,H and I) should be conducted after the model routines have been improved. As seen in Fig. 4, the overall model performance seems fairly low with a median KGE barely above 0. If the models routine is not good enough, also a better precipitation input, for example, cannot improve runoff in a snowmelt-dominated catchment. Please think of moving the evaluation of the different meteorological input data sets at the end of you workflow.

Response: Thank you for your suggestion. We appreciate the feedback and have carefully considered moving the meteorological analysis to the end of the analysis. However, we have finally decided to keep the order of experiments as in the original manuscript because of the following reasons:

First, the main part of the model does not change by changing the input dataset: we do not alter any parameters related to landcover, groundwater, routing or human water use. As mentioned before, PCR-GLOBWB is generally not calibrated (Sutanudjaja et al., 2018) and these parameters are derived from specific data sources, e.g. from datasets on landcover. Second, the largest changes in model performance we found were related to meteorological products, albeit with mixed performance changes. Although improving the different model routines leads to general model performance gains, the model forcing has a major effect on model performance. This highlights that a model can only improve as much as the quality of the model input allows. Therefore, we think that it is crucial to first identify the best forcing dataset. Once this dataset has been identified, we then test various other ways of potential model improvement by adjusting different model routines. Furthermore, as the forcing dataset has a very large effect compared to the other modelling choices, the general differences in model performance between meteorological forcing datasets would persist regardless of when we evaluate the various meteorological forcings.

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.

Line 411: I do not see a general improvement of discharge. It looks a bit random to me. How do you come to the conclusion that the performance is 'decent overall'. Please quantify.

Response: Thank you for pointing this out. We removed the phrase "although performance is decent overall" from the text as this is subjective and quantified the discharge performance in the previous sentence instead.

Line 426: I am not sure I can see this improvement in the Fig. 5 G and H. Looking at Fig. G and H I do not see any improvement in model performance with increasing complexity of the snow routine.

Response: Thank you for making us aware of this. Indeed, there is no improvement for the observational stations in 5G and H. We have now explicitly mentioned this in the text, and at the same time added an explicit quantification of the regional performance changes per elevation zone.

"Without calibration, further structural changes to the snow module (i.e. the seasonally varying DDF, exponential temperature dependence, and a rain-to-snowfall transition temperature range) ***did not improve performance against observational SWE stations (see Figure 5G and H). Averaged over elevation zones, however, this change*** *lead to an improvement of the SWE representation ..."*

Line 430: I do not see the evaluation of 'melt rates'. You calibrate a DDF which is the same for all the area. How are there different melt rates depending on elevation? Please explain.

Response: Indeed, we don't directly evaluate melt rates. What we mean by melt rates is the amount of snow melted per day (with unit m/day), which is different from the DDF (with unit m/day/deg C). As the DDF varies with the season and snowmelt timing varies with elevation, this leads to elevation-dependent differences in the melt rates at the same temperatures. Therefore, we conclude that the differences in modelled SWE capture such elevation-dependent patterns.

We have further clarified this in the text:

*"These differences between elevation zones suggest that the model structure **of a time-varying DDF combined with an elevation-dependent timing of the melting season leads to average melt rates varying with elevation.**"*

Line 462: Please explain the performance decrease in the south-western Switzerland.

Response: We clarified in the text that such performance changes are likely due to reservoir regulation:
*"Furthermore, our results indicate that in certain regions, especially around the **heavily-regulated** Rhône river in south-western Switzerland, discharge performance can decrease with the addition of glaciers (see Figure 6E). Such performance decrease generally occurs in catchments with a negative **WB** or reservoir **regulation** (see Figure 7F)."*

Line 491: 'These changes have similar magnitudes as the changes due to different forcing data'. This is an interesting sentence, as you previously state that the forcing data has a very strong influence. Does this mean also the selection of the evaluation period has a strong influence?

Response: Whereas changes for individual catchments indeed depend on the chosen evaluation period, performance over the whole domain does not vary much with the choice of evaluation period (see Figure 11F).
We have clarified this in the text:
*"These changes have similar magnitudes as the changes due to different forcing datasets (compare to Figure 4G,H, and I**), suggesting that the selected evaluation period is important for model performance in individual catchments. For the domain as a whole,** the warmer Evaluation II period shows a slight drop in median performance…"*

Note that this text on the Transferability analysis will be moved to the Supporting Information as mentioned before.

Line 494 and Line 630: 'Transferability to warmer climate conditions': I do not think that the analysis at the current state sufficiently supports this statement. You use an highly empirical approach for snowmelt and calibrate it to a specific time period. I do not see how the comparison of 10-year time slices can prove that the DDFs will be the same end of the century. In my opinion, the discussion of the transferability of DDFs in time also is not the focus of the study. As there are a lot of other interesting aspect to focus on, please consider to shorten/revise/remove this part.

Response: We agree that this is not the main focus of this study. Therefore, we have decided to move the "Transferability" section to the Supporting Information and removed or weakened any claims related to this section in the text:

In Discussion:

*"**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).**"*