Author’s Response to Reviews of

Modelling runoff in a glacierized catchment: the role of forcing product and spatial model resolution

Alexandra von der Esch, Matthias Huss, Marit van Tiel, Justine Berg, Daniel Farinotti

Dear Editor,

We thank you for your feedback on our manuscript. We appreciate the suggestions by you and the three reviewers and have implemented them into our manuscript. Below, we address all the reviewer comments. Our Authors' replies are written in blue. The provided line numbering refers to the manuscript sent in after the first review. New text parts are highlighted by underlining them. RV1 refers to Reviewer 1, RV2 refers to Reviewer 2 and RV3 refers to Reviewer 3.

Summary of changes:

RV1:

- We clarified the structure of GERMs runoff routing and added a dedicated section in the discussion to discuss its structural limitations.
- We clarified that we did not use SMB for calibration but Geodetic glacier ice volume change. Furthermore, we clarified that we did not include snow observations for calibration to remain within the study scope to replicate a setting in a data scarce environment. While we do agree that these products are more commonly used nowadays, we highlight that they are still not available in many mountainous regions other than the geodetic glacier ice volume change.
- We clarified that the Precipitation lapse rate is primarily based on literature values from previous studies in this catchment and only validated using accumulation data.
- We have addressed the minor comments.

RV2: We have addressed the minor comments

RV3: We have addressed the technical corrections

RV1

The study of von der Esch et al. examines and compares different meteorological data sets for glacio-hydrological modeling in a well-instrumented catchment in Switzerland. In addition, a second experiment investigates the significance of model resolution for reliable model results. Both aspects address key issues in the hydrological investigation of glacierized high mountain areas, namely the quality of forcing products (in situ and gridded) and the compromise between model discretization and computational time. Both aspects are evaluated using a single and multi-data approach. The revised manuscript is well written and structured. The figures and tables in the manuscript and Supplementary Material are well understandable, despite the two different experiments and the large number of different model setups. In addition, the authors have conducted additional sub-studies to strengthen the comprehensibility and argumentation of the presented study.

In my opinion, only one major point and a few minor aspects should be considered.

Major comment:

I recommend adding another subsection about the constraints of the experiments performed with the GERM model for investigating the spatiotemporal precipitation distribution and hydrological response to the discussion. This would be especially beneficial for further studies with a similar research focus and question. This subsection should critically discuss the following aspects:

1. How does the model complexity of GERM impact the spatial representation of hydrological processes? According to my understanding, GERM consists of a distributed snow and glacier model with an integrated lumped runoff routine for routing melt water and liquid precipitation to the catchment outlet. The only spatial heterogeneity considered is the surface properties in the form of land cover. However, no topographical characteristics and drainage network are considered. This is a major constraint in evaluating the sensitivity of spatial model resolution. In the case of fully distributed hydrological models, the spatial resampling to coarser resolutions can lead to significant changes in the flow directions, accumulations, and times in the complex terrain of high-elevation catchments. In the case of the GERM, the resampling only leads to differences in the catchment elevation, as well as the location and number of land cover classes.

We agree that a critical reflection on the structural constraints of GERM would strengthen the discussion and improve the interpretation of the results. We have therefore added a dedicated subsection in the discussion addressing this aspect.

We would like to clarify that GERMs runoff module is not lumped. Instead, runoff is calculated at the grid-cell level using a linear-reservoir routing approach (see e.g. Farinotti et al., 2012). Each grid cell contains several reservoirs (snow, interception, fast, slow and “groundwater”, depending on the surface type), which represent distinct runoff components. Local runoff is computed for each cell and subsequently aggregated at the catchment outlet. This setup allows for a distributed representation of runoff generation. Nevertheless, structural simplifications remain. While land-cover classes are explicitly resolved, GERM does not incorporate drainage networks or topographic flow pathways. This limits the degree to which spatial resampling affects hydrological connectivity compared to fully distributed models that route water explicitly along slopes and channels. In GERM, resampling mainly influences the catchment elevation and the number and distribution of land-cover classes. We now highlight these constrains in the revised discussion.

New section in the discussion: 5.3 Model limitations.

"The experiments presented in this study are subject to structural limitations of GERM. In the model, runoff is routed at the grid-cell level using a linear-reservoir approach. Each cell contains a set of reservoirs that represent different runoff components, depending on the local surface type. The runoff from each grid cell is computed individually and then aggregated at the catchment outlet. This structure allows for a spatially distributed simulation of runoff generation. At the same time, GERM uses some simplifications. Although spatial heterogeneity in the form of land-cover classes is explicitly represented, drainage networks and topographic flow pathways are not resolved. As a result, changes in spatial resolution primarily affect the mean catchment elevation and the distribution of land-cover classes, whereas flow directions, accumulation zones, and travel times remain unaffected. This is in contrast to fully distributed hydrological models, where resampling to coarser resolutions can significantly alter hydrological connectivity and runoff dynamics in complex terrain (Cao et al., 2021; Erdbruegger et al., 2021). In flat terrain, this resampling can markedly affect flow directions (Erdbruegger et al.,2021). In contrast, in steep and spatially limited catchments such as Gletsch, this effect is expected to be minor – particularly at a daily temporal resolution – since most water reaches the outlet within a day. These simplifications do not diminish the value of the experiments, but they need to be considered when interpreting the results and when transferring the findings to other settings. In particular, applications in larger or more topographically complex basins may require model structures that explicitly resolve drainage pathways in order to capture the full sensitivity of the hydrological response to spatial resolution."

2. How representative/suitable is the observed runoff for calibrating the melt parameters? There is not only the uncertainty of the runoff records that limits the suitability for model calibration, as already mentioned in the manuscript. As the runoff is an integrated information of different hydrological storages and processes with temporal lags and complex interdependencies, a process-based calibration and evaluation is needed for reducing the general issue of equifinality. You have considered seasonal mass balances (SMB) and runoff observations for constraining the model parameters in the case of multi-data. However, it is common practice nowadays to include snow observations from different products (in situ and/or from remote sensing) for constraining the snow ablation parameters and evaluating the seasonal evolution. This data is generally more available for other areas and regions than observed SMB data. In addition, the SMB data lacks temporal resolution as they represent aggregated information on snow accumulation and snow/ice ablation.

We agree that using additional process-based observations such as snow cover could in principle help to reduce equifinality and better constrain seasonal dynamics (Schaefli & Huss, 2001; Barandun et al., 2021; Cremona et al., 2025). However, in the present study we deliberately refrained from including snow cover observations. Our aim was to test model behaviour and calibration strategies under data-scarce conditions that resemble many glacierized catchments outside well monitored regions, such as in High Mountain Asia. Snow observations are not consistently available for such settings – geodetic glacier ice volume change, however, is locally and globally accessible (e.g. local: GLAMOS 2024b; global: Hugonnet et al., 2021). We also clarify that we did not use seasonal surface mass balance (SMB) measurements for calibration. Instead, the calibration was based on the geodetic glacier ice volume change derived from multi-annual DEM differencing, which provides a robust and spatially integrated constraint on long-term glacier change. In the single-data calibration, we used such geodetic glacier ice volume changes only, while in the multi data calibration we added runoff observations as a second constraint. SMB measurements were only applied for model evaluation. By focusing on geodetic mass balance and runoff, our calibration strategy aligns with the study objectives: to evaluate how different meteorological forcing products and model resolutions simulate mass balance and runoff, particularly under limited data availability.

To better clarify the above, we added the following to Section 3.5, Line 275: "Although it is becoming increasingly common to also include snow cover or snow depth observations as additional constraints during model calibration (e.g., Schaefli & Huss, 2011; Barandun et al., 2021; Cremona et al., 2025), we deliberately refrained from doing so in this stud, as our aim is to replicate data-scarce conditions. In such settings, consistent and spatially representative snow observations are rarely available, whereas geodetic glacier volume change data are more accessible both locally and globally (e.g., GLAMOS, 2024b; Hugonnet et al., 2021). To maintain methodological consistency with these conditions, we based the model calibration solely on geodetic glacier volume change in the single-data calibration, complemented by runoff observations in the multi-data calibration setup."

3. Is the winter precipitation on the glacier representative of the spatiotemporal precipitation distribution of the entire catchment? As the precipitation lapse rate is estimated based on yearly winter accumulation on the glacier, it may not be representative of the summer (ablation) period. Moreover, it only represents the highest glaciated area of the catchment and neglects the lower-elevation parts of the catchment. This simplification (i.e. linear lapse rate) may overestimate the precipitation in lower elevations or leeward parts of the catchment. Therefore, it is important to consider the spatial precipitation heterogeneity of the precipitation products (when available) in the modeling. As GERM can only consider point meteorological inputs, the aggregation can lead to a reduced spatial variability, as visible in S1 A and B for the MSgrid product.

In our setup, the precipitation lapse rate is primarily derived from literature values and nearby meteorological stations and thus, represents a regional or synoptic-scale gradient rather than a glacier-specific one. The in-situ snow accumulation data on Rhonegletscher were used solely for validation of this gradient, not for its derivation. We agree that this was not clear in the text, and we now clarified this. To account for local-scale variability, the regional lapse rate is combined with a terrain-based redistribution scheme that considers slope and curvature effects, which has been shown to capture accumulation patterns in alpine catchments. This approach ensures that while the large-scale precipitation pattern reflects regional conditions, the finer-scale spatial heterogeneity associated with topography is still represented in the model. Importantly, in our model framework, the ability of the forcing product to capture total precipitation amounts and temporal variability is more influential for model outcomes than the imposed spatial distribution (Figure S1). Where available, higher-resolution products already incorporate spatial variability in their interpolation, and we explicitly tested the effect of upscaling them (Supplementary Figure S1 & S2), finding limited sensitivity of our model results to their internal spatial resolution.

We clarified in Section 3.1 "climate forcing", line 199f: "Precipitation is distributed across the catchment by applying an overall correction factor (C_prec) and an annual fixed precipitation lapse rate (dP/dz) based on regional literature values (e.g. Farinotti et al., 2012) and validate by in situ snow accumulation data over the glaciers elevation range."

Minor comments:

1. As the data quality of meteorological stations is essential for reliable model results, please add some information about the quality control of the Grimsel station data to section 2.2.1. Are there any corrections applied to the time series, especially on the precipitation records?

The Grimsel station data were not further corrected during preprocessing. MeteoSwiss applies quality control but no homogenization before providing the data. In GERM, the time series are adjusted using elevation gradients and calibration parameters. We will clarify this in section 2.2.1.

Line 132: "Furthermore, we did not perform any additional bias correction to the gridded data products or the station data prior to their use as model inputs, in order to emulate a data-scarce environment. Bias adjustments are instead handled internally by the model through its parameters and lapse rates, as described in detail in section 3.1."

Information about the quality control and homogenization of the data from MeteoSwiss: https://www.meteoswiss.admin.ch/climate/climate-change/changes-in-temperature-precipitation-and-sunshine/homogeneous-data-series-since-1864/homogenisation-of-series-of-climatic-measurements.html

2. Is data from the Grimsel station assimilated into the MeteoSwiss grid data? That is essential information for the comparability of the two products.

The Grimsel station is indeed one of the many stations assimilated by MeteoSwiss to generate the gridded MS grid product, as it is based on the interpolation of the available station data. As we aimed to simulate a data-scarce case, none of our applied meteo products were additionally bias-corrected using station data. This reflects conditions in many mountainous regions where high-altitude meteo observations are sparse or absent.

We clarified this in line 132, see comment above. And have added another column in table 2 summarizing the data sources for all applied products.

3. Are there any data gaps in the hydrometeorological observation records? If so, please give information on how these are considered in the model simulations.

Both runoff and meteo time series were continuous and we thus used them in the study without any adjustments. We now clarified this in the manuscript.

Line 128: "With these four continuous data products......"

Line 172: "For catchment runoff, we used continuous daily observations....."

4. Can you please provide more information on how the product-specific monthly temperature lapse rates were computed for the Grimsel station and the ERA5 Reanalysis? Both products only have one elevation information for the study area.

For the ERA5-Reanalysis, monthly temperature lapse rates were derived using the same approach as for ERA5-Land, by applying a linear regression between temperature and elevation across the surrounding grid cells of the study area. For ERA5-Reanalysis we just added additional cells around the cell covering the catchment. For the Grimsel station data, monthly lapse rates were computed through linear regression from nearby meteo stations with a 50 km radius.

We clarified this in section 3.1., line 198: "For each meteorological product, temperature lapse rates were computed as monthly constants by performing a linear regression of air temperature against elevation of grid cells that fall within the catchment. For ERA5-Reanalysis, neighbouring grid cells outside the catchment were also included, since the catchment itself is covered by only a single grid cell. For the Grimsel station data, we used observations from surrounding stations within a 50 km radius to perform the linear regression."

5. Although the glacier routine (deltaH) does not require a spin-up period, the hydrological (linear) storages usually do need a spin-up time to achieve a steady state. According to Table S2, the slow runoff response can be impacted by the initialization up to 80 days. Do you consider the hydrological storages as full or empty at the start of the simulation period?

We initialize the hydrological storages as empty at the start of the simulations. The spin-up time differs by storage type: for the fast reservoirs it is only a few days and therefore negligible for the analysis presented here, while for the groundwater storage it is on the order of a few months to fill the reservoir to equilibrium. To account for this, we discard the first simulation year from the analysis. We now clarified this in the text.

Line 283, section 3.6: Note that, due to the spin-up of the hydrological storages (up to one year for groundwater in this catchment), the first simulation year was discarded from the evaluation."

6. Why is the simulated glacier area in Fig. 5F constant for the 2000 m experiment? Compared to the other spatial resolution, this behavior is quite unexpected.

The constant glacier area at 2000 m resolution is due to the coarse spatial discretization. At such resolution, the glacierized grid cells represent only averaged topography, which tends to shift their mean elevation upward and smooth out valleys that actually host glacier surfaces at a lower elevation. This reduces the representation of low-lying ablation areas and increases the relative share of the accumulation area, resulting in a more balanced mass budget. Consequently, the simulated glacier area remains unrealistically stable at this spatial resolution. We consider this to be a direct effect of coarse grid discretization. We clarified this now in the manuscript.

Line 309: "Furthermore, the initial glacier area in coarse-resolution simulations diverges from observed values by approximately +- 2 km2 (for the 1000 m and 3000 m resolutions) while the glacier area for the 2000 m resolution remains constant."

Line 434: "As a result, the mean glacier elevation shifts upward. One of the effects noticed from this, is the apparent glacier area stability observed for the 2000 m resolution. With an average higher glacier elevation, the low-lying ablation areas are under-represented, and the relative proportion of the accumulation area increases, leading to an unrealistically balance mass budget and a stable glacier extent. "

7. The terms evaporation and evapotranspiration are both used when describing the GERM's model components. In section 3, you introduce the GERM components and mention evapotranspiration. In section 3.4, only evaporation is mentioned. And in the discussion (5.1), you refer to evapotranspiration as a useful point for validation. Table S3 presents simulated average evaporation sums in the Supplementary Material (1.2). Please clarify whether GERM considers transpiration from vegetation in the evaporation module and which component is evaluated in your study.

Thanks for pointing out this inconsistency. In GERM, the evaporation module accounts for evapotranspiration over vegetated surfaces, in addition to evaporation from snow, ice and rock surfaces. In our study, however, we do not evaluate individual components separately but only consider the total evapotranspiration (ET) as it enters the water balance. To avoid confusion, we revised the terminology in the manuscript.

Line 237, section 3.4: "GERM uses a runoff routing scheme that integrates meltwater and rainfall, with evapotranspiration subtracted at each time step….."

8. If possible, please add a table with the simulated mean annual water balance components (i.e. Q, total P, liquid and solid fraction of P, snow melt, ice melt, ET, and storage change) from the different experiments to the Supplementary Material. This would enhance the comparison of absolute values.

We have now added this in the supplementary material in Table S4.

Spelling and typos:

Line 48: space is missing after "Station data" -done

Line 86: change a to the before "well-instrumented Gletsch catchment" -done

Line 100: remove second the -done

Tab.1 Please add the area size (km^2) of the bounding box used for extracting the gridded products in the table or caption - done

Please use a uniform abbreviation for m a.s.l. or m asl. -done

Line 125: the cross-reference is probably to Fig. 2A, E and not to Fig. 3 -corrected

Line 165: remove space after "September"-done

Line 224: move T before "is"-corrected

Line 263: missing .-corrected

Line 284: remove space after "(NSE)" -corrected

Line 329: space after " runoff)."-corrected

Line 352f: remove additional comma and points (i.e.., a delayed onset of melting),, -corrected

Line 428: as the resolutiong resolution -corrected

1.2 of the Supplementary Material: include a space in the table description (Table S3) "evaporationaccording"-corrected

RV2

Dear authors,

All the issues identified in the first draft of the article have now been resolved. The quality of the manuscript has improved significantly. The figures and tables in the new version are clear and easy to understand.

I have just a few comments at this stage:

Comment 1:

I am still surprised by the poor correlation between the simulated and observed flows between January and March (see Figure 7). A sentence should be added to the text regarding this issue.

The seemingly poor correlation between simulated and observed flows during January to March (Figure 7) arises from the very low absolute discharge values in winter. As shown in Figure 6, even small absolute difference between simulated and observed flows translates into large relative differences during this period. While the model reproduces the general low-flow conditions well, the relative performance metrics appear disproportionately poor in winter months. We edited the sentence in Line 333f to make this clearer.

Line 333f: "The seasonal inconsistency in runoff totals is further underscored by the monthly NSE metric (Fig. 7), focusing on April to October when runoff data is sufficiently reliable, as low inter flows introduce uncertainties and even small absolute differences between modelled and simulated can translate into large relative differences."

Minor comments:

Figure 1: Please add the latitude and longitude. – was added to the figure.

Table 3: It would also be helpful to indicate the temperature correction gradient.

Please note the temperatures lapse rates were already provided in the Supplementary Table S1.

Table S3: Please specify the year used for the different products/data.

We added in the caption: "Values are given in mm per year averaged over the modelling period"

RV3

I have received the revised manuscript by von der Esch et al. The authors have thoroughly revised the manuscript and addressed all my major concerns. I have just a few technical suggestions and fully support publication of this manuscript.

We would like to thank the reviewer for the positive appreciation of our work. All suggestions have been taken into account as elaborated below.

Specific comments

Line 14-15: I think that this sentence is still not quite accurate given that the spatial resolution of the input data is not crucial as the authors mention in Section 3. The formulation as in Line 467 would be more helpful here.

We edited the sentence, Line 14-15: "These findings highlight the trade offs between computational efficiency and model reliability, emphasizing the need for high resolution forcing data, particularly precipitation amount and seasonal variability, and careful calibration strategies to capture glacio-hydrological processes accurately."

Line 58: "their" instead of "it's"

We corrected this.

Line 129-130: It might be worth to highlight here that MeteoSwiss is not only high-resolution product but also based directly on observations.

To address this, we have added this information as an additional column in Table 2: "Data sources".

Line 474: maybe "erroneous" instead of "unwanted"? – Corrected.

Line 479: high-resolution forcing based on in-situ observations. – Corrected.