Response to RC1:

Overall, this is robust research presented in a well written manuscript. I was impressed with the thorough analysis you performed when comparing the simulated results to the in situ field data. You also produced clear visuals to report the results. I have some comments below in response to a few places in the paper where I thought you could add additional context.

Answer: Thanks very much for the helpful review and appraisal of the value of our work.

Introduction

Comment #1: A paragraph is needed introducing, explaining, and justifying the use of the Water Balance Simulation Model (WaSiM) as the hydrologic model for this study. Given that this research is the first to apply WaSiM in high-elevation environment, this explanation/knowledge gap should also be addressed in the introduction.

Answer: We agree with the comment. In the revision, we have elaborated and strengthened the explanation on this aspect at L 79-87 in the Introduction. To answer the comment here, we adopted WaSiM in this research because WaSiM is one of the very few physics-based and fully-integrated hydrological models that enable to model both surface and subsurface in detail in glaciated catchments by updating the glacier volume. Such multi-variate, spatially-distributed, and physics-based modeling in both surface and subsurface potentially allows new hydrological process understanding and plausible flow and storage quantification, but it has been rarely done in **high-elevation glaciated environments**. Given this crucial knowledge gap, the novelty of this work lies in being the first to apply a **fully-distributed, fully-integrated surface-subsurface** hydrological modeling approach using WaSiM, in which the model parameters are constrained with extensively collected hydro-meteorological data of both surface (glacier, snow, river discharge) and subsurface, including the rarely monitored in-situ groundwater levels in high elevations in the central European Alps (Martell Valley in Northern Italy).

Comment #2: Additionally, I think a paragraph is needed explaining how in situ data can be used in these fully-integrated models would be helpful – particularly by highlighting the rarity of in-situ groundwater data.

Answer: We thank the reviewer for this suggestion. We agree with the suggestion and have enhanced the elaboration on this aspect at L 68-78 in the Introduction. Specifically, we summarized the challenges of using in-situ, point-scale data to constrain distributed models (scale-gap) and even more so in highly heterogeneous landscapes, with supported references to snow heights, soil moisture and groundwater levels (the three most commonly used types of in-situ data for these models).

Study Area

Comment #3: Figure 1b – The basemap is present showing the Germany label cut off. I would recommend fixing your map.

Answer: Thanks for pointing this out. We have fixed this issue on the map.

Comment #4: Section 3.2 You mention using the sen slope and Mann-Kendall test – please also add your p-value significance threshold, confidence intervals, etc.

Answer: We have applied a standard p-value significance threshold of 0.05 and the 95% confidence interval (CI). This detail was added to L 155-156 in Section 3.2. The p-value and CI are now added along with the slopes in the Result Section 5.1.

Comment #5: Section 4.4.3 It seems to me that there should be additional statistical significance tests such as RMSE, MAE and p-values. I am recommending adding the RMSE and MAE analysis. Also, I recommend expanding section 4.4.3 to include the NSE, spearman rank correlation coefficient, and p-value thresholds that you considered "significant" in this analysis.

Answer: We agree and have added the RMSE and MAE analysis and the p-value threshold (i.e. 0.05), besides the NSE and Spearman rank correlation coefficient, to help evaluate model performance. The revisions are provided at L 314-316 in Method Section 4.4.3 and in the Result Sections 5.1, 5.3, and 5.4.

Comment #6: Line 368: please provide the river gauge and borehole ID 4479 water fluctuation in order to compare it to the 0.8 reported for the borehold ID 4478.

Answer: The water level fluctuation at the river gauge is 0.24 m, that at borehole ID 4479 is 0.26 m and that at borehole ID 4478 is 0.88 m. We have added these details to facilitate the comparison at L 378-379 in Section 5.2.

Comment #7: Line 565: I'm not sure that the hypothesis that "subsurface lateral flow may have limited contribution to streamflow generation at the high-elevation headwater landscapes." Is fully explained in the introduction. Therefore, I would recommend ensuring that this narrative is present throughout the paper, including an additional explanation in the introduction. I'm also not sure how this is an "outrageous hypothesis" – perhaps expand on why this hypothesis is unusual for hydrologic research?

Answer: We agree and have added this context and explanation at L 46-67 in the Introduction and rediscussed this aspect in detail in the Discussion Section 6.2. In particular, we have referred the reader to the widely accepted runoff generation classification diagram by Dunne (Dunne, 1983; Trancoso et al, 2016), which states that in humid climates and steep slopes, runoff generation is dominated by lateral subsurface stormflow (in combination with saturation-excess overland flow). However, the streamflow generation process in the steep terrain in the cryosphere is still not fully understood and whether the Dunne diagram is applicable in such environment.

It is noteworthy that our conclusion refers specifically to the modeled dominant contributions to river discharge at the daily catchment scale. Under the simulated conditions, streamflow can be adequately explained by groundwater baseflow and surface runoff, while subsurface lateral flow (or "interflow", in the unsaturated zone) is not needed by the model to explain the observed dynamics, while considering both the river and the aquifer systems. This might be "outrageous" or "counter-intuitive" for many readers: in fact, in alpine environments, we have steep slopes and

can assume high contrast between hydraulic conductivities in the subsurface, which typically favors subsurface lateral flow (McGuire et al., 2024). However, we also have shallow soils, i.e., limited storage potential, which can favor percolation to aquifers, which in turn can be highly connected to the stream network and show high dynamics (Hayashi, 2019). As such, our model-based conclusion is plausible, in particular given that it reproduces streamflow (flux) and groundwater dynamics (storage).

As reviewed by McGuire et al. (2024), observing subsurface lateral flow remains a considerable challenge. The limited availability of observational data has substantially constrained our ability to accurately predict interflow and quantify its contribution to streamflow across diverse hydrological settings. There remain significant opportunities to improve our understanding of interflow through both advances in hydrological modelling and observational approaches.

Comment #8: Conclusion: I think you can add a paragraph on how the findings of this research can be applied to similar high alpine ecosystems throughout the world in order to give the research a "global application"

Answer: Thanks for this helpful suggestion. We have elaborated more on the applicability of the research findings in a global context at L 711-721 in the Conclusion Section. As the first physics-based, spatially-distributed, and fully-integrated surface-subsurface hydrological modeling study with WaSiM in the high-elevation glaciated environment, the achieved new insights on hydrological processes should be further tested and confirmed in other high Alpine settings. The insights from this study are most likely applicable to the mountain ranges with similar climatic, geological, and landscape settings and the sites with similar subsurface permeabilities, such as the central European Alps, Scandinavian Mountains, and Canadian Rockies. The high-elevation headwater catchments in these non-carbonate mountain ranges share key characteristics, where Quaternary sediments host shallow groundwater systems, impermeable bedrock restricts deep percolation, and subsurface lateral flow is often limited. These are opposed to settings with distinct subsurface permeabilities, e.g. the Hindu Kush Himalaya or with very different climatic settings, e.g. the Andean Alps. Future investigations can refer to the demonstrated workflow and recommendations (Sec. 6.3) and potentially overcome the limitations (Sec. 6.4). Such global modeling case studies and intercomparison contribute to a holistic understanding of hydrological processes in high Alpine environments.

References

Dunne, T. (1983). Relation of field studies and modeling in the prediction of storm runoff. *Journal of hydrology*, *65*(1-3), 25-48.

Hayashi, M.: Alpine Hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. Groundwater, 58 (4), 498-510. https://doi.org/10.1111/gwat.12965, 2019.

McGuire, K. J., Klaus, J., & Jackson, C. R. (2024). JAMES BUTTLE REVIEW: Interflow, subsurface stormflow and throughflow: A synthesis of field work and modelling. *Hydrological Processes*, *38*(9), e15263.

Trancoso, R., Larsen, J. R., McAlpine, C., McVicar, T. R., & Phinn, S. (2016). Linking the Budyko framework and the Dunne diagram. *Journal of Hydrology*, *535*, 581-597.

Response to RC2:

Review of the Manuscript

This is a good study that demonstrates significant effort in conducting extensive fieldwork and modeling works. However, I believe the manuscript still needs improvements. I've outlined several comments to assist in this process, as there is potential for enhancement during revision (before publication).

Answer: Thank you for taking the time to review our manuscript. We have revised our manuscript accordingly and provided point-to-point response below.

Major Issues

Comment #1: Structure and language: The organization and clarity of the paper should be enhanced to improve comprehension.

Answer: In the revision, we have focused on improving the organization and clarity of the manuscript by following the reviewer's detailed comments below. We have especially rewritten and reorganized the Introduction, Results, and Discussion sections in the revised manuscript.

Comment #2: Subsurface flow concerns: I remain unconvinced by the claims regarding limited subsurface flow contributing to river flow.

Answer: We acknowledge the concern of the reviewer, which is also mentioned in the specific comments below at L11-13 (Comment #4) and L657-659 (Comment #18). We agree that the limited contribution of subsurface lateral flow to river discharge needs to be justified more clearly and stated with appropriate caution. In our study, this interpretation is not based on direct subsurface lateral flow measurements. Rather, it emerges from the ensemble of hydrological simulations used for model calibration, which was evaluated against both groundwater level and river discharge observations. Across this ensemble, subsurface lateral flow appeared to be consistently calibrated to low values, as shown in Figs. S9-S13 of the Supplementary Material. We have revised the manuscript to make this explicit and to avoid overstating that rapid subsurface flow processes are absent; rather; rapid subsurface lateral flow (i.e., the associated response dynamic) is not needed to explain catchment-scale daily discharge and local-scale groundwater dynamics. This is compatible with, for example, the observation discussed by Hayashi (2019) that alpine aquifer systems may have highly dynamic, hydraulically active parts that contribute to streamflow. Given the limited number of integrated hydrological modeling studies in high-elevation glaciated catchments, additional applications in other high Alpine environments are needed to determine whether this behaviour is site-specific or reflects a more general characteristic of high-elevation hydrogeological systems.

Our conclusion refers specifically to the modeled dominant contributions to river discharge at the daily catchment scale. Under the simulated conditions, streamflow can be adequately explained by groundwater baseflow and surface runoff, while subsurface lateral flow (or "interflow", in the unsaturated zone) is not needed by the model to explain the observed dynamics, while

considering both the river and the aquifer systems. This might be counter-intuitive for many readers: in fact, in alpine environments, we have steep slopes and can assume high contrast between hydraulic conductivities in the subsurface, which typically favors subsurface lateral flow (McGuire et al., 2024). However, we also have shallow soils, i.e., limited storage potential, which can favor percolation to aquifers, which in turn can be highly connected to the stream network and show high dynamics (Hayashi, 2019). As such, our model-based conclusion is plausible, in particular given that it reproduces streamflow (flux) and groundwater dynamics (storage).

As reviewed by McGuire et al. (2024), observing subsurface lateral flow (interflow, in the unsaturated zone) remains a considerable challenge. The limited availability of observational data has substantially constrained our ability to accurately predict interflow and quantify its contribution to streamflow across diverse hydrological settings. There remain significant opportunities to improve our understanding of interflow through both advances in hydrological modelling and observational approaches.

Specific Comments:

Comment #3: Abstract Lines 9-11: The phrase “highlighting the need for improved subsurface parameterization” is confusing. Did you address this in your manuscript? If so, how? If it's a future recommendation, how do you plan to achieve it?

Answer: This is indeed a key improvement in our study. This study is among the first to adopt a **fully-distributed** and **fully-integrated surface-subsurface hydrological modeling** approach in a **glaciated high Alpine setting**. In contrast to the widely adopted simplified representation of subsurface processes, such as by adopting a linear reservoir or by transferring the outflow from unsaturated zone directly to the streams without a groundwater module (Coderre et al., 2026), our study adopts a spatially-explicit and fully-distributed hydrological model which solves the hydrological processes in the unsaturated zone with the Richards equation and solves the groundwater flow with Darcy’s equations in two dimensions. These methods require gridded parameterization of the subsurface and they solve the subsurface flow in a much more detailed manner compared with the above-mentioned widely adopted simplified approaches (e.g., one-dimensional linear reservoir, or simply no groundwater module). More details of how we address subsurface parameterization and modeling are provided in detail in Section 4.2 in the manuscript. In the revision, we enhanced this context to improve comprehension at L 27-36 in the Introduction. We also answered in more detail the same question mentioned again in Comment #10 about subsurface parameterization of the reviewer.

Comment #4: Abstract Lines 11-13: The statement about subsurface lateral flow being constrained to zero and its limited influence on river discharge is questionable. Could this subsurface flow be as rapid as overland flow? What is your model's temporal resolution? If overland flow dominates, can you cite evidence for this, such as land erosion?

Answer: We agree that the role and potential timescale of subsurface lateral flow should be stated more carefully. We have interpreted this result more carefullly in the revised Abstract to avoid easy misinterpretation. It is now framed as “Surprisingly, model simulations with limited subsurface

lateral flow (interflow, in the unsaturated zone) contributions best fit the observed groundwater levels and streamflow, suggesting that groundwater flow (in combination with surface runoff) is dynamic enough to explain the observed daily streamflow dynamics in the computational settings.".

In our model, subsurface lateral flow was not assumed to be zero a priori; rather, it was constrained to zero after evaluating a large ensemble of hydrological simulations, as shown in Figs. S9-S13 in the Supplementary Material. This implies that, at the daily temporal resolution and 25 m x 25 m spatial resolution of our model, subsurface lateral flow does not make a significant contribution to explain the simulated streamflow dynamics in comparison with the contributions of groundwater baseflow and surface runoff.

The model represents subsurface lateral flow, groundwater baseflow, and surface runoff as separate flow components, and their magnitudes and temporal dynamics are quantified explicitly separately (see Section 4.3). One of the ensemble model setups includes the same parameterization for surface runoff and subsurface lateral flow (i.e., both flow components are the same fast), however, cannot reproduce the observed groundwater dynamics. Moreover, such model setup (with two flow components parameterized the same) is not parsimonious and experiences equifinality issue.

In our calibrated model simulations, the rapid runoff responses are primarily attributed to surface runoff, while the groundwater component represents slower baseflow contributions from the headwater floodplain. We acknowledge, however, that our model resolution (1 day) does not allow us to rule out very rapid preferential subsurface flow processes at shorter, event-based timescales. We have revised the manuscript in Discussion Section 6.2 to clarify that our conclusion refers to the dominant modeled contributions to discharge at the daily catchment scale, rather than to the complete absence of fast subsurface flow processes in the unsaturated zone.

As supporting evidence for the importance of surface processes, a recent high-temporal-resolution soil erosion study within the study site reported high erosion potential near the Zufallferner during rainfall and melt events (Altmann et al., 2025). This is consistent with the interpretation that overland flow can be an important process in the upper headwater floodplains. In the revision, we have added this reference at L601-603 in Discussion Section 6.2. We have also highlighted to use alternative model structures, independent observations, and event-scale analyses as relevant future works to further test the role of subsurface lateral flow (interflow, in the unsaturated zone) in high-elevation headwater floodplains (L534-538 in Discussion Section 6.1).

Comment #5: Abstract Lines 14-16: Please clarify how point scale groundwater observations were incorporated in your modeling. Or did you just use them for model validation or calibration?

Answer: We have used the point-scale time-series groundwater level observations to calibrate the subsurface module parameters in the model (see Sections 4.2, 4.4.1, and 4.4.3 for explanations on subsurface modeling, parameterization, and reference signals of calibration). Specifically, for calibration of the groundwater module using point-scale groundwater level observations, we extracted an ensemble of simulated groundwater level time series from model grid cells located

in the vicinity of each observation borehole within the model domain. The resulting ensemble of grid-cell simulations was then compared against the observed groundwater level records at the corresponding borehole locations (see Fig.5 and Fig.S14 in the Supplementary Material). Subsurface model parameters were subsequently adjusted based on the discrepancies between simulated and observed groundwater levels to improve model performance. We enhanced the elaboration on how to use point-scale groundwater observations to calibrate the groundwater module at L 288-292 in Section 4.4.2 in the revision.

Comment #6: Research Questions: The connection between research questions and the main text is unclear. Specifically, what do you mean by "the commonly adopted more simplified" in research question (iii)? How does your study differ from existing researches (literature)? Please clarify them. Clarification on how point-scale groundwater observations address your earlier identified research gaps is essential.

Answer: "The commonly adopted more simplified" approach points to the overly simplified representation of the subsurface (i.e., the unsaturated zone and the groundwater) in many hydrological models, such as the lumped-parameter hydrological models (e.g. HBV (Seibert and Bergström, 2022)) and semi-distributed hydrological models (e.g. mHM (Kumar et al., 2013), THREW-T (Nan et al., 2021)). In these models, the subsurface is generally simulated with a single linear reservoir. In another more simplified manner, the percolation (leaked water) from the bottom layer of the soil is even directly transferred to the stream to fulfill the water balance due to the lack of a groundwater module (Coderre et al, 2026). These clarifications are added at L27-36 in the Introduction in the revision.

In comparison with the above-mentioned simplified modeling approach, our study adopted a much more detailed hydrological modeling approach, that is, the physics-based, **fully-distributed**, and **fully-integrated surface-subsurface** hydrological modeling approach to model the **high-elevation glaciated environment** for the first time. Specifically, in the spatially-gridded subsurface the unsaturated zone is solved with the Richards equation and the groundwater model is solved with the Darcy's equation in two dimensions in detail (Schulla, 2024). This marks a key improvement compared with the widely adopted one-dimensional linear reservoir approach or the simply ignorance of the groundwater module. These points and the research novelty are now clarified at L 46-87 in the Introduction.

Furthermore, we enhanced the clarification on how in-situ data can be used in fully-integrated models at L 68-78 in the Introduction. Specifically, we summarized the challenges of using in-situ, point-scale data to constrain distributed models (scale-gap) and even more so in highly heterogenous landscapes, with supported references to snow heights, soil moisture and groundwater levels (the three most commonly used types of in-situ data for these models). We also enhanced the elaboration on how to use point-scale groundwater observations to calibrate the groundwater module at L288-292 in Section 4.4.2 in the revision and have provided recommendations on groundwater piezometer installation and evaluation in Section 6.3.

Comment #7: Line 75: The term "alone" needs clarification.

Answer: Here the “alone” means an integrated surface-subsurface hydrological model is adopted, which allows for modeling both surface and subsurface processes temporally and spatially continuously within WaSiM itself. This is opposed to the approach of adopting a hydrological model (e.g. SWAT) and coupling it externally with another groundwater model (e.g. MODFLOW). We have revised the research question for clarity.

Comment #8: Line 76: Could you explain “external coupling framework”?

Answer: The “external coupling framework” means, for example, to couple a hydrological model (e.g. SWAT) with an ice model or a groundwater model (e.g. MODFLOW) (Ice model + hydrological model + groundwater model). This happens when a hydrological model does not have these specific modules, but it needs to be applied to model the environment with ice cover or for improved groundwater simulations. To apply this type of hydrological model in the glaciated environment, the researchers need to first run the ice model, then give the outputs of the ice model as the inputs to the hydrological model and finally give the outputs of the hydrological model as inputs to the groundwater model without internal feedback. The external coupling approach often disconnects the interactions and processes spatially between the surface and subsurface, which are simulated separately with different models. We have revised the research question for clarity.

Comment #9: Line 81: The phrase “with all its limits” should be reconsidered for clarity. “A missing piece in the literature”: This phrasing may imply you’re the first; please revise, as while you may be the first to implement WaSim in high-elevation glaciated environments, you may not be the first to model in this setting?

Answer: No model is perfect with pros and cons, especially given the challenges of modeling complex topography and processes in a high-elevation glaciated catchment. The WaSiM model, as one of the most advanced hydrological models for simulating high Alpine environments, also has its limits besides the numerous advantages. We have discussed the limits of applying WaSiM in detail in Section 6.4 Limitations and we removed this unnecessary phrase here in the Introduction.

Regarding the novelty, it is not only because of implementing WaSiM in such a complex environment, but it is more because a **spatially fully-distributed and fully-integrated surface-subsurface hydrological modeling approach** in general has been rarely applied to modeling such an environment and calibrated with both streamflow and in-situ groundwater level data. We revised this phrase as “rarely studied” in the Introduction.

Comment #10: Table 4: There appear to be no spatial variations for the parameters presented. Could you clarify this assumption? If no spatial variation exists, your claim of an “improved” subsurface schematization is, perhaps, questionable?

Answer: The subsurface parameters are not fully spatially homogeneous. We applied the essential spatial subsurface heterogeneity by calibrating the hydraulic conductivity for each soil layer - vertical spatial heterogeneity, and for the soil in each land cover category (rocks, grassland,

forest) – horizontal spatial heterogeneity. The "ksat" in Table 4 is provided for the soil in each land cover, and its range indicates the variations of hydraulic conductivity between the soil layers vertically. We clarified this point in the caption of Table 4 in the revision.

Furthermore, given that all five installed groundwater piezometers are located in the same hydrogeology (sedimentary aquifer, see Fig. 1a) and fluctuate around similar depths, other parameters are hence assumed to be the same because we cannot constrain them with the available dataset. The modeling results are with high confidence in the upper headwater floodplain where the piezometers are installed, and the results for the whole catchment aim to provide an overview despite lower confidence (L269-273 in Section 4.4.1). Compared with the widely adopted simplified representation of subsurface in hydrological models, such as a single one-dimensional linear reservoir or no groundwater model at all (revised at L27-36 in the Introduction), our implemented spatially-distributed subsurface parameterization with key variations in hydraulic conductivity for different soil layers and different land covers represents a key step forward for understanding surface-subsurface hydrological interactions and for improving water partitioning between surface and subsurface.

Besides that, more detailed heterogeneity of the subsurface characteristics is challenging to apply due to the rarely available soil and aquifer data in the glaciated high-elevation catchments. Many studies on the topic of stochastic hydrogeology show the challenges of estimating the spatially variable subsurface parameters even within an extremely small experimental area, starting from the detailed investigation of the famous sites of Borden and Columbus, for example (Rehfeldt et al., 1992; Woodbury et al., 1991). Assigning heterogeneous subsurface properties in the ungauged areas involves numerous assumptions. We therefore applied a conservative approach to ensure the key subsurface heterogeneity. Exploring the impact of even high-level spatially variable subsurface parameters is out of the scope of this study. The limitations on this aspect have been discussed in detail at L 650-662 in the Discussion Section 6.4.

Comment #11: Chapter 4: What is the temporal resolution of your model? I couldn't locate this information in the manuscript.

Answer: The temporal resolution of the model is daily, which is mentioned at L 122 in Section 3.1 when we introduce the daily meteorological data as model inputs. We also added this information at L5 in the Abstract and L 179 in Section 4 in the revision.

Comment #12: For soil analysis, I'm just wondering why not using SoliGrids (250 m resolution) for accuracy.

Answer: We acknowledge that the SoliGrids provides a finer resolution of 250m compared with the one (HWSD) that we adopted of 1km. Both datasets are global soil datasets. Given the uncertainty of global soil datasets in high mountain regions, using a slightly finer global soil product would not necessarily improve the simulations significantly. This is because in the high mountainous regions, a finer-resolution global soil product such as SoliGrids still contains high uncertainty due to sparse field observations. Besides that, the HWSD product remains widely used in regional modeling because of harmonized and internally consistent soil attributes across broad spatial extents. In this study, we also used the land use map (e.g., the presence of rocks,

forests, and grasslands) to evaluate and modify the soil type data, as another layer of validation of data quality. We strengthened the discussion of this limitation at L 676-683 in the Discussion Section 6.4.

Comment #13: Section 6.1: I think many paragraphs of this section should be moved to the Introduction. This part showed what has been done in previous researches and how your study cover the research gaps.

Answer: We thank the reviewer for this suggestion. We have revised Section 6.1 by removing the general literature review and research gap discussion. This section has been rewritten with a clean focus on the critical comparison of methods and results similarities and differences between the proposed model and existing approaches. We have also rewritten the Introduction Section for clarity and avoiding the repetitions.

Comment #14: Lines 350-354: The flow of these sentences could be improved; the conclusion about declining water availability seems unsupported by the previous sentences. While glacier melt (10 mm/year) may offset declining rainfall (-4.5 mm/year), could there be times when glacier melt ceases?

Answer: We have improved the flow of these sentences and clarity in Section 5.1. We also agree to revise the uncertain conclusion about declining water availability, as the trends are insignificant at the 95% confidence level. Note that besides the annual trend of rainfall and glacier melt, we also quantified the snowmelt trend on an annual scale. Here is our revision: The estimated annual trends of the flux components are −5 mm/year for rainfall (p-value = 0.50; 95% CI: −19 to +4 mm/year), −26 mm/year for snowmelt (p-value = 0.10; 95% CI: −62 to +4 mm/year), and +10 mm/year for glacier melt (p-value = 0.16; 95% CI: −5 to +30 mm/year). These results indicate that none of the flux components exhibit statistically significant trends over the study period. Among them, snowmelt shows the strongest decreasing tendency, whereas rainfall and glacier melt display comparatively weaker trends. However, the confidence intervals of all three components span both negative and positive values, indicating high uncertainty in the magnitude and direction of the long-term changes (L355-364 in Section 5.1).

Comment #15: Figure 8: The groundwater heads appear to represent groundwater depth (or depth to the water table).

Answer: The groundwater heads represent the depth to the water table, i.e. the distance between the ground surface and the groundwater table. We have clarified this point in the caption of Figure 8 in the revision.

Comment #16: Lines 474-476: Clarification is needed regarding recharge. are you implying that some of it is released outside your study area? How significant is this recharge?

Answer: We have improved the clarity of these sentences at L490-494 in Section 5.4. The “recharge” refers to the water flowing from groundwater to the river (i.e., the river is recharged by the groundwater). Regarding the magnitude, our results show that shallow groundwater

contributes significantly (up to 40%, see Fig.7) to the streamflow in winter. Deep groundwater processes were not directly constrained in this study because groundwater observations and model representations were limited to the shallow aquifer. However, we consider it unlikely that deeply circulating groundwater from the hillslopes of the proglacial outwash plain provides a substantial contribution to the stream network in the upper headwater floodplain. Such deeper groundwater contributions are more likely to occur in the valley floor, where longer flow paths may connect groundwater with the river network (see Fig.1b in the work of van Tiel et al. (2024)).

Comment #17: Lines 645-646: The reasoning for being unable to model permafrost due to a lack of soil temperature data is unclear. If such data are not existing, can we just simulate these temperature?

Answer: We have revised this sentence at L 686-690 in Section 6.4 to improve clarity. To answer the question here, the permafrost-related freezing and thawing processes are not included in our simulations because the corresponding WaSiM parameters could not be reliably constrained without soil temperature observations. Although the model can simulate soil freezing and thawing, spatially distributed soil temperature measurements are needed for calibration and evaluation of freezing and thawing depths, and such observations are rarely available in high-elevation catchments.

Comment #18: Lines 657-659: What evidence supports the claim of limited subsurface flow? If fast flow is primarily overland, then soil erosion may also provide insights.

Answer: Thanks for the comment and suggestion. We have answered this specific comment together with the major comment #2 and specific comment #4 on this topic. To restate it here, we agree that the statement regarding limited subsurface flow requires clearer justification. In this study, the inference of reduced subsurface flow contribution is not based on direct measurements, but rather on hydrological simulations and physical landscape characteristics that influence flow partitioning, including steep topography, shallow soil depth, and limited soil water storage capacity in the study area (see the rewritten Discussion Section 6.2). These conditions are generally associated with reduced subsurface storage and enhanced rapid runoff generation. We also agree that soil erosion processes can provide additional indirect evidence for surface-dominated flow pathways. High erosion potential has been shown in the Zufallferner within the study site (Altmann et al., 2025). While erosion was not explicitly quantified in the current analysis, we acknowledge that integrating erosion indicators would strengthen the interpretation of dominant flow pathways. We have revised the manuscript to clarify that our interpretation is based on hydrological modeling rather than direct subsurface flow measurements, and we have highlighted erosion and event-scale runoff generation as relevant future works to further test the role of subsurface lateral flow in high-elevation headwater floodplains. All these revisions are now provided in the Discussion Section 6.2.

References

Altmann, M., Fischer, P., Haas, F., Pfeiffer, M., Ramskogler, K., Rom, J., Kara-Timmermann, D., Himmelstoß, T., Heckmann, T., and Becht, M.: Simulation of soil erosion in the Little Ice Age

glacier foreland of the Zufallferner (South Tyrol, Italy) using Erosion-3D, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-5734, https://doi.org/10.5194/egusphere-egu25-5734, 2025.

Coderre, P., Knoben, W., Thébault, C., Vásquez, N., Clark, M., and Pietroniro, A.: The impact of benchmark selection on spatial patterns of model evaluation metrics, EGU General Assembly 2026, Vienna, Austria, 3–8 May 2026, EGU26-15343, https://doi.org/10.5194/egusphere-egu26-15343, 2026.

Hayashi, M.: Alpine Hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. Groundwater, 58 (4), 498-510. https://doi.org/10.1111/gwat.12965, 2019.

Kumar, R., Samaniego, L., & Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resources Research, *49*(1), 360-379. https://doi.org/10.1029/2012wr012195, 2013.

McGuire, K. J., Klaus, J., and Jackson, C.R.: JAMES BUTTLE REVIEW: Interflow, subsurface stormflow and throughflow: A synthesis of field work and modelling, Hydrological Processes, 38 (9), e15263. https://doi.org/10.1002/hyp.15263, 2024.

Nan, Y., Tian, L., He, Z., Tian, F., and Shao, L.: The value of water isotope data on improving process understanding in a glacierized catchment on the Tibetan Plateau, Hydrology and Earth System Sciences, 25, 3653–3673, https://doi.org/10.5194/hess-25-3653-2021, 2021.

Rehfeldt, K. R., Boggs, J. M., & Gelhar, L. W.: Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. Water resources research, 28(12), 3309-3324. https://doi.org/10.1029/92WR01758, 1992.

Seibert, J. and Bergström, S.: A retrospective on hydrological catchment modelling based on half a century with the HBV model, Hydrology and Earth System Sciences, 26, 1371–1388, https://doi.org/10.5194/hess-26-1371-2022, 2022.

Schulla, J.: Model Description WaSiM (Water balance Simulation Model), http://www.wasim.ch/downloads/doku/wasim/wasim_2024_en.pdf, 2024.

van Tiel, M., Aubry-Wake, C., Somers, L., Andermann, C., Avanzi, F., Baraer, M., ... & Yapiyev, V.: Cryosphere–groundwater connectivity is a missing link in the mountain water cycle. Nature Water, 2(7), 624-637. https://doi.org/10.1038/s44221-024-00277-8, 2024.

Woodbury, A.D. & Sudicky, E.A.: The geostatistical characteristics of the Borden aquifer. Water Resources Research, 27(4), 533-546. https://doi.org/10.1029/90WR02545, 1991.

Response to RC3:

Review for manuscript "New insights from fully-integrated surface-subsurface hydrological modeling in high-elevation glaciated environments" by Fan et al., 2026

Comment #1: Overall, I found the manuscript well written and clearly structured. The modelling decisions are generally well supported, and the approach seems appropriate given the complexity of the system and the amount of available data. The model appears to be reasonably well constrained, especially with respect to glacier and snow processes.

I particularly appreciated the comparison between simulations with and without lateral flow. This provides a clear and interesting result: lateral flow appears to need to be inhibited in order to reproduce observed groundwater levels. However, I found the implications of this result somewhat difficult to interpret. Does this suggest that, in this system, most water transfer occurs vertically rather than laterally between cells? Or could this be a structural artefact of the modelling framework? I think this result is important and would benefit from a more in-depth discussion of its physical interpretation, limitations, and broader implications.

Answer: Thank you for your interest and for taking the time to review our manuscript. The above comment is completed together with Comment #25:

Comment #25: Line 518 I would like to see a more detailed discussion of the implications of lateral flow. Why does this result matter? What physical conditions could explain the need to inhibit lateral flow? Alternatively, could this reflect limitations of the model structure or spatial discretization?

Answer: These comments have raised several key points: how to interpret the result? What are the implications? What physics could explain it? Is it an artefact due to model structure and spatial discretization?

First of all, we would like to underline that our finding indicates that, within the current model structure and parameterization, explicitly simulated subsurface lateral flow (i.e. interflow, in the unsaturated domain) is not needed and even leads to a poorer reproduction of the observed streamflow and groundwater dynamics. This can be a model artefact, primarily because of model complexity: we could simply not have found the calibration that leads to a good representation of streamflow and groundwater; our systematic testing, however, suggests that this is unlikely to be the case (see Fig. S9-S13 in the Supplementary Material).

The interpretation of the results is an interesting point and is addressed now in Section 6.2 and discussed below:

What could explain a "true" absence of subsurface lateral flow in the catchment? As discussed in the response to reviewer 2, in alpine environments, we have steep slopes and can assume high contrast between hydraulic conductivities in the subsurface, which typically favors subsurface lateral flow (McGuire et al., 2024). However, we also have shallow soils, i.e. limited storage potential, which can favor percolation to aquifers, which in turn can be highly connected to the

stream network and show high dynamics (Hayashi, 2019). Our model-based conclusion is a plausible representation of high alpine streamflow generation dynamics, in particular given that it reproduces streamflow (flux) and groundwater dynamics (storage).

Our result suggests that groundwater flow (in combination with surface runoff) is dynamic enough to explain the observed streamflow dynamics; this is in particular in line with the above-mentioned (review-based) discussion of highly active groundwater contribution to streamflow in alpine areas in the review papers of Hayashi (2019), Somers et al. (2020) and McGuire et al. (2024). The implications are that we should, in the future, design field campaigns that can challenge this result, in particular based on tracer observations (Xie et al., 2024). Furthermore, from a modelling perspective, we show that importance of calibrating any model against observations of streamflow and storage dynamics to avoid representing the observed streamflow dynamics with processes that might not be at play.

Comment #2: **Figure 1** Could the meteorological station symbols be made larger or shown in a brighter colour? It took me some time to identify the 12 stations. The piezometer labels are also somewhat difficult to read against the background; adding a light label background or halo might improve readability.

Answer: We have enlarged the meteorological station symbols (note that only 9 meteo stations are displayed in Figure 1a to allow enough space for the zoomed-in details to be shown within the catchment. The other 3 stations a bit further outside of the catchments were not shown but their coordinates are provided in Table 1). We have also added a light label background for the piezometers labels to improve readability.

Comment #3: **Line 116** The phrase “are requested” sounds slightly awkward. Would “are obtained” or “were obtained” be more appropriate? Also, because the stations are not labelled in Figure 1, it is difficult for the reader to know where specific stations such as Zufall or Careser are located. It may be helpful either to label them in Figure 1 or to move some of this information to the data availability statement.

Answer: We used “are obtained” in the revision. To avoid making Figure 1 too crowded with many labels, we have moved the data acquisition information of specific stations to the data availability statement.

Comment #4: **Table 1** Why is wind speed abbreviated as “WG”? “WS” may be more intuitive, unless there is a specific reason for using “WG”.

Answer: We used “WS” as the new abbreviation for wind speed in the revision.

Comment #5: **Line 142** I am not sure the reader needs all the details about who installed each station. It may be sufficient to include the years of operation and where the data can be accessed.

The installation details could potentially be moved to the data availability statement. Similarly, if the model ultimately uses daily data, is the measurement interval necessary here?

Answer: We agree and have moved the data owner information to the data availability statement. We also agree that mentioning the daily interval is sufficient here and the measurement interval is unnecessary, which is removed for clarity in Section 3.2.

Comment #6: **Line 147** November is not technically winter, so "at the onset of the low-flow season" may be more accurate. I would also be cautious about stating that a single water sample collected at the end of fall helps characterize winter interactions. This may be true in a limited sense, but the current wording may overstate what can be inferred from one sampling period.

Answer: We agree and adopt the "at the onset of the low-flow season" instead of "winter". Regarding the water sampling, they were collected in two sampling periods, i.e., in November 2022 and November 2023. We agree to be cautious with this statement and have acknowledged the uncertainty associated with the limited number of water samples at L 488-489 in Section 5.4 in the revised manuscript.

Comment #7: **Line 159** Could you clarify how reasonable the soil profile and soil type assumptions are? Were these compared with any field measurements? The 1 km × 1 km resolution seems coarse compared with some of the other spatial inputs, so it would be helpful to briefly justify this choice or discuss its implications.

Answer: The adopted spatial resolution of soil data is rather coarse, which is primarily limited by the sparse field observations available in the high mountain environments. A slightly higher resolution of soil product (e.g., SoliGrids in 250m*250m) may not necessarily improve the simulations significantly because of limited field observations in the high elevations. In this study, we used the land use map (such as the presence of rocks, forests, and grasslands) to evaluate and modify the soil type data, as another layer of validation of the soil data. The HWSD product was adopted as it remains widely used in regional modeling due to harmonized and internally consistent soil attributes across broad spatial extents. We justified this choice and discussed the potential implications at L 676-683 in the Discussion Section 6.4 Limitations.

Comment #8: **Line 165** It sounds like the model is calibrated to glacier area change and evaluated against mass balance. This is useful information, but I wonder whether it belongs in the spatial data section or would fit better in the modelling approach section.

Answer: We agree with the suggestion and have moved these calibration and evaluation details of the glacier module to L308 in the Modeling Subsection 4.4.3.

Comment #9: **Line 192** "Flexibly activated" is sufficient; "activated/deactivated" is probably not necessary, since deactivation is implied.

Answer: We adopted the “flexibly activated” and removed “deactivated” to avoid repetition.

Comment #10: **Line 197** This paragraph seems to mix the general model description with specific parameter choices. For example, why was the range of -2 to +2 selected? Were these values calibrated or assumed? This information may fit better in Section 4.4, where model parameterization and calibration decisions are discussed.

Answer: We agree that this information belongs to the model parameterization section. To clarify, the parameter “Trans” is a calibrated parameter and it governs the temperature transition range from snow to rainfall. The “Trans” calibrated to the value of 2 means a temperature transition range of -2 to +2 °C in the model, when the “TRS” is equal to 0 °C (which is also a calibrated parameter, meaning the temperature at which 50% of precipitation falls as snow). We have moved this information to Section 4.4.1 about calibration parameters.

Comment #11: **Line 199** If the bucket approach is not used, it may not need to be mentioned here.

Answer: We agree and have removed this sentence.

Comment #12: **Line 201** Could the algorithms be named more specifically, if they have formal names? The phrase “state of the art” feels somewhat overstated unless the methods are identified and their status is briefly justified.

Answer: There is indeed no formal name for these algorithms for calculating wind and gravitational redistribution of snow. This could be because these algorithms have been directly integrated into the hydrological model without being published as a stand-alone software. Therefore, we only mentioned the related references. We removed the word “state of the art” to avoid overstatement.

Comment #13: **Line 332** The sentence beginning “as global warming continues…” could likely be removed, as it introduces a broader climate-change framing that is not needed for the calibration discussion. If you want to contextualize the simulated value of -126 mm yr^{-1}, it may be more useful to compare it with existing glacier mass-balance estimates, for example from Dussaillant et al. (2025).

Dussaillant, I., Hugonnet, R., Huss, M., Berthier, E., Bannwart, J., Paul, F., and Zemp, M.: Annual mass change of the world's glaciers from 1976 to 2024 by temporal downscaling of satellite data with in situ observations, Earth System Science Data, 17, 1977–1991, 2025. https://doi.org/10.5194/essd-17-1977-2025

Answer: We have removed this sentence. Instead, we compared our result with the average glacier mass changing rate in central Europe in the suggested literature (-1.21 m vs. -1.29 m w.e. per year). The revision is provided at L 336-340 in Section 5.1 in the manuscript.

Comment #14: **Line 342** Could you clarify how snowmelt and glacier melt are distinguished in the model? For example, is snowmelt occurring on glacierized cells counted as snowmelt, or is it included in glacier melt? This distinction is important for interpreting the reported melt contributions.

Answer: The hydrological model WaSiM calculates and outputs the snowmelt on the non-glacierized cells, snowmelt on the glacierized cells, firn melt on the glacierized cells, and ice melt on the glacierized cells. We calculate the total snowmelt as the sum of snowmelt occurring both on the glacierized and non-glacierized cells. We clarified this point in the manuscript at L 351-355 in Section 5.1.

Comment #15: **Line 352** The sentence beginning "Overall…" may not be necessary here. The point seems more appropriate for the introduction or broader framing rather than this section.

Answer: We agree and have removed this sentence in this section.

Comment #16: **Figure 2** "Warm-up period" is understandable, but "spin-up period" is more commonly used in the modelling literature. Consider changing the terminology for consistency.

Answer: We agree and have adopted "Spin-up period" instead of "Warm-up period" in the revision.

Comment #17: **Figure 3** Could you clarify the reference level for river water level? The river appears to be above 0, so it is not immediately clear what 0 represents. Is this water level above the riverbed, with 0 indicating dry conditions? Or is it referenced to ground surface or another datum? A short clarification in the caption or text would help.

Answer: The reference level of 0 is the riverbed, with 0 indicating dry conditions and no water in the riverbed. We clarified this point in the figure caption in the revision.

Comment #18: **Figure 5** Could the line colours in panels c and d reflect the TWI classes shown in panels a and b? It may be clearer to group all simulations by TWI rather than by selected groundwater ID. At present, it is difficult to identify the IDs in panels a and b, and the groundwater borehole cell colour is similar to the TWI colour scale, which makes the figure harder to interpret. Also, where are the time series for wells 4473 and 4472?

Answer: We agree with the suggestion and we have colored the lines in (c) and (d) based on the TWI classes. We have used another color (pink) for groundwater boreholes to make them more easily identifiable. The simulations for wells 4473, 4476, and 4672 are shown in Fig.S14 in the Supplementary Material.

Comment #19: **Line 420** I found this conclusion somewhat vague. What does this result imply about the use of individual grid cells for comparison with groundwater observations? It could suggest that well-location selection needs to be done very carefully, that multiple neighbouring cells should be considered, or that some mismatch between real-world point observations and model grid cells is unavoidable. Since this issue is discussed more fully in Section 6.3, you could either clarify the implication here or leave the interpretation for the discussion section.

Answer: We agree with the suggestion. Both implications mentioned by the reviewer are indeed correct. We have clarified the implications of these results at L 433-437 in Result Section 5.3 and also pointed out to the readers to refer to Section 6.3 for further discussion and recommendations.

Comment #20: **Figure 6 / Line 432** I think the description of low-flow performance at Plima needs to be more careful. In Figure 6, there appears to be winter flow at S7, while the model simulates zero flow. Describing this as a "slight underestimate" seems misleading if the model is simulating no flow during periods when observations indicate some winter discharge. It also looks as though some winter periods may have measurement issues, while others, such as 2019–2020, show measurable winter flow. The 2021 winter drop also looks suspicious and may need explanation. Overall, I suggest revising the text to more explicitly acknowledge that the model does not fully capture winter low-flow dynamics at S7.

Answer: We agree and we have revised the text to be more explicit when interpreting the model performance during low-flow period at the Plima Station. The revision is provided at L 445-455 in Section 5.4.

Comment #21: **Line 457** The difficulty of the modelling task has already been mentioned several times. This sentence could probably be shortened or removed to avoid repetition.

Answer: We have shortened this sentence to avoid repetition in the revision.

Comment #22: **Figure 8** I found the distinction between percolation and groundwater recharge confusing. Percolation is often understood as water draining toward groundwater, while groundwater recharge is commonly interpreted as the amount of water entering the groundwater system. In the figure, groundwater recharge appears much lower than percolation, which makes me wonder whether "groundwater recharge" is actually being used to mean net groundwater storage change or net groundwater volume change. If so, a different term may be clearer. Otherwise, the current terminology may be misleading.

I also wondered whether the model assumes a constant head beneath the glacier. If so, was this described earlier in the manuscript? Is that assumption based on model structure, observations, or another constraint?

Answer: In the WaSiM model, the "Percolation" output is defined as the vertical flux into the groundwater system, while the "groundwater recharge" output is defined as the net change of

groundwater storage after calculating both vertical and horizontal inflow and outflow. We have adopted the same terminology as the WaSiM model and its user manual when showing this result. To avoid confusion, we have clarified the definition of "percolation" and "groundwater recharge" in the caption of the Figure 8 in the revision.

The groundwater heads are assumed to be constant beneath the glacierized cells, and this assumption is a built-in feature in the WaSiM model. Given that there is rare continuous observation of groundwater level beneath the glacier, it is challenging to estimate how groundwater may (or may not) fluctuate beneath the glacier covered region. We have clarified this assumption in the model description at L 236-237 in Section 4.2 in the revision.

Comment #23: **Figure 9** When discussing this figure in the text, could you also include the terms "gaining stream" and "losing stream"? These terms are widely used to describe groundwater–surface-water exchange and would help make the interpretation clearer.

Answer: Thanks for the suggestion. We have included the terms "gaining stream" and "losing stream" when interpreting the results of this figure in the Section 5.5.

Comment #24: **Line 515** The direct quotation feels somewhat unnecessary. The same point could likely be paraphrased and cited instead.

Answer: We have paraphrased this point instead of a direct quotation in the revision at L 529-530 in Section 6.1.

Comment #25: Line 518 I would like to see a more detailed discussion of the implications of lateral flow. Why does this result matter? What physical conditions could explain the need to inhibit lateral flow? Alternatively, could this reflect limitations of the model structure or spatial discretization?

Answer: we answered this together with Comment #1.

Comment #26: **Line 525** You could also cite Aubry-Wake et al. (2024), who used shallow groundwater head observations in a fully integrated hydrological modelling study of glacier–groundwater interactions. That study was HRU-based rather than fully distributed, but it is still relevant to the use of groundwater observations for constraining glacierized basin models:

Aubry-Wake, C., McNamara, G., Somers, L. D., McKenzie, J. M., Pomeroy, J. W., & Hellström, R. (2024). *Sensitivity of surface water and groundwater contributions to streamflow in a tropical glacierized basin under climate change scenarios*. Environmental Research Letters, 19, 114036. https://doi.org/10.1088/1748-9326/ad7c68

Answer: Thanks for mentioning this relevant literature. We have discussed this work in our manuscript in the revision at L 539-546 in Section 6.1.

Comment #27: **Line 567** "Outrageous" feels too strong and informal. Possible alternatives include "unlikely," "physically improbable," "counterintuitive," or "inconsistent with established understanding of flow behaviour."

Answer: We have adopted "counter-intuitive" in the revision.

Comment #28: **Line 779** I really liked this section. It was clear and helpful.

Answer: Thanks for the appraisal of the value of our work. Glad to hear that.

References

Altmann, M., Fischer, P., Haas, F., Pfeiffer, M., Ramskogler, K., Rom, J., Kara-Timmermann, D., Himmelstoß, T., Heckmann, T., and Becht, M.: Simulation of soil erosion in the Little Ice Age glacier foreland of the Zufallferner (South Tyrol, Italy) using Erosion-3D, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-5734, https://doi.org/10.5194/egusphere-egu25-5734, 2025.

Hayashi, M.: Alpine Hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. Groundwater, 58 (4), 498-510. https://doi.org/10.1111/gwat.12965, 2019.

McGuire, K. J., Klaus, J., and Jackson, C.R.: JAMES BUTTLE REVIEW: Interflow, subsurface stormflow and throughflow: A synthesis of field work and modelling, Hydrological Processes, 38 (9), e15263. https://doi.org/10.1002/hyp.15263, 2024.

Schulla, J.: Model Description WaSiM (Water balance Simulation Model), http://www.wasim.ch/downloads/doku/wasim/wasim_2024_en.pdf, 2024.

Somers, L. D. and McKenzie, J. M.: A review of groundwater in high mountain environments. Wiley Interdisciplinary Reviews: Water, 7(6), e1475. https://doi.org/10.1002/wat2.1475, 2020.

Xie, S., Zeng, C., Zhang, F., Wang, G., and Xiao, X.: Defining lateral subsurface flow and identifying its water sources in an alpine-permafrost hillslope. Catena, 236, 107765. https://doi.org/10.1016/j.catena.2023.107765, 2024.

Response to CC1:

This study implements a fully-integrated and fully-distributed surface-subsurface hydrological model using the WaSiM framework to examine water exchange processes in the high-elevation, glaciated Martell Valley in the Italian Alps. The authors utilize extensive field data, including rare in-situ high-elevation groundwater level observations—to constrain model parameters across snow, glacier, and subsurface modules.

This paper is worthy of publication as it addresses a significant gap in mountain hydrology: the quantification of surface-subsurface interactions in cryosphere-dominated environments. The integration of groundwater-level observations into a distributed, physics-based model for a high-alpine catchment is a notable technical achievement.

Answer: Thanks for your interest and appraisal of the value of our work.

Comment #1- While the model achieved satisfactory performance with subsurface lateral flow constrained to zero, this remains a counter-intuitive finding for steep alpine terrains. The authors acknowledge that their piezometers were placed in relatively flat floodplains. The finding might be a localized artifact of piezometer placement rather than a catchment-wide process. More discussion is needed on how this parameterization might affect the model's ability to simulate hillslope-to-stream connectivity in steeper parts of the catchment.

Answer: In the revision, we have enhanced discussion in more detail on the interpretation, implications, and physical explanations associated with this result in Discussion Subsection 6.2. Specifically, we enhanced clarification in the manuscript and marked in Figure 1 that the results are with high confidence in the upper subcatchments where the piezometers are installed, while we have also show the results for the whole catchment to provide an overview despite a lower confidence. The finding is unlikely to be a local artifact but a landscape-scale impact in the alpine headwater environment where all piezometers observations and simulations in this study supported this hypothesis. If sustained subsurface lateral flow from steep hillslopes was a dominant control, it would be expected to manifest in the downstream valley floor but not in the upper headwater landscapes (see the conceptual diagram in Fig. 1 in the work of van Tiel, et al. 2024).

In terms of the counter-intuitive finding, in alpine environments, we have steep slopes and can assume high contrast between hydraulic conductivities in the subsurface, which typically favors subsurface lateral flow (McGuire et al., 2024). However, we also have shallow soils in the study site, i.e. limited storage potential, which can favor percolation to aquifers, which in turn can be highly connected to the stream network and show high dynamics (Hayashi, 2019). Our model-based conclusion is hence a plausible representation of high alpine streamflow generation dynamics, in particular given that it reproduces streamflow (flux) and groundwater dynamics (storage).

Comment #2- The study notes a specific numerical requirement in WaSiM where the lower boundary of the soil layers must be identical to the lower boundary of the aquifer (calibrated here to 1.30 m). This 1.30 m constraint is quite shallow for an entire catchment. While it may represent the shallow porous aquifers observed, it automatically excludes any simulation of deeper groundwater flow paths. The authors should explicitly clarify if this thickness (1.30 m) is applied

uniformly across the entire 77 km2 catchment, including the valley floor where sedimentary deposits are likely much deeper.

Answer: The lower boundaries of the soil layers are calibrated with the observed groundwater heads at all piezometers (see Fig.S3-S8 in the Supplementary Material). Unfortunately there is no deep groundwater observations available in the valley floor where the deep groundwater may exist. We therefore applied the calibrated lower boundary depth in the gauged area to the ungauged area and presented the results in the manuscript explicitly in two ways: for the upper headwater landscapes (including Subcatchments S7,8,9) with piezometers observations where the results have higher confidence, and for the whole catchment where the results may contain higher uncertainty due to the exclusion of deep aquifers but are still useful to provide a catchment overview.

Excluding deep groundwater observations, however, does not compromise the merit of this study, as our aim is to apply an integrated hydrological model to achieve better partitioning of the liquid flux (sum of rainfall and melts) at the interface between surface and subsurface, i.e., a step forward compared with the commonly adopted linear reservoir or conceptual approach for subsurface representation. Despite the shallow soil formation, the soil zone plays a crucial role in the surface-subsurface water partitioning. Such fully-integrated hydrological modeling contributes to key hydrological process understanding in high Alpine landscapes.

Comment #3- In Section 6.3, the recommendation to match the TWI spatial resolution to the model resolution (25m×25m) is clear. It might be helpful to include a brief sentence advising how practitioners should handle areas where the 25m grid might smooth over critical local topographic features (like small moraines) mentioned earlier in the text.

Answer: The 25m×25m spatial resolution in fully-distributed hydrological modeling is considered high enough for investigating the hydrological processes in this study given the trade-off between the available observational information and the computational demands. For practitioners who are interested in the impact of even smaller local features, it is recommended to develop a model focusing on their interested local features with higher spatial resolution, if more information on their interested local features is available for performing such detailed study. However, this task will involve significantly increased computational efforts.

Comment #4- It would be beneficial to explicitly discuss if this "zero interflow" finding is a physical characteristic of the Martell Valley (e.g., due to high vertical conductivity in porous aquifers) or if it highlights a structural limitation in how current physics-based models distribute lateral flow in steep, shallow-soil alpine terrains. In Section 5.3, consider clarifying if "zero interflow" refers to the model parameter dr being set to 0, and briefly reiterate the physical implication for the reader.

Answer: In the revision, we have specifically enhanced the discussion about the physical explanation, interpretation, and implications of the result regarding the simulated subsurface lateral flow in the Discussion Subsection 6.2. First of all, we would like to underline that our finding indicates that, within the current model structure and parameterization, explicitly simulated subsurface lateral flow (i.e. interflow, in the unsaturated domain) is not needed and even leads to a poorer reproduction of the observed streamflow and groundwater dynamics. This

can be a model artefact, primarily because of model complexity. Our systematic testing, however, suggests that this is unlikely to be the case (see Fig. S9-S13 in the Supplementary Material).

Regarding the physical characteristics, as answered in Comment #1, in alpine environments, we have steep slopes and can assume high contrast between hydraulic conductivities in the subsurface, which typically favors subsurface lateral flow (McGuire et al., 2024). However, we also have shallow soils in the study site, i.e. limited storage potential, which can favor percolation to aquifers, which in turn can be highly connected to the stream network and show high dynamics (Hayashi, 2019). Our model-based conclusion is a plausible representation of high alpine streamflow generation dynamics, in particular given that it reproduces streamflow (flux) and groundwater dynamics (storage).

The parameter dr is a highly sensitive parameter that governs the subsurface lateral flow, but it is not the only governing parameter. The physical meanings of all relevant parameters that govern the flow components are summarized in Table 4 in the manuscript.

Comment #5- As you are working on surface-subsurface hydrological model I do strongly recommend to broad your literature review and cite "Assimilation of sentinel-based leaf area index for modeling surface-ground water interactions in irrigation districts"

References

Hayashi, M.: Alpine Hydrogeology: The critical role of groundwater in sourcing the headwaters of the world. Groundwater, 58 (4), 498-510. https://doi.org/10.1111/gwat.12965, 2019.

McGuire, K. J., Klaus, J., and Jackson, C.R.: JAMES BUTTLE REVIEW: Interflow, subsurface stormflow and throughflow: A synthesis of field work and modelling, Hydrological Processes, 38 (9), e15263. https://doi.org/10.1002/hyp.15263, 2024.

Schulla, J.: Model Description WaSiM (Water balance Simulation Model), http://www.wasim.ch/downloads/doku/wasim/wasim_2024_en.pdf, 2024.

van Tiel, M., Aubry-Wake, C., Somers, L., Andermann, C., Avanzi, F., Baraer, M., ... & Yapiyev, V. (2024). Cryosphere–groundwater connectivity is a missing link in the mountain water cycle. Nature Water, 2(7), 624-637.