It was a please to read the manuscript on "Modelling runoff in a glacierized catchment: the role of forcing product and spatial model resolution". The study analyzes model performance as a function of spatial resolution of the modeling domain and the choice of the precipitation products. The findings of the study are essential for finding optima between computational effort and minimum resolution needed for accurate glacio-hydrological simulations. The manuscript is well-structured and is well-written.

I appreciate that the authors have distinguished between spatial resolution of input data (i.e., precipitation) and resolution of model elements. However, I find that the effect of precipitation resolution is not well isolated in this study as it compares two things simultaneously, namely different source of precipitation (i.e., interpolated gauged and reanalysis) and different spatial resolution associated with each of the selected dataset. I think this aspect can be easily addressed by running additional simulations. Please find my detailed comments below.

Kind regards,

Larisa Tarasova

Detailed Changes:

Major comments:

MC1: 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.*

Thank you 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.

MC2: 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.

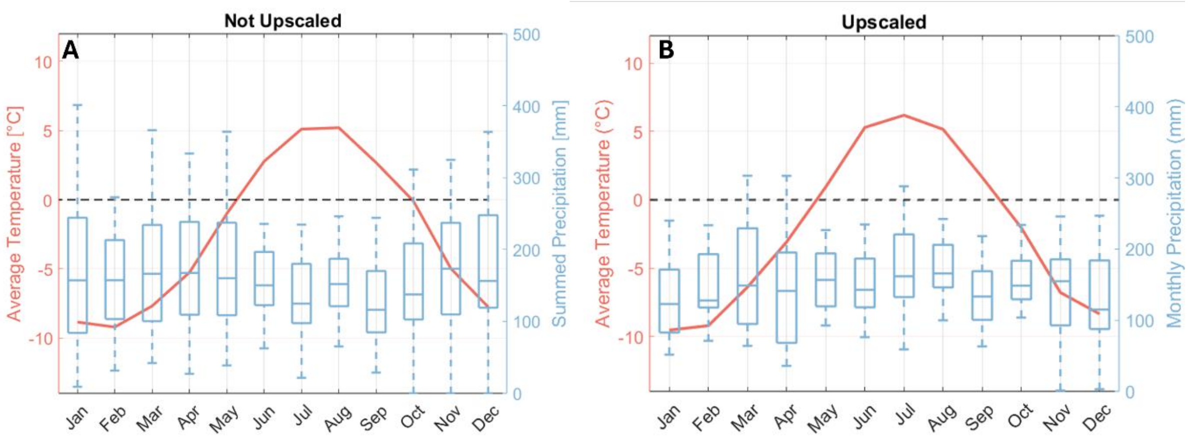

Figure S1*: Average monthly temperature and precipitation from the MS_{grid} for the period 2000-2022. (A) Temperature and precipitation from the product's original spatial resolution (1 km) aggregated over the catchment. (B) Temperature and precipitation aggregated over the catchment after degrading 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 average temperature lapse rate provided in Supplementary Table S1. Precipitation is plotted as the mean catchment precipitation.*

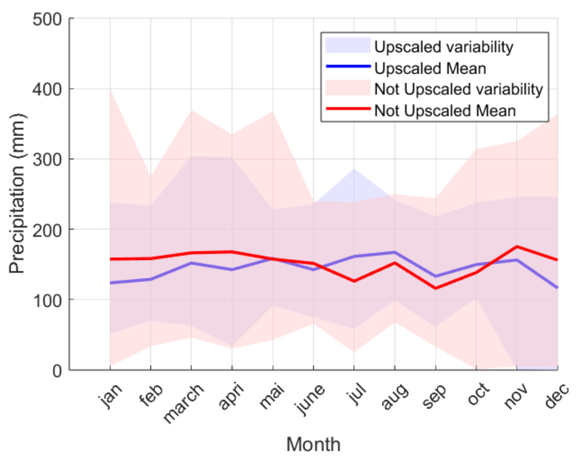

Figure S2*: Comparison between the mean 2000-2022 precipitation from the MS_{grid} product for both the upscaled (blueish) and not upscaled (reddish) methods. Colored area shows the variability of precipitation, while the line corresponds to the mean precipitation.*

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

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 [....]."

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.*

Thank you for your comment 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**

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)."

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)."

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, 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. The catchment outline is provided by the Federal Office for the Environment (FOEN)."

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 mass loss change between 2013 and 2021. The geodetic mass loss was determined by differentiating two high-resolution DEMs for Rhonegletscher acquired by dedicated monitoring flights on 21 Aug. 2013 and 20 Aug. 2021 (GLAMOS, 2024b). The resulting 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)"

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 of snow accumulation and ice melt across the entire glacier surface both in late April and September. Winter snow observations from 150 up to 300 snow-sounding locations were converted to water equivalent using snow density measurements. Measurements of local annual mass balance at a network of 10 ablation stakes 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."

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.*

Thank you for your 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). For each meteorological product, temperature lapse rates were computed as monthly averages 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 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."

***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

***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*

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

Thank you 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*."

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). [.....]"*