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 two 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 original manuscript. New text parts are highlighted by underlining them. RV1 refers to Reviewer 1 and RV2 refers to Reviewer 2.

Summary of changes:

RV1

- We have considered the use of remote sensing derived meteorological products and now provide an explanation in the manuscript for why we chose not to use them. The main reasons are that no single product provides both temperature and precipitation data; temperature products often represent surface temperature rather than air temperature; and precipitation products are typically associated with high uncertainty and misclassification in mountainous regions.
- We also acknowledge the concern that differences between precipitation datasets are not only related to the data source but also to their spatial resolution and distribution. We now clarify in the manuscript why the spatial resolution of the meteorological product has limited impact on our model results, as we aggregate the gridded data to a catchment-average value for model forcing. To further support this, we tested how the products compare when rescaled to the same resolution (30 km) and found that the catchment-average precipitation does not change significantly (see Fig. S1 & S2).
- We provide the temperature lapse rates used in our model in the supplementary material (see Table S1). We also clarified the approach for computing the lapse rates.
- The constructive, specific edits suggested in the reviewer’s detailed comments have been taken on board.

RV2

- For the importance of the conclusions: We have clarified in the manuscript that the key contribution of our study is to systematically quantify how model performance degrades with reduced data availability and spatial resolution. This is particularly important for applying our model in data-scarce regions, like the Himalayas for example, where data-availability can be limited.
- We addressed the concerns regarding the representation of non-glacierized areas by adding a detailed description of their treatment in the model (Section 3.4). More specifically, we now include information on the routing scheme, storage parameters, and evapotranspiration estimates. We also compared modelled evaporation with historical

observations (Bernath, 1989), these values now being included in a newly proposed Supplementary Table (Table S3).

- We clarified how the spatial distribution of the meteorological forcing is dealt with, particularly explaining how temperature and precipitation are aggregated and redistributed. We now also provide the lapse rates (Table S1) and more detailed descriptions.
- We included the full list of model parameters used in our simulations and referenced an existing study for an analysis of their sensitivity (Farinotti et al., 2012). We also addressed the question of model spin-up by explaining the use of the dh-parameterization used for computing glacier geometry change (Huss et al., 2010). Indeed, the latter enables transient glacier adjustment without requiring any spin-up period (Section 3.3).
- To clarify how forcing data affects model performance, we included new figures in the supplementary material (Figures S3–S5) showing results from uncalibrated runs. These runs help isolate the influence of meteorological products.
- In terms of model evaluation metrics, we now include both NSE and KGE in the results. We adjusted the interpretation of the performance metrics accordingly, and moderated claims where needed. We also clarified why we focus our evaluation on the melt season (April–September) — the period most relevant to glacier hydrology — and revised the relevant figure captions and text (e.g., Figure 7). To highlight potential uncertainties in the runoff measurements, we added a shaded uncertainty range in Figures 6 and 8, based on the assessments of Bernath (1989).
- Further, specific edits suggested in the reviewer's detailed comments have all been taken on board.

RV1

Major comments (MC):

RV1MC1: Choice of the precipitation products for the comparison: *The rationale for selecting exactly these datasets (interpolated gauge-based dataset and two reanalysis ERA5 and ERA5 Land) is not clear to me. Particularly, it is not clear why two reanalysis products are compared, while the satellite and hybrid products are not selected. Moreover, the Section 2.2.1 does not provide any information whether their performance was tested with the in-situ observations in the region. Please revise and clarify.*

We thank the reviewer for this comment. We will add a mention of the satellite derived meteorological products in the introduction, also stating why we didn't consider them for our study (see revised text section below).

Proposed revision line 40-45:

"They are typically generated through interpolation of available weather station measurements (e.g. Dorninger et al., 2008; Frei, 2014), or by estimating the conditions in non-monitored areas with numerical modelling in combination with the observed data from nearby stations (e.g. Muñoz Sabater, 2019; Hersbach et al., 2020). Alternatively, satellite observations can provide remote sensing estimates of precipitation and temperature with broad spatial and temporal coverage. For example, satellite precipitation products from missions such as the Integrated Multi-satellitE Retrievals for GPM (IMERG) (e.g. Huffman et al., 2015) or the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (e.g. Funk et al., 2015) rely on active and passive microwave sensors. However, both gridded climate products and satellite-derived estimates face important limitations in complex mountainous regions. Gridded products often have coarse spatial resolutions (typically 1–30 km or larger), which can lead to significant uncertainties in precipitation estimates due to unresolved orographic effects and local variability in precipitation patterns (Palazzi et al., 2013; Tarasova et al., 2016; Chen et al., 2021; Peña-Guerrero et al., 2022). Similarly, satellite-based products are affected by retrieval uncertainties in high-altitude regions, misclassification of the precipitation phase, and limited ground validation (e.g. Li et al., 2023; Nepal et al., 2024). In addition, satellite temperature products generally provide land surface temperature (e.g., from MODIS, Wan et al. (2006)) rather than near-surface (2_m) air temperature. For these reasons, and because no single satellite product consistently provides both precipitation and temperature variables, we opted not to use satellite-derived climate data as forcing in this study, but only the interpolation and reanalysis products."

Our focus was on comparing commonly used gridded meteorological products which often have varying spatial resolutions and data-generation methods. We chose not to include satellite-only or hybrid products (e.g., IMERG, CHIRPS) for the following reasons: Most remote sensing datasets offer only one of the required meteorological variables — typically precipitation — while near-surface air temperature is generally derived from different platforms, such as MODIS or AIRS. Importantly, there is no single remote sensing dataset that provides both air temperature and precipitation simultaneously and consistently across the time span needed for our model. For this study, using forcing products where both variables originate from the same source (e.g., ERA5 or MeteoSwiss) was a choice to ensure internal consistency and avoid introducing further uncertainty from cross-dataset blending.

Lastly, Reanalysis and regional gridded products are widely used in **glacio**-hydrological studies across various regions (e.g. Naz et al., 2014; Engelhardt et al., 2017; Huss & Hock, 2018; Rounce et al., 2020; Wimberly et al., 2025). Their comprehensive temporal coverage, physical consistency, and widespread availability make them a suitable benchmark for evaluating model sensitivity to meteorological forcing. This choice also enables the broader applicability of our findings to data-sparse regions, where reanalysis products may often be the only viable source of temperature and precipitation.

RV1MC2: Spatial resolution of precipitation: *The narrative of the manuscript indicates that the goal is to investigate the effect of spatial resolution of precipitation input. However, in the experiments it is not only the resolution changes, but also the source of precipitation. In Figure 2 it is clearly visible that datasets are associated with different seasonality of precipitation among interpolated and reanalysis products. Given how different are the sources of precipitation, the effect of spatial resolution cannot be isolated. I think this can be easily fixed by upscaling (i.e., artificially increasing the resolution) of the same product (e.g., interpolated gauge-based precipitation) by several factors.*

We thank the reviewer for this comment regarding the spatial resolution of precipitation and its impact on our study. We acknowledge the concern that not only the data source changes across different precipitation datasets but also the resolution/spatial distribution. However, we would like to clarify why the spatial distribution does not significantly affect our model setup and how we have tested this issue.

GERM does not utilize the distributed spatial information of meteorological data directly. Instead, the meteorological inputs are aggregated to the catchment average and then distributed according to the spatial resolution of GERM. This distribution is applied solely to the meteorological time series using the corresponding lapse rate (see clarification on this below when addressing the specific comment on this). Therefore, in our model setup, the spatial resolution of the input precipitation data itself does not influence the results as much as how well the meteorological product resolves precipitation amount and estimates its timing.

To further address this concern, we performed a test where we artificially upscaled the MeteoSwiss gridded precipitation (1 km) to the 30 km grid resolution of the ERA 5 Reanalysis (will be added to the supplementary material, Figure S1 & S2). The results indicate upscaling the gridded products to the same resolution, in order to isolate the effect of spatial resolution/distribution of the product, does not introduce significant changes in seasonality or precipitation estimates at the aggregated level, supporting our claim that the resolution of the precipitation product does not substantially alter the results in this model setup.

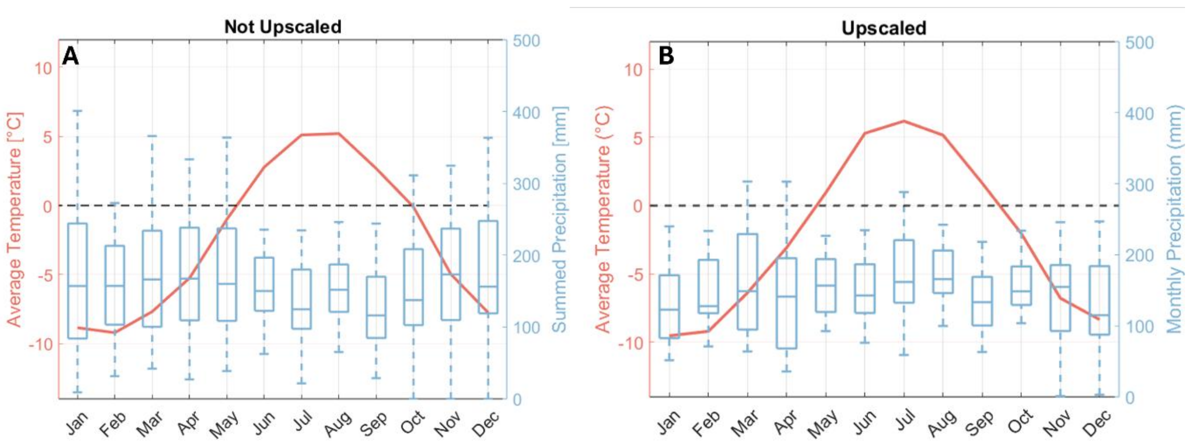


***Figure S1**: Average monthly temperature and precipitation from the MS_{grid} for the period 2000-2022. (A) Temperature and precipitation from the product's original spatial resolution (1 km) aggregated over the catchment. (B) Temperature and precipitation aggregated over the catchment after degrading the product to the 30 km resolution of the coarsest meteorological product used in this study. In both panels temperature was then corrected to the mean catchment elevation using the product-specific monthly constant temperature lapse rate provided in Supplementary Table S1. Precipitation is plotted as the mean catchment precipitation.*

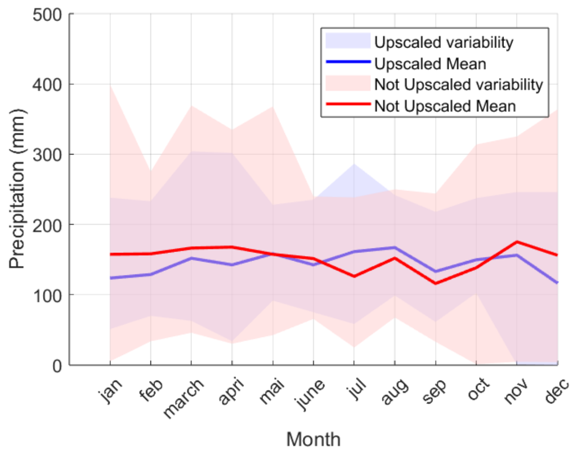


***Figure S2**: Comparison between the mean 2000-2022 precipitation from the MS_{grid} product for both the upscaled (blueish) and not upscaled (reddish) methods. Coloured area shows the variability of precipitation, while the line corresponds to the mean precipitation.*

To clarify we removed part of this sentence, section 3 Methods: “Our workflow (Fig. 4) contains two main experiments performed with GERM. Experiment 1 assesses the impact of the choice of meteorological forcing data on model outputs. To do so, the model is forced using four distinct meteorological products ~~with different spatial resolutions~~, while maintaining a fixed model (GERM) geometry at 25m resolution".

We added this clarification at the end of section 3.1. Climate forcing: “In this setup, the spatial distribution of precipitation within the original product has a limited effect on the catchment-averaged time series applied in the model. This was tested by upscaling the high-resolution products to a coarser resolution prior to extracting the catchment-averaged precipitation time series (cf. Supplementary Figure 1 & 2). Consequently, in our model configuration, the ability of the precipitation product to accurately capture total amounts and temporal variability is of greater importance than its spatial resolution."

Specific comments RV1

RV1 Line 12-13: *At this point in the manuscript, it is not quite clear what is meant here by the constant precipitation adjustment. Please revise and clarify this part.*

We clarified this section by adding "temporally" in line 13:

"Calibrating the model on multi-data, [...] but is limited by temporally constant precipitation adjustments [....]."

RV1 Line 40-45: *It is important to mention here that gridded datasets are not always interpolated products but can also be reanalysis and satellite data.*

Regarding the classification of gridded datasets. We acknowledge that gridded datasets can also be derived from satellite-based products, in addition to interpolation and numerical modelling (reanalysis).

In the manuscript, we have already listed both interpolation-based and reanalysis-based products, as these are the types of datasets used in our study. For completeness, we will add a mention of satellite-based products in this section, while clarifying that they were not included in our analysis.

See revised text in **MC1**

RV1 Line 47: *It might be worth mentioning here the work of Pena-Guerrero et al. 2022 (doi: 10.1002/joc.7548) that compares the performance of different global precipitation products over complex terrain.*

We have now included the work of Peña-Guerrero et al. (2022) in the revised manuscript.

Text edits, line 47: "This introduces uncertainty to the product, especially when estimating precipitation at high altitudes in complex mountainous topography, missing orographic effects, and local variability in precipitation patterns (Palazzi et al., 2013; Tarasova et al., 2016; Chen et al., 2021; Peña-Guerrero et al., 2022)."

RV1 Line 119-120: *Please explain this method in more detail and provide the corresponding reference.*

We have now clarified how the MeteoSwiss gridded product interpolates for temperature.

"We used the gridded MeteoSwiss TabsD and RhiresD datasets. TabsD provides daily mean air temperature at 2 m above the surface, using data from about 90 long-term station series across Switzerland since 1961. The dataset applies a deterministic analysis method for temperature interpolation in high-altitude regions with a spatial resolution of 1 km, capturing daily temperature variations (Frei, 2014). The interpolation procedure combines a two-dimensional lapse-rate regression to represent vertical temperature gradients with a subsequent horizontal interpolation to account for spatial variability (Frei, 2014)."

RV1 Figure 1: *Please explain acronym ELA in the caption*

We have now spelled out the abbreviation **ELA** in the figure caption.

New caption Figure 1:

" [.....] The hypsometry (middle-left panel) represents the distribution of catchment area and glacier area across elevation bands based on data from 2016, <u>with the equilibrium-line altitude (ELA) indicated as dashed black line. The ELA marks the elevation at which annual accumulation equals annual ablation, effectively dividing the glacier into zones of net mass gain and loss.</u> The catchment outline is provided by the Federal Office for the Environment (FOEN)."

RV1 Line 142-145: *Please explain this method in more detail and provide the corresponding reference.*

We have already provided the reference in the original manuscript. We have now edited the text to make the method clearer.

Revised text: "For model calibration, we relied on geodetically-derived glacier <u>mass loss</u> change between 2013 and 2021. <u>The geodetic mass loss was determined by differentiating two high-resolution</u> DEMs for Rhonegletscher acquired by dedicated monitoring flights on 21 Aug. 2013 and 20 Aug. 2021 (GLAMOS, 2024b). <u>The resulting</u> ice volume change of –0.1354 km3 was found for the respective time period referring to the main glacier in the catchment (Rhonegletscher). The ice volume change was converted to a mass change by assuming a density of volume change of 850 kg m−3 (Huss, 2013)"

RV1 Line 151: *Please explain how the extrapolation is done.*

We have edited the text to clarify the extrapolation procedure.

"To evaluate model results, we used annual and seasonal glacier-wide mass balance measurements for Rhonegletscher, covering the period 2007–2024 (GLAMOS, 2024a). This data is based on spatially distributed in-situ measurements <u>of snow accumulation and ice melt</u> across the entire glacier surface both in late April and September. Winter snow observations from <u>150</u> up to 300 snow-sounding locations were converted to water equivalent using snow density measurements. Measurements <u>of local annual mass balance</u> at a network of 10 ablation stakes <u>were extrapolated to the entire glacier surface with a model-based approach (Huss et al., 2021). Herein, a daily distributed mass balance model is optimized to match all point observations of winter and annual mass balance and thus extrapolates to unmeasured regions based on calibrated physical relations. Furthermore, the utilized approach provides a homogenization of arbitrary measurement dates to the fixed dates of the hydrological year. The so-obtained data set thus allows for straight-forward comparison to model results acquired in the present study.</u>"

RV1 Line 175-180: *Please clarify how the lapse rates are computed and whether or not they are recomputed for different spatial resolutions. Please provide the estimates.*

We thank the reviewer for the valuable comment. We have clarified the methodology for deriving and applying lapse rates in the manuscript (section 3.1 Climate forcing).

Text revisions (line 188 following):

"GERM is driven by a point time series of temperature and precipitation, either near or within the catchment area, which are subsequently distributed across the catchment using a monthly-averaged temperature lapse rate (cf. Supplementary Table S1) and a constant precipitation lapse rate to every grid cell at the specified model resolution. 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. These monthly lapse rates were then used to downscale the temperature time series across the model domain. Precipitation is distributed across the catchment by applying an overall correction factor (C_prec) and an annually fixed precipitation lapse rate (dP/dz) generally derived from in situ snow accumulation data over the glacier's elevation range, as well as literature values (e.g. Farinotti et al., 2012). For capturing the small-scale spatial variability of snow accumulation, a distribution matrix derived from terrain characteristics (slope and curvature) is superimposed on spatialized precipitation (Huss et al., 2008a). "

Figure 2 caption correction: "[......] Temperature and precipitation of the gridded products were spatially averaged over the catchment. Temperature was then corrected to the mean catchment elevation using a product-specific monthly constant temperature lapse rate (cf. Supplementary Table S1) while precipitation is given as the mean catchment precipitation. For the box plots, the 22-year daily precipitation series was aggregated to mean monthly sums."

***Table S1**: Applied monthly temperature lapse rates (in °C per 100m of elevation; kept constant over the entire modeling period) for each meteorological product applied in this study. MS_grid refers to the gridded product of MeteoSwiss. The sequence of months reflects the hydrological year. The lapse rate for the Grimsel station data was obtained based on surrounding meteorological stations.*

Month	Grimsel	MS_grid	ERA5-Land	ERA5-Reanalysis
October	-0.52	-0.47	-0.44	–0.41
November	-0.53	-0.45	-0.42	-0.39
December	-0.60	-0.43	-0.41	-0.38
January	-0.64	-0.43	-0.42	-0.37
February	-0.65	-0.44	-0.42	-0.38
March	-0.65	-0.49	-0.45	-0.41
April	-0.65	-0.52	-0.48	-0.43
May	-0.62	-0.53	-0.48	-0.44
June	-0.59	-0.55	-0.49	-0.45
July	-0.56	-0.55	-0.5	-0.46
August	-0.53	-0.54	-0.48	-0.44
September	-0.56	-0.51	-0.45	-0.41

***RV1 Line 183:** It is not clear how this is done. Please clarify.*

We have clarified it in the text as mentioned in the reply to MC2 and the specific comment to line *Line 175-180*

***RV1 Line 228:** It is not clear why precipitation correction factor represents accumulation parameter. Please clarify.*

We thank the reviewer for pointing this out. We agree that the terminology could have been better clarified. In our model setup, the precipitation correction factor (C_prec) directly influences the total precipitation input, including both liquid and solid components. Since snow accumulation in the model is entirely driven by solid precipitation, scaling total precipitation with C_prec also scales the snow accumulation accordingly.

To avoid confusion, we will no longer refer to C_prec as an "accumulation parameter" and instead consistently refer to it as the *precipitation correction factor*. However, we clarify in the revised text that its role in controlling accumulation arises from its direct influence on solid precipitation, which drives accumulation in the model.

Revised text (Line 228): *"At the same time, the precipitation correction factor (C_prec) is optimized within bounds of [0.6, 1.5]. C_prec is a constant parameter that adjusts the daily* *catchment precipitation—both liquid and solid—by a fixed percentage, thereby increasing or decreasing it uniformly over the modeling period. Since accumulation in GERM is entirely determined by solid precipitation, and C_prec directly scales this input, it effectively also controls the magnitude of accumulation in the model*."

RV1 Table 3: *Please clarify if these are best calibrated parameters.*

Yes, the values shown in Table 3 represent the final, best-calibrated parameter sets resulting from the respective calibration procedures (single-data and multi-data) for each forcing product and model resolution. We have clarified this in the manuscript and table caption.

Table 3 heading: "*Single- and multi- data calibration: Final best-calibrated parameter values from the single- and multi-data calibration for each Experiment 1 (top) Experiment 2 (bottom). [.....]"*

RV2

General comment: *Von der Esch et al. present an important and interesting work in terms of modelling, which aims to simulate the glaciological and hydrological functioning of a catchment area of 39.4 km2, of which 16.7 km2 (44%) are glaciated. However, both the novelty and the relevance of their conclusions are not immediately obvious. The conclusions that the model simulates the runoff better when it is calibrated against this runoff, that reducing the model resolution reduces its capacity, and that a model resolution should be adapted to the size of the simulated object seem so trivial that more information is needed to convince the reader that this is not the case. To improve the manuscript, some important issues need to be addressed (Major comments 1, 2 and 3), and some specific comments should be considered (see below).*

We thank the reviewer for this important feedback. We acknowledge that some of the conclusions—such as the model performing better when calibrated with runoff data, and reduced performance at coarser spatial resolutions—may appear intuitive at first glance. However, the primary objective of our study was to assess these expected outcomes in a controlled, well-instrumented environment as to rigorously test the Glacier Evolution Runoff Model (GERM) under varying data availability and model setups. This was essential groundwork for our intended application of the model in data-scarce regions like the Himalayas.

The novelty lies in systematically quantifying the magnitude of performance loss under reduced data and resolution conditions, using a model specifically designed for glaciated catchments. This includes, for example, understanding (i) to what extent omitting runoff data affects seasonal dynamics, (ii) at which resolution critical glacier and topographic features become underrepresented, or (iii) how the forcing influences the model results. These insights are vital for informing model applications in regions where high-resolution and high-quality input data and runoff observations are not available—a situation common in many high-mountain areas globally.

While the broader motivation for applying the model to remote regions was already mentioned in the manuscript, we realize that the practical motivation behind the design of our experiments—namely to simulate the limitations we would encounter in the Himalayas—could be made more explicit. We have now revised the introduction to better communicate this motivation and the broader relevance of our findings.

Text edit line 75-80: "By using a catchment with robust data availability, we aim to assess how these modelling choices perform in a controlled setting and to provide insights relevant for data-limited, high-altitude regions. <u>While the experiments are conducted in a well-instrumented Alpine catchment, the design of this study reflects the limitations commonly encountered in remote regions, such as the Himalayan Mountain range for example. Understanding how the model performance is affected by the absence of high-resolution input or runoff data, and systematically quantifying the magnitude of performance loss, is crucial for evaluating the reliability of glacio-hydrological models under such constraints, especially when applied in ungauged or poorly monitored environments."</u>

Major comments:

RV2 MC 1*: Ability of the model to reproduce the hydro-glaciological functioning of the catchment*

Since the model used here is a glacio-hydrological model, and that less than 50% of the simulated catchment is glacierized, hydrological conditions simulated for the non-glacierized part of the catchment are important on a daily time scale.

1. *Non glacierized part of the model*

- *The description of the model, how it works and how it is calibrated is completely lacking for this non-glaciarized part. For example, what are the runoff coefficients chosen, how is the subterranean compartment considered, etc.... This can be important in term of hydrological functioning, particularly during summer rainfall events or during low flows periods.*

We have added a description in Section 3.4, explicitly stating that non-glacierized surfaces (e.g., rock, vegetation, snow-covered areas) are included in GERM using a reservoir-type routing scheme. These components use fixed storage and retention parameters, following the conceptual structure originally described in Huss et al. (2008) and Farinotti et al. (2012). These parameters are not optimized or calibrated separately, as the model focuses primarily on glacier-related processes. The parameters used for these reservoirs are derived from previous applications of the model referred to above and are included in a new supplementary table (Table S3), along with a short description of their physical meaning. An extended sensitivity analysis of these parameters was already performed in Farinotti et al. (2012), and while we do not see a need of repeating that analysis, we now make reference to its key findings in the method description (Section 3.4) and discussion (Section 5.1). The amended text blocks read as follows:

Text revisions section 3.4: "GERM uses a runoff routing scheme that integrates meltwater and rainfall, with evaporation subtracted at each time step (see Farinotti et al., 2012, for a detailed description of this model component). The scheme is structured around the concept of linear reservoirs (Langbein, 1958) and simulates the water balance of every grid cell and time step across diverse surface types—including ice, snow, rock, vegetation, and groundwater—by routing water through type-specific reservoirs with fixed retention constants. Each land surface type is assigned to a reservoir and associated with specific fixed retention and storage parameters, originally described in Huss et al. (2008) and Farinotti et al. (2012). These parameters are not calibrated in this study but are based on validated applications of GERM to similar catchments, including the Gletsch basin (e.g. Huss et al., 2010; Farinotti et al., 2012). A detailed list of the parameter values used is provided in Supplementary Table S3. This representation captures both rapid surface runoff and delayed subsurface flow components, which are particularly relevant during summer rainfall events and low-flow conditions. The total discharge is obtained by summing the outflows from all reservoirs at the catchment level, enabling a fully distributed, partitioned hydrograph simulation (Farinotti et al., 2012)."

Text revision section 5.1 line 335 following:" In line with the finding that meteorological variables are the main source of uncertainty, the parameter sensitivity analysis of GERM by Farinotti et al. (2012) in the Gletsch catchment showed that constant retention and storage capacity parameters have a relatively minor impact compared to temperature lapse rate, precipitation correction, and ablation parameters. This justifies the decision not to calibrate reservoir-specific parameters individually, as previously described. Instead, calibration efforts are best focused on accurately estimating temperature gradients and ablation dynamics, which contribute most significantly to uncertainty in runoff projections."

- *In addition, evaporation is low in such a mountainous environment, except in summer when it reduces the contribution of precipitation to runoff. How are the meteorological forcings applied to this part of the catchment, and how they differ from the glacier model part? How do these forcings compare with local observations (e.g. André Bernath has made precipitation and evaporation measurements in this catchment; and the Hydrological Atlas gives an estimate of the evaporation term)?*

We agree that evaporation plays an important role in shaping runoff, particularly during summer. To address this, we compared modelled annual average evaporation values with historical measurements by Bernath (1989), which are now provided in the supplementary material (Table S3). The study by Bernath (1989) focused on the water balance of the Gletsch catchment, among other catchments, in the Swiss Central Alps. It provides a particularly relevant comparison for our work because it includes detailed, independent measurements of precipitation, evaporation, and discharge over several years (1979–1983) in the Gletsch catchment. Our modelled evaporation values (Table S3) are in good agreement with Bernath's estimate of 131-240 mm/yr. To reflect this comparison, we added a short discussion in the text:

Revised text Section 5.1, line 343 following: "<u>However, the model's representation of evapotranspiration provides a useful point of validation. While evapotranspiration plays a relatively small role in this high-alpine environment, it becomes relevant during summer in non-glacierized areas. Modelled annual evapotranspiration values (173–206 mm/year) are consistent with the historical range of 131–240 mm/year reported by Bernath (1989) (Table S3), indicating that this process is well represented. This suggests that the main sources of uncertainty in summer runoff simulations are not due to evapotranspiration losses, but rather arise from reservoirs more directly affected by meteorological forcing—such as glacier and snow components—which are also more sensitive to calibration parameters</u>"

Table S3: *Estimated evapotranspiration based on measured summer and estimated winter evapotranspiration from Bernath (1989) and average modelled annual evapotranspiration in GERM for each applied meteorological forcing. Values are given in mm per year.*

Bernath (1989)	Grimsel	MS_grid	ERA5-Land	ERA5
131-240	179.5	173.1	181.9	206.5

We also updated the manuscript to clarify that a physically-informed approach is used to distribute the meteorological time series across the catchment. More specifically, temperature is distributed by using a monthly temperature lapse rate which is specific to each meteorological product (the lapse rates are now included in Table S1). For precipitation, we follow the method described in Huss et al. (2008b), which involves applying a constant correction factor to the catchment-mean time series and includes an altitudinal precipitation gradient and a spatial distribution matrix for solid precipitation based on topographic characteristics (slope and curvature). Based on this approach, also the meteorological forcing applied to non-glacierized areas is dependent on the topographic characteristics.

Text revisions (line 188 following):

"GERM is driven by a point time series of temperature and precipitation, either near or within the catchment area, which are subsequently distributed across the catchment using a monthly-averaged temperature lapse rate (cf. Supplementary Table S1) and a constant precipitation lapse rate to every grid cell at the specified model resolution. 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. These monthly lapse rates were then used to downscale the temperature time series across the model domain. Precipitation is distributed across the catchment by applying an overall correction factor (C_prec) and an annually fixed precipitation lapse rate (dP/dz) generally derived from in situ snow accumulation data over the glacier's elevation range, as well as literature values (e.g. Farinotti et al., 2012). For capturing the small-scale spatial variability of snow accumulation, a distribution matrix derived from terrain characteristics (slope and curvature) is superimposed on spatialized precipitation (Huss et al., 2008a). "

Figure 2 caption correction: "[......] Temperature and precipitation of the gridded products were spatially averaged over the catchment. Temperature was then corrected to the mean catchment elevation using a product-specific monthly constant temperature lapse rate (cf. Supplementary Table S1) while precipitation is given as the mean catchment precipitation. For the box plots, the 22-year daily precipitation series was aggregated to mean monthly sums."

- *Finally, this non-glacial part will have an impact on the separation of the types of flow (surface, underground, ice melt and snow melt). For the moment this is noted on lines 284 to 286 so there is a need to provide much more information.*

In our study, we primarily focus on the snow melt and ice melt components of the catchment's hydrology, which are represented as direct outflow within the model's reservoir-based routing framework (Farinotti et al., 2012). According to this scheme, these components are routed through dedicated reservoirs. This structure ensures a one-way routing configuration, where snow and ice melt contributions are passed directly to the catchment outlet without additional modification from subsurface or slower flow components. We acknowledge that this aspect was not sufficiently clear in the original manuscript and have revised Section 3.4 to provide more clarity. See previously mentioned text revision on Section 3.4

1. ***Glacierized part of the model***

The model chosen is a good choice as well as the methodology for investigating the sensitivity to the resolution of the input meteorological data and the multi-objective calibration based on mass balances and flow rates.

- *However, many parameters are not detailed and are not evaluated through a sensitivity study. This is the case for temperature and precipitation lapse rates (see also the next comment). The values of all the parameters should be given and the sensitivity tests carried out should be indicated, showing the ranges of consecutive values for simulated mass balances and flow rates.*

We thank the reviewer for the appreciative comment about our model choice and our methodological design, and we agree that the temperature and precipitation lapse rates, among other parameters, are crucial for the glacio-hydrological model performance. While a dedicated sensitivity study was not performed in the current work, we draw on the extensive analysis presented in Farinotti et al. (2012), which used the same model framework across several high-Alpine catchments, including the Rhone Glacier. In that study, lapse rates, ablation parameters, as well as other model parameters were systematically varied in a factorial experiment, and the influence on both mean annual runoff and model performance was quantified.

To better inform readers of the above, we now include a paragraph summarizing the findings of Farinotti et al. (2012). Similarly, we now included Supplementary Table S2, listing the key parameter values used in our simulations. The summarizing paragraph is found, in the Discussion section and reads:

Text revision section 5.1 line 335 following: <u>"In line with the finding that meteorological variables are the main source of uncertainty, the parameter sensitivity analysis of GERM by Farinotti et al. (2012) in the Gletsch catchment showed that constant retention and storage capacity parameters have a relatively minor impact compared to temperature lapse rate, precipitation correction, and ablation parameters. This justifies the decision not to calibrate reservoir-specific parameters individually. Instead, calibration efforts are best focused on accurately estimating temperature gradients and ablation dynamics, which contribute most significantly to uncertainty in runoff projections."</u>

- *It appears that no spin-up was performed to bring the Rhone Glacier into equilibrium with the simulated mass balance (since all simulations started with the same area: Fig.5C). As the annual mass balance varies between simulations, part of the area change (Fig.5) is due to the initial imbalance. The simulated daily discharge is mostly a function of the daily melt rate applied to the glacier surface. Since the Rhône glacier has a time response of several decades, its surface area (and volume) is due to the initial simulation conditions and not to the prescribed accumulation rate, unless a long spin-up run has been applied to equilibrate the glacier with the prescribed forcing. This problem is mentioned very briefly (pp. 348-349) but not discussed.*

We thank the reviewer for raising this important point regarding glacier equilibrium and the potential need for a spin-up to balance the glacier geometry with the applied climate forcing.

While we acknowledge that spin-up procedures are common in glacier modeling, we did not perform such spin-up in our simulations for the Rhone Glacier. This is because of the structure of our modeling framework, which relies on the so-called dh-parameterization (Huss et al., 2010) for updating the glacier geometry. In a nutshell, this method imprints observed elevation-change patterns on the annual glacier geometry, by honouring mass conservation and the glacier mass changes computed by the model's mass balance module. This enables the glacier to dynamically respond to the annual mass balance forcing without the need for long-term equilibration runs.

Huss et al. (2010) demonstrated that even over multi-decadal time scales, the dh-parameterization closely reproduces the results of a 3-D finite element ice flow model in terms of both glacier area and surface elevation changes. Their validation specifically included Rhone Glacier (see their Fig. 7–9) and shows that despite no explicit spin-up, the parameterized model can accurately reflect long-term glacier evolution under changing climatic conditions.

We have now clarified this in the method section 3.3

Text edits: " Glacier geometry and area are updated annually using the dh-parameterization (Huss et al., 2010). It approximates changes in glacier surface elevation and glacier area in response to annual mass balance. This empirical approach redistributes net mass changes across the glacier based on a normalized elevation-dependent function (dh) derived from observed surface elevation changes in the past. The parameterization is mass-conserving and reflects typical glacier behavior, producing the largest and smallest elevation changes in the ablation and accumulation area, respectively. It adjusts the glacier extent by removing glacier sections where the surface elevation falls below the bedrock. Albeit the dh-parameterization does not explicitly simulate dynamic processes, it has been shown to closely replicate the results of a 3-D finite element flow model in terms of glacier volume, length, and area evolution over decadal scales (Huss et al., 2010). Since the dh-approach allows the glacier to transiently adjust to the imposed climate forcing as an immediate response, no spin-up time was applied in our simulations. "

- *For the whole model, the modelling strategy for calibration and validation (or evaluation) is not well explained. There are numerous methods of data set selection (e.g. split sample tests).*

We agree that some additional explanation can be helpful to understand our work even better. In our study, we chose a fixed calibration period (2013–2021), based on the availability of precise geodetic glacier volume change data. This period allowed us to calibrate the model using spatially integrated glacier mass balance information. We then evaluated the model over the full simulation period (2000–2022) to test its performance under varying climate conditions, including years outside the calibration range. We have clarified this in the method section 3.5.

Suggested Addition to Section 3.5 at the end: "In this study, model calibration was performed over the period 2013–2021, which aligns with the availability of high-resolution geodetic glacier volume change data. This period serves as the calibration window for both the single- and multi-data calibration approaches. Model evaluation was then conducted over the full simulation period (2000–2022), allowing assessment of long-term model performance, seasonal variability, and year-to-year consistency. This fixed calibration–evaluation approach was selected to maintain consistency across experiments."

RV2 MC 2. *Impact of meteorological forcing and spatial model resolution on the accuracy of glacio-hydrological simulations*

A first question is what is meant by "accuracy" or "reliability" of a simulation? This depends entirely on the context. For operational forecasting of e.g. hydropower, these daily simulations are far too coarse, whereas for centennial simulations even the weakest resolution is sufficient (since the annual mass balance is correct).

We agree that model accuracy is context-dependent. In our study, accuracy refers to how well the model reproduces observed glacier mass balance (annual and seasonal) and runoff over a historical period. We clarified this in section 3.6

In line 241 following: "In our study, model accuracy refers to how well simulated glacier mass balance (annual and seasonal) and catchment runoff match corresponding observations over effect of single versus multi-data calibration <u>on the accuracy of the</u> model results, we evaluate the simulated glacier mass balance and runoff against observational data for both."

As raised in the previous comment we can ask the following question: are really the meteorological forcing and the spatial resolution responsible of the accuracy differences among simulations? Objectively, the Grimsel and MSgrid meteorological series are more accurate than the ERA5 at 30km resolution. Objectively, the 25m resolution model describes much more accurately the catchment than the 3km model. However, the meteorological series have been independently corrected, and the model calibrated differently for each setup, so that the link between each forcing or resolution and the corresponding simulation is not obvious. Especially, the elevation correction applied to precipitation is crucial. It is well known that lapse rates are not consistent in the Alps. Hence, the basis and magnitude of these corrections, their interplay with the model Cprec, are important questions here. Further, the precipitation correction factor, Cprec, is exactly 1 for simulations with a varying resolution (Table 3: 100-1000m), and much lower than 1 for lower resolutions: this seems at odd with precipitation being too low compared to glacier accumulation (as it is generally noticed). In fact, Fig.5B-E shows winter accumulation of 2m, hence an annual rate of precipitation of more than 3m, not found in the precipitation products (Fig.2). Even in Switzerland, which has the best observational network and the best knowledge, the question of snow measurements underestimation has been in debate for decades (the Boris Sevruk version of the Swiss precipitation Atlas had a correction by +20-30%, whereas the more recent Ch. Frei version has not.). Also, looking at Figs.6 to 8it is not obvious that the objectively more accurate forcings and resolutions lead to 'more accurate' simulations?

So, some clarification is required on the magnitude of the precipitation correction, and how corrected precipitation compares with estimates. (The Gletsch catchment has been extensively studied, see Bernath 1989; Klok et al. and references therein). Some clarification is also required to understand how a 3km-resolution catchment could 'work so well', indeed. Especially, Fig.3 shows that the area of the catchment varies with its resolution, so that simulated and observed runoffs should not compare in absolute unit (in m3/s; as in Figures 6 and 8), but only in specific unit (mm/d). Some correction of the area has been obviously done?

We thank the reviewer for this important comment. To isolate the effect of the meteorological forcing itself, we have added figures in the supplementary material with uncalibrated model runs. In this setting, "uncalibrated" means that we used the parameter combination obtained from a default run (in this case the run in which the model is forced with the Grimsel station, which - as the reviewer correctly noticed - can be considered to be the most accurate or nearby meteorological information available for our study area) to all other model runs too, no matter the forcing product.

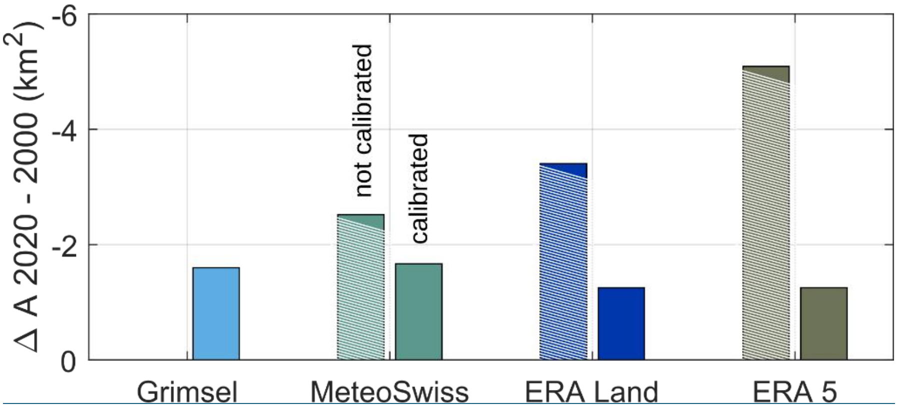


Figure S3: *Effect of the forcing products considered in Experiment 1 on the glacier area change simulated for the period 2000-2020 when the model is (right bars) or is not (left bars) calibrated to glacio-hydrological observations . For this sensitivity analysis, "Grimsel" is the default, meaning that only the calibrated results exist.*

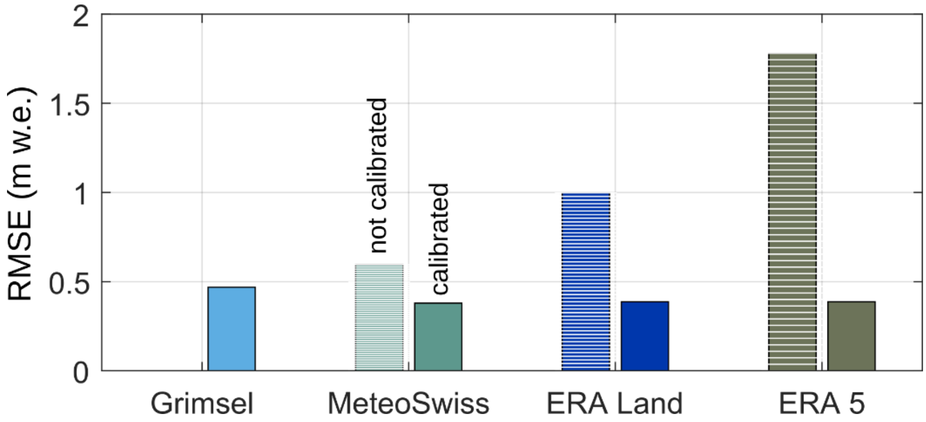


Figure S4: *Same as Figure S3 but showing the root-mean-squared-error (RMSE) between modeled and observed glacier-wide annual glacier mass balance in the period 2007–2022.*

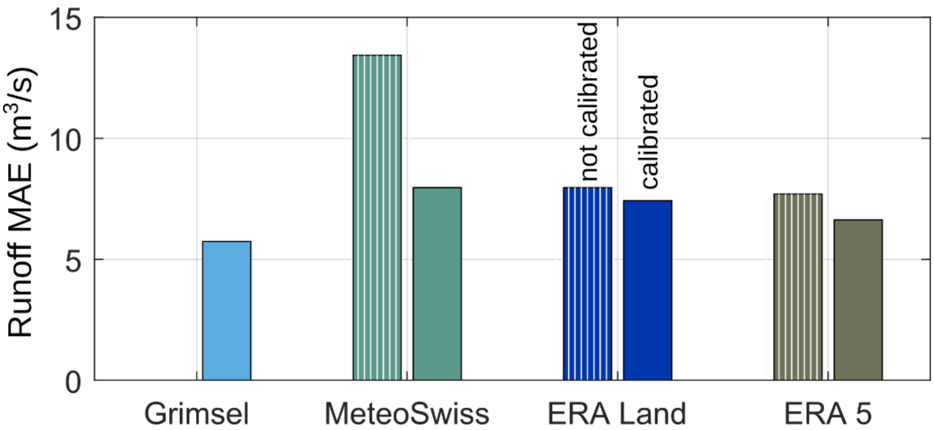


Figure S5: *Same as Figure S3 but showing the mean absolute error (MAE) in modeled annual runoff for the period 2000–2022.*

These figures show that without calibration, the deviations of the simulated variables from the observed would be even larger. After calibration, the differences between each simulation become smaller, but still exist, which implies that the resulting difference in model performance must stem from the difference in applied meteorological products.

Furthermore, we clarify (see our reply to the reviewer's MC1) that in our model, precipitation increases linearly with elevation and small-scale accumulation variability is accounted for based on topographical indices (curvature and slope). In coarser resolution model runs, the glacier area is often shifted to higher (and thus colder) elevations because of the spatial aggregation. Since more precipitation is then falling as snow than it would at a lower elevation, the precipitation that is needed to achieve a similar mass balance is lowered, effectively resulting in smaller values for the parameter "Cprec". This can, as correctly noted by the reviewer, lead to a mismatch with observed accumulation and thus to an underestimation of annual runoff, while still capturing the seasonal runoff pattern. To clarify this, we added the following in the revised manuscript:

Text revision at the end of Section 5.1.: "To further isolate the impact of the meteorological forcing, we conducted additional model runs without re-calibrating model parameters to each forcing product. The results (Figures S3–S5 in the supplementary material) show that in the absence of calibration, the deviations between modelled and observed glacier area, mass balance, and runoff are even larger. Calibration reduces these differences but does not eliminate them, confirming that the choice of meteorological forcing product remains a primary driver of model performance"

For what the units of Figures 6 and 8 are concerned, we agree that comparing absolute runoff volumes (m^3/s) across resolutions can be misleading. We now corrected this by calculating specific runoff amounts for the respective catchment area (in units of mm/a) and updated the figures accordingly.

RV2 MC 3. Uncertainty on runoff measurements and Nash-Sutcliffe criterion choice

The caption to figure 7 states that "...the grey shaded areas indicate the months considered in this study", but this fact is not specified in the text. This choice to evaluate only the summer months is highly questionable and more details are needed.

We thank the reviewer for the comment. We have revised both the figure caption and the corresponding section to clarify and better justify our focus on the melt season (April–September). This period was chosen because it is when snow- and glacier-melt dominated processes are most active, making it particularly relevant for assessing model performance in the context of our study. While winter runoff data are also affected by higher uncertainty due to low flows and the practical difficulties in measuring them (Alpine streams can then be partially covered in ice and snow), our primary motivation for selecting the melt season is to evaluate the model's ability to capture runoff dynamics driven by snow and glacier melt - in line with our research objectives.

Revised Caption for Figure 7: "(A, B) Monthly NSE values for each experiment. Grey-shaded areas indicate the melt season (April–September), which is considered for model evaluation. This period aligns with the time of year when glacier- and snowmelt-driven runoff dominates. Winter runoff values are excluded due to both their high uncertainty and their limited contribution to annual discharge. (C) CV of the annual runoff sums [.....]."

Revised Sentence in the Text (addition to line 244): The monthly Nash–Sutcliffe efficiency (NSE) and monthly relative difference (%) are used to quantify the agreement between observed and simulated runoff and capture seasonal variations. This model evaluation focuses on the melt season (April–September), when snow and glacier melt dominate the hydrological response.

This period is most relevant to our study objectives, which center on glacier-influenced hydrology. Winter runoff is excluded due to its limited relevance and higher associated uncertainty from low flows.

The choice of the Nash-S parameter to evaluate the model is highly controversial. The study by Althoff and Rodrigues, JoH, 2021, shows that this coefficient should be avoided. Other options exist, such as the KGE. Could you please provide other metrics to evaluate the model?

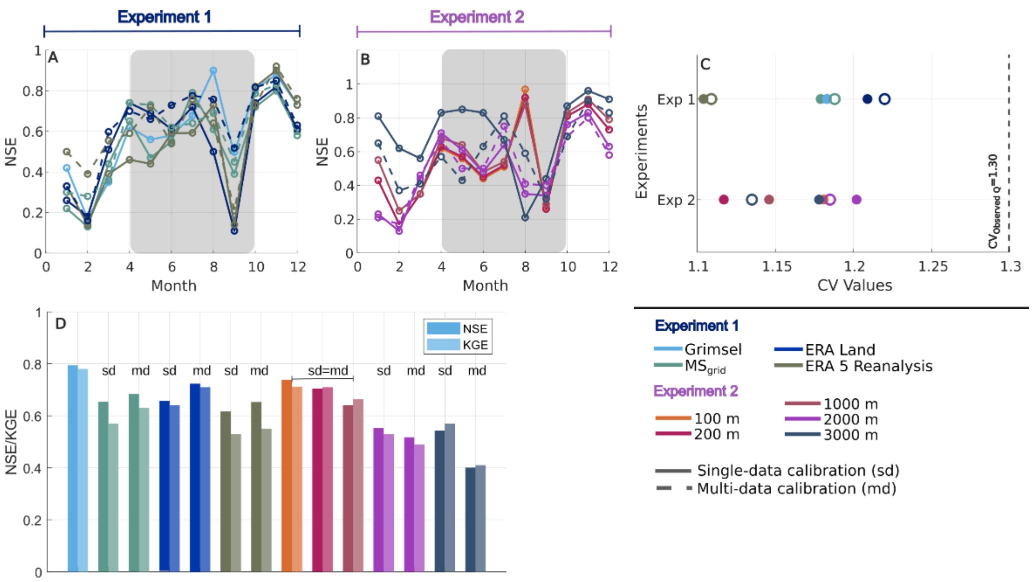
We appreciate the reviewers comment and acknowledge the ongoing discussion regarding the limitations of the NSE as a model evaluation metric. We now also included the KGE alongside NSE to provide a more balanced view on model performance. This resulted in an addition of a plot to Figure 7 and minor additions in the Figure caption, Section 3.6 and in the Results and Discussion.

Text revision section 3.6 (line 241 following): “The runoff simulations are assessed against measured daily catchment runoff at Gletsch over the period 2000–2022, while glacier mass balance is evaluated using annual and seasonal measurements spanning 2007–2022. The monthly and annual Nash–Sutcliffe efficiency (NSE), annual Kling–Gupta efficiency (KGE) and monthly relative difference (%) are used to quantify the agreement between observed and simulated runoff and capture seasonal variations.”

Text revision in section 4.2: “Ultimately, across seasonal and annual scales, the single-data calibration consistently underestimated runoff (Figure 8). Similarly, the KGE values reflect a comparable fit, with the annual values indicating a bias primarily driven by the systematic underestimation of runoff (Figure 7D).”

Updated Figure 7 caption: “[....] (D) Annual NSE (left bar) and KGE (right bar) for each simulation. sd indicates a simulation performed with the single-data calibration, md with the multi-data calibration. [...] “

Updated Figure 7:



Text revisions Section 5.1 after the new addition from MC1: "Similar to NSE, KGE values are highest for simulations forced with Grimsel data and decline when using gridded meteorological products (Figure 7D). Multi-data calibration generally improves KGE values across all simulations. However, the ranking between forcing products remains consistent with that seen for NSE, reinforcing the conclusion that meteorological forcing quality impacts the reliability of runoff simulations."

Text revisions Section 5.2, line 361 following: "The observed strong decline in KGE (Figure 7D) values with coarser spatial model resolution, i.e. larger than 100 m, supports this interpretation. While NSE primarily reflects timing and shape agreement, KGE additionally penalizes deviations in runoff magnitude and variability. Thus, the decreasing KGE at coarser resolutions emphasizes that errors in total runoff volumes increase as spatial detail is lost."

Lines 306 to 309, it is written: 'For most resolutions (except 3000 m), the NSE decreases from April to June, probably due to delayed runoff timing, ...' This conclusion is questionable as the model is run at a daily time step. An hourly time step should be used to draw this conclusion.

We agree that sub-daily runoff patterns are characteristic for glacierized catchments, and that an hourly model time step would be necessary to analyze diurnal runoff dynamics such as the timing of daily peaks. However, our statement was not intended to imply such sub-daily behavior. Instead, we were referring to a seasonal shift in runoff timing, which can still affect monthly NSE values, even when the model is run at a daily time step.

To avoid confusion, we revised the sentence as follow:

lines 306–309: "For most resolutions (except 3000 m), NSE declines from April to June, likely due to a seasonal shift in runoff timing (i.e., a delayed onset of melting), then improves markedly from June to August before dropping again in September."

To compare simulated and measured runoff, the uncertainty on measurements should be accounted for. Measuring runoff in this highly variable environment is difficult. Also, the question of a potential water underflow not measured at the Gletsch gauge station was discussed by Bernath (1989).

We added a grey shaded uncertainty range to the observed runoff in Figure 6 and 8. The shading is based on Bernath (1989), who quantified the relative random error in water level measurements, considering instrument precision and natural fluctuations such as wave effects. Based on this study, we applied a ±0.9% uncertainty range to the observed runoff, as it provides a suitable estimate for measurement-based runoff errors in our setting.

Finally, line 340 rightly mentions the concept and definition of equifinality, and this principle should guide this study by testing most of the parameters.

As mentioned in our reply to the reviewer's MC1, an extensive parameter sensitivity analysis was already performed in Farinotti et al. (2012). We therefore only include a supplementary table listing the key parameter values used in our simulations and paraphrase the findings of Farinotti et al. (2012) in our discussion. For the revised text, see our answer to MC1.

Specific comments:

RV2 SP1: *daily time scale needs to be specified more clearly (abstract, introduction, etc...)*

L4-6: “This study assesses the reliability of glacio-hydrological simulations in a glacierized catchment (39.4 km2) in Switzerland using the Glacier Evolution Runoff Model (GERM) at daily temporal resolution.”

L85: “To answer these questions, we simulate the glacier mass balance and runoff of the small-scale Gletsch catchment (44% glacierized, Rhonegletscher) at daily resolution over a 22-year period, using the the Glacier Evolution Runoff Model (GERM, Huss et al., 2008b; Farinotti et al., 2012).”

RV2 SP2: L75-77*: please specify the name of the river/catchment*

revised text: *“In this study, we investigate the impact of meteorological forcing products and spatial model resolution on the reliability of simulated glacier mass balance and runoff within the well-instrumented Gletsch catchment, a 39.4 km^2 glacierized headwater basin of the Rhone River in the Swiss Alps”*

RV2 SP3: caption of table 1: *please specify the name of the glacier*

The name of the glacier was already specified. Original text:” Summary of the catchment (Gletsch) and glacier (Rhonegletscher, including the main glacier and 10 small glaciers in the same catchment) characteristics [....]”

RV2 SP4: figure 1*: please add the river more clearly*

The figure and the corresponding caption is updated to now show the river network and the proglacial lake

Revised caption Figure 1: “Gletsch headwater catchment. The blue dot in the upper-left inset marks the location of the catchment within Switzerland. The right panel shows the catchment area, with glacierized area (in white) and contour lines (100-meter intervals, in cyan) over the glacier for the year 2016 according to the (Linsbauer et al., 2021, Swiss Glacier Inventory (SGI)). Contour lines are shown only for the glacierized area. The red dot marks the location of the catchment outlet and the gauging station at Gletsch. The rivers and the proglacial lake shown on the map are taken from the HydroRIVERS (Lehner and Grill, 2013) and HydroLAKES datasets (Messager et al., 2016), respectively. [....]”

RV2 SP5: Figure 2: *it is not clear how the box plot is made (temporal vs. spatial aggregation) please give more details*

The figure caption was updated to explain this in more detail:

Figure 2 caption correction (already including the correction from MC1): “[....]. Temperature and precipitation of the gridded products were spatially averaged over the catchment. Temperature was then corrected to the mean catchment elevation using a product-specific monthly average temperature lapse rate (cf. Supplementary Table S1) while precipitation is given as the mean catchment precipitation. For the box plots, the 22-year daily precipitation series was aggregated to mean monthly sums.”

RV2 SP6: line 156*: please add the calibration and validation periods*

At the end of section 3.5: “In this study, model calibration was performed over the period 2013–2021, which aligns with the availability of high-resolution geodetic glacier volume change data. This period serves as the calibration window for both the single- and multi-data calibration approaches. Model evaluation was then conducted over the full simulation period (2000–2022), allowing assessment of long-term model performance, seasonal variability, and year-to-year consistency. This fixed calibration–evaluation approach was selected to maintain consistency across experiments.”

RV2 SP7: lines 159-163: *please add the land cover areas*

See our reply to MC1: we now included a more detailed description of the runoff routing/handling of the non-glacierized areas and point more clearly at the original works by Huss et al. (2008b) and Farinotti et al. (2012), where the full details are given.

RV2 SP8: line 178*: please give the values for the lapse rates (evolving in time or not?)*

We now provide the applied temperature lapse rates in Supplementary Table S1 and clarify that we use monthly average temperature lapse rates derived from each of the meteorological products and a fixed precipitation lapse rate derived from the grimsel meteorological station and surrounding stations and previous studies (e.g. Farinotti et al., 2012). For the revised text, see our answer to MC1.

The table is now included in the supplementary material and can be viewed in the response to Reviewer 1.

RV2 SP9: line 189*: please give a reference used to select the T° values.*

The threshold values are based on Hock (1999), which is now referenced in the text.

RV2 SP10: figure 4: *please redo it more readable (two small font).*

We increased the font size of the Figure.

RV2 SP11: Line 226*: How are the values chosen?*

We now clarify that with the following wording, Line 222: “Geodetic glacier mass change serves as the primary constraint, and additional constraints can include measured runoff data. During the calibration process, the model adjusts the ablation parameter, which includes the melt factor(FM) and the radiation factors for ice and snow (rice/snow) in an automated procedure. FM and rice/snow have a fixed relation to each other (rice/FM = 0.024; rsnow/rice=0.66). The ratio between the parameters was adopted from earlier applications of the same model, which demonstrated their suitability for glacierized catchments in the Swiss Alps (Farinotti et al., 2012).[......].”

RV2 SP12: Table 3: *please add the values of NSE (and other metrics, see MC3)*

We added the specific values for both NSE and KGE in Supplementary Table S4. The KGE is now also shown in all relevant figures of the manuscript. Furthermore, we have added a full list of the relevant model parameters to the supplementary material Table S2.

RV2 SP13: Lines 285-286: *...'shows that ice melt may be underestimated...' How could you conclude that? Indeed it is not possible to quantify this term 'ice melt' on the basis of observed runoff alone.*

We have now clarified this sentence in Lines 284-286: "When forced with MSgrid and ERA5-Reanalysis, the model produces up to 20% less ice melt than when forced with Grimsel (which yields the results that are most consistent with the observed total runoff)."

RV2 SP14: L294 : *0.6 and 0.8 for NSE are not 'good' , please moderate.*

We have moderated our wording and now describe NSE values between 0.6 and 0.8 as "indicative of moderate performance".

RV2 SP15: figure caption of figure 7*, it is not possible to select only a selected period to draw conclusion. One can have some doubts about the hydro-glaciological model with NSE below 0.2 for some months.*

See our reply to MC3: we edited the figure caption and section 3.6 to clarify why we only select the summer period.

RV2 SP16: figure 8. *Please add the value for 2011 (which should be 91.9 million m-3).*

We apologise this was a plotting mistake and we thank the reviewer for spotting it. The value is now added.