We thank the reviewer for their constructive comments, which have helped to improve the manuscript. Below we provide a point-by-point response to each comment. Reviewer comments are shown in *grey italics*, followed by our responses.

# General Comment

*This manuscript presents a solid study that integrates the global glacier model GloGEM with the HBV model implemented in the Raven framework, together with a snow redistribution scheme, across 14 glacierized headwater catchments in Switzerland. The study offers a systematic assessment of how glacier constraints, snow redistribution, and precipitation correction affect melt partitioning and parameter behavior at the catchment scale. Overall, the work is carefully conducted and valuable.*

Thanks for the overall positive appraisal.

# Main Comments

*That said, I have several concerns regarding the experimental design and the novelty of the main findings. My main comments are as follows:*

***The main contribution of the paper seems stronger as a systematic regional assessment than as a genuinely new methodological advance.*** *The study is solid and useful, but the novelty of the main findings appears somewhat limited. The main messages of the paper are that (i) good streamflow performance does not necessarily imply realistic cryospheric process representation, (ii) uncoupled HBV can compensate forcing biases through excessive glacier melt, and (iii) introducing a precipitation correction factor can aggravate equifinality and uncertainty. These are important points, but they are not new in cold-region hydrology and glacier-hydrology modeling. In this sense, the main value of the manuscript seems to lie in the systematic multi-catchment comparison across 14 headwater basins, rather than in delivering new methodological or process-level insight.*

We thank the reviewer for the summary of our main contributions and we agree that we could better emphasize how our work goes beyond the state-of-the-art. We acknowledge that the individual findings, such as the disconnect between streamflow performance and cryospheric process realism, or equifinality introduced by precipitation correction, are known challenges in cold-region hydrology. However, we would like to clarify and emphasize what we believe constitutes the novel contribution of this work:

1. To the best of our knowledge, no study has yet integrated glacier melt simulations from a global glacier model (GloGEM) directly as forcing input into a conceptual hydrological model (Raven) at the local catchment scale. As shown in our introduction, existing studies coupling global glacier model output with hydrological models are limited to large-scale global applications, and this study explicitly addresses that gap. We will make this clearer in the revised version.

2. We agree with the reviewer that the systematic multi-catchment evaluation across 14

diverse headwater basins represents a core value of this study. However, it should not just be seen as a comparison: rather, this contributes to demonstrating the transferability and limitations of this coupled approach across a range of catchment characteristics. We will emphasize this point in the revised version.

***The role of the precipitation correction factor (PCF) needs to be explained more clearly.*** *The manuscript lists PRECIP_CORR as a precipitation correction factor in Table 2 and treats it as a separate configuration dimension in Table 3 (e.g., HBV-PCF and HBV-GloGEM-PCF). However, its hydrological meaning and implementation remain insufficiently explained. At present, the manuscript provides a label, but does not clearly explain what exactly is corrected, how the factor operates within the HBV/Raven setup, or what physical bias it is intended to represent. It is also unclear why calibration of this parameter is treated as a dedicated scenario rather than simply as one additional calibration parameter.*

Thanks for pointing out the lack of details regarding this important aspect of our work. Precipitation in high elevation catchments is notoriously difficult. Two main difficulties arise: i) observed point precipitation can be biased due to gauge undercatch (Goodison et al., 1998), which might lead to up to 50 % bias for solid precipitation (Rasmussen et al., 2012). ii) spatially distributed precipitation in high-elevation catchments remains uncertain, as gauge networks are typically strongly biased towards lower elevations (Frei and Schär, 1998). Our precipitation correction factor (PCF) accounts for both aspects, as in many similar model applications (van Tiel et al., 2020), and this will be made clearer in the revised manuscript.

Practically, we implemented precipitation correction by multiplying all observed precipitation values (i.e. all observed values at all time steps) with a spatially and temporally uniform precipitation multiplier, ranging from 0.8 to 1.8.

Furthermore, we would like to explain why the PCF is introduced as a dedicated scenario. In the uncoupled HBV setup, the model can compensate for precipitation biases by producing excessive glacier melt. In the coupled HBV-GloGEM setup, however, glacier melt is given as an input to the hydrological model, meaning it can no longer compensate for potential precipitation biases in this way. The PCF scenario was therefore introduced to test whether correcting the precipitation input could simultaneously reduce the overestimation of glacier melt in HBV and the streamflow underestimation in HBV-GloGEM. Understanding this compensation effect requires a dedicated configuration rather than simply an additional calibration parameter: each configuration in this study corresponds to a separate calibration run with a distinct parameter set. This will be made clear in the revised version by reformulating parts of the methods section.

***The contrast between the two model setups may not primarily reflect "whether GloGEM is coupled," but rather whether glacier information is used to constrain the glacier component.*** *In the current design, the coupled setup benefits from glacier-specific calibration against geodetic mass balance, whereas the original HBV setup is calibrated only against streamflow. This makes it difficult to attribute the differences solely*

*to coupling with GloGEM. In practical terms, both the GloGEM-based runoff and the melt generated by the simplified HBV glacier routine ultimately act as inputs to the hydrological simulation. It is therefore plausible that, if the native HBV glacier routine were also constrained using glacier data, it could produce more comparable glacier-melt estimates. The manuscript should discuss this point more explicitly and clarify whether the key distinction is really coupling itself, or the use of glacier observations to constrain the glacier contribution.*

Thanks for commenting on this important point that deserves explicit discussion in the manuscript, and we will address it more thoroughly in the revised version in the discussion section.

As rightly summarized above, the two setups differ in how glacier melt is simulated (by GloGEM or by the native HBV glacier routine via Raven), and also in the information used to constrain the models (GloGEM was calibrated against a global geodetic mass balance product, the HBV glacier model only against streamflow).

Calibrating the native HBV glacier routine against geodetic mass balance would face two fundamental obstacles: (i) the semi-distributed model structure lacks glacier-specific resolution: glacier melt is governed by a single set of degree-day parameters applied across all glacier HRUs, whereas geodetic mass balance data (Hugonnet et al., 2021) is resolved per individual glacier. Aggregating per-glacier estimates into a catchment-wide constraint would average out glaciers with different characteristics, providing only a weak calibration target. GloGEM's advantage lies not in a fundamentally different melt formulation, as both rely on temperature-index methods, but in its ability to calibrate parameters individually for each glacier using its specific area-altitude distribution. (ii) the uncoupled model treats glacier ice as an unlimited reservoir that is never depleted, and lacks a mechanism to convert accumulated snow into glacier ice. The glacier can therefore only lose mass through melt but never gain it through accumulation, meaning the model cannot close the glacier mass budget. Comparing such an incomplete mass balance against geodetic observations, which integrate both accumulation and ablation as well as geometric change, would not constitute a meaningful calibration target.

These limitations are not unique to our setup. Most hydrological models represent glaciers through simplified routines that neither resolve individual glaciers nor simulate the full glacier mass balance, making it inherently difficult to incorporate glacier-specific observations into their calibration. This is exactly one of the gaps this study aims to address, by exploring if glacier runoff from a global glacier model can be used as input for a hydrological model and by providing a mass-conservative, transferable and model-agnostic method to feed the glacier melt into the hydrological model that has a different spatial set-up.

We agree, however, that it would be valuable to test whether the observed differences between the two setups stem from the coupling itself or from the different calibration constraints used. We will highlight this explicitly in the revised manuscript as a direction for future work.

***The snow redistribution module should be introduced and visualized more clearly.*** *The manuscript explains why snow redistribution is needed and briefly describes how it is implemented. However, its role is not clearly shown in the workflow figure, which makes the overall methodology harder to follow. I suggest either adding this component explicitly to Figure 2 or clarifying its position and function more clearly in the text.*

We agree that the snow redistribution module was not sufficiently visible in the workflow. A more detailed description of the module will be added to an appendix to improve the clarity. In brief, the module is based on a parameterization originally developed to simulate avalanche-driven snow transport (SnowSlide, Bernhardt and Schulz (2010)), but is used here not to represent avalanching explicitly, but rather to avoid unrealistic snow accumulation and to redistribute excess snow to lower elevation HRUs where warmer temperatures allow melt to occur. Snow transport is triggered when a slope-dependent snow-holding depth threshold is exceeded, and HRU connectivity is derived during catchment preprocessing using D8 flow routing, with redistribution weights based on upstream contributing area. No additional calibrated parameters are introduced.

***The treatment of snow-covered area (SCA) remains unclear, and the lack of SCA-based evaluation weakens confidence in the snow and melt partitioning results.*** *I may have overlooked it, but I could not find a clear explanation of how snow-covered area is determined in the model. In addition, the model does not appear to be evaluated against SCA data. This limits confidence in the snowmelt simulations. Even if catchment-average SWE is reproduced reasonably well, the spatial distribution of snow and the timing and location of melt may still remain uncertain without an independent constraint on snow extent. I therefore recommend that the authors clarify whether SCA is represented or diagnosed in the current workflow, and discuss the implications of not evaluating SCA for the interpretation of snow storage and melt contributions.*

Thanks for pointing out that it is not entirely clear in the paper how snow covered area (SCA) is computed. In catchment-scale hydrological modelling, there are two options to compute catchment-average snow accumulation and melt: i) in models without explicit spatial reference for snow simulation (i.e. in fully lumped models), snow accumulation and melt is computed for a reference elevation (from temperature and precipitation) and then linked to SCA via pre-defined "snow height - snow covered area curves" (Martinec et al., 1983). ii) in spatially-explicit models (semi-lumped, semi-)distributed, in particular models with elevation bands, snow accumulation and melt is computed per spatial unit and SCA is obtained directly at each time step from the sum of all spatial units that are snow covered.

Our model setup uses HRUs: these are regrouped "patches" of the catchment with the same land use type and elevation band. Each HRU has its mean elevation and we compute snow accumulation and melt for this mean elevation. In this model setup, SCA is computed internally at each time step as the sum of snow-covered HRUs, but was not used as an evaluation metric. Instead, snow simulations are evaluated against a gridded SWE product (Marty et al., 2025), using catchment-average SWE as the presented metric (Section 4.3, Figure 5). We acknowledge that this approach has two limitations: i) the reference SWE

product is itself model-based, although it assimilates in-situ snow depth observations from 350 stations via an Ensemble Kalman Filter (Magnusson et al., 2014), which we consider a sufficiently robust observational constraint for the purposes of our study; ii) as noted by the reviewer, a model could in principle reproduce catchment-average SWE well while misrepresenting SCA. Purely satellite-based snow products would in principle allow for an additional evaluation, but currently available products such as MODIS provide SCA rather than distributed SWE. Moreover, since the hydrological model is HRU-based and only knows the mean elevation of each HRU but not its exact location in the catchment, the spatial extent of snow cover has no direct meaning for the model structure. Given the known hysteresis between SCA and catchment-average SWE (Riboust et al., 2019), evaluation based on SCA alone also gives an incomplete assessment of snow routines. Accordingly, we believe that catchment-average SWE provides a robust measure of snow storage and its seasonal dynamics, capturing the information most relevant to streamflow generation. In the revised manuscript, we will clarify the computation of SCA in the model description and discuss the use of observed SCA for model evaluation as an outlook for settings where no reference SWE product is available.

# Minor Comments

*Please clarify explicitly whether SWE was used only for validation or also during calibration. From the current text, SWE appears to have been used for validation rather than calibration. Since SWE results are shown prominently, this point should be stated unambiguously.*

No SWE data is used in the calibration procedure. The gridded SWE product is only used to validate simulated SWE outputs from the snow parameters calibrated on streamflow. This will be made more explicit in the Manuscript.

All other minor issues like naming concerns and typos will be addressed in the revision.

# References


Bernhardt, M. and Schulz, K.: SnowSlide: A simple routine for calculating gravitational snow transport, Geophysical Research Letters, 37, 2010GL043 086, https://doi.org/10.1029/2010GL043086, 2010.

Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, International Journal of Climatology, 18, 873–900, https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8⟨873::AID-JOC255⟩3.0.CO;2-9, 1998.

Goodison, B., Louie, P., and Yang, D.: The WMO solid precipitation measurement intercomparison, Tech. Rep. WMO/TD No. 872, World Meteorological Organization, Geneva, 1998.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., and Kääb, A.: Accelerated global glacier mass loss in the early twenty-first century, Nature, 592, 726–731, https://doi.org/10.1038/s41586-021-03436-z, 2021.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods, Water Resources Research, 50, 7816–7835, https://doi.org/10.1002/2014WR015302, 2014.

Martinec, J., Rango, A., and Major, E.: The Snowmelt-Runoff Model (SRM) user's manual, Tech. Rep. NASA Reference Publication 1100, NASA Goddard Space Flight Center, Greenbelt, MD, 1983.

Marty, C., Michel, A., Jonas, T., Steijn, C., Muelchi, R., and Kotlarski, S.: SPASS – new gridded climatological snow datasets for Switzerland: potential and limitations, The Cryosphere, 19, 4391–4407, https://doi.org/10.5194/tc-19-4391-2025, 2025.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E.: How well are we measuring snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed, Bulletin of the American Meteorological Society, 93, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1, 2012.

Riboust, P., Thirel, G., Le Moine, N., and Ribstein, P.: Revisiting a simple degree-day model for integrating satellite data: implementation of SWE-SCA hystereses, Journal of Hydrology and Hydromechanics, 67, 70–81, https://doi.org/10.2478/johh-2018-0004, 2019.

van Tiel, M., Stahl, K., Freudiger, D., and Seibert, J.: Glacio-hydrological model calibration and evaluation, WIREs Water, 7, e1483, https://doi.org/10.1002/wat2.1483, 2020.