Detailed Changes:

General comments

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:

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) <u>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)."</u>

Text revision section 5.1 line 335 following:" In<u> 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."</u>

- *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). 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."

- *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: ”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.”

- *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."

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 on the accuracy of the 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.

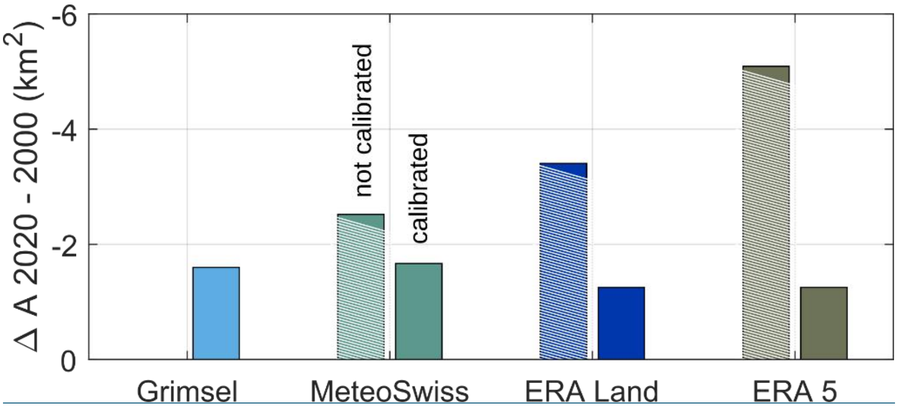


Figure S3: *Effect of the forcing products considered in Experiment 1 on the glacier area change simulated for the period 2000-2020 when the model is (right bars) or is not (left bars) calibrated to glacio-hydrological observations . For this sensitivity analysis, "Grimsel" is the default, meaning that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

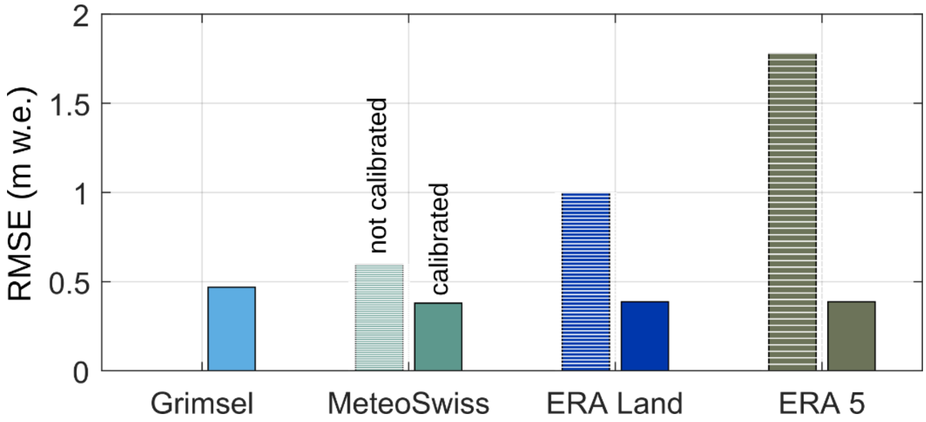


Figure S4: *Same as Figure S3 but showing the root-mean-squared-error (RMSE) between modeled and observed glacier-wide annual glacier mass balance in the period 2007–2022.*

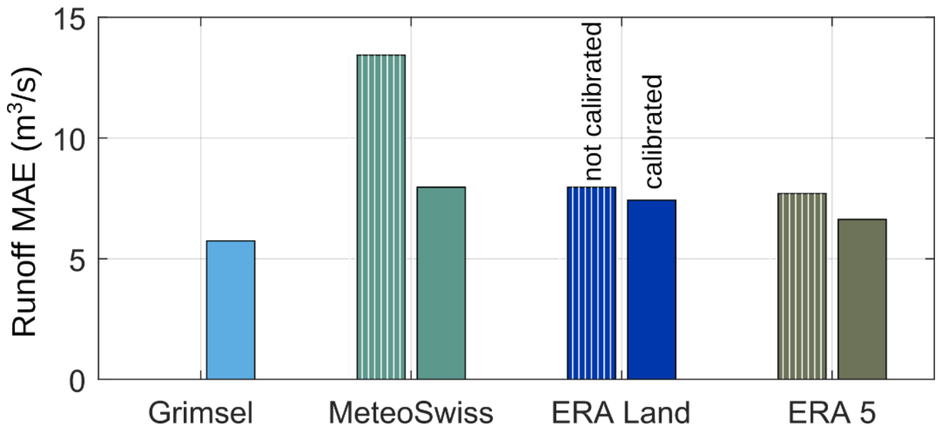


Figure S5: *Same as Figure S3 but showing the mean absolute error (MAE) in modeled annual runoff for the period 2000–2021.*

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.

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. 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?

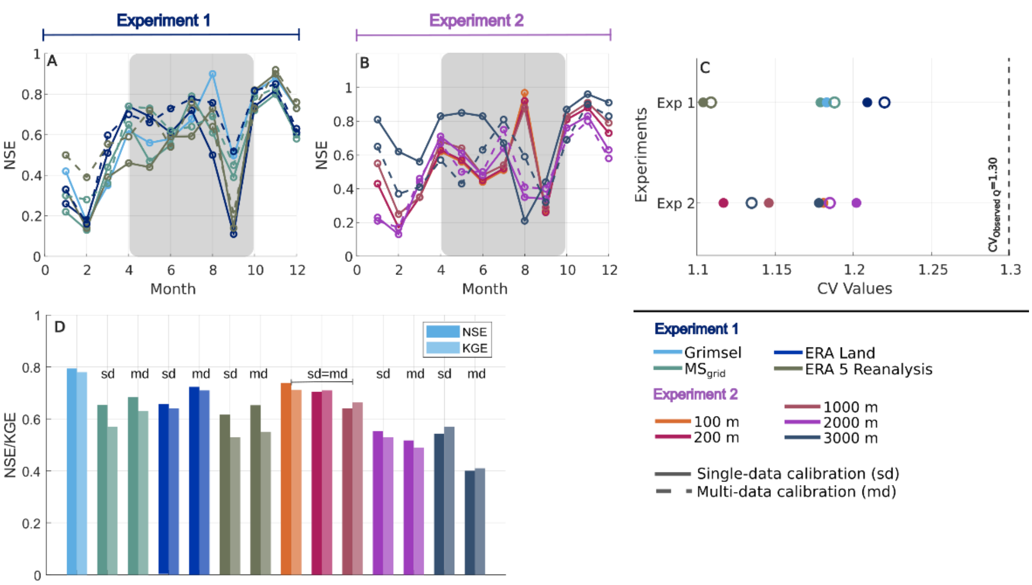
We appreciate the reviewer's comment and acknowledge the ongoing discussion regarding the limitations of the NSE as a model evaluation metric. We now also included the KGE alongside NSE to provide a more balanced view on model performance. This resulted in an addition of a plot to Figure 7 and minor 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 and KGE for each simulation. sd indicates a simulation performed with the single-data calibration, md with the multi-data calibration. For the simulations of Experiment 2 with the 100-1000 m resolution the sd and md NSE and KGE are the same.

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:

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)."

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"*

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 [....]"

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. [....]"

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."

***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.”

***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.

***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.

***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.

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

We increased the font size of the Figure.

***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).[……].”

***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.

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).”

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”.

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.

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.