Comment #1: The reliance on manual calibration for a complex, fully-distributed model with numerous parameters is a significant limitation. While the authors justify this choice due to computational constraints and data availability, the manuscript would benefit from a more thorough discussion of the implications of manual calibration on parameter uncertainty and model robustness. For instance, how sensitive are the key findings (e.g., minor role of subsurface lateral flow) to parameter choices? A sensitivity analysis, even if limited, could strengthen the credibility of the results.

Comment #2: The multi-signal calibration approach is a strength, but the sequential calibration process (surface to subsurface) may introduce biases. The authors should discuss potential dependencies between modules (e.g., snow module influencing groundwater recharge) and how these were addressed to ensure consistency across the calibration steps.

Answer to #1 and #2: Thanks for the suggestion. We calibrated the physics-based hydrologic model manually by following a top-down approach and module by module. We first performed manual sensitivity analysis on the key parameters in each module, and then focused on calibrating the sensitive parameters in detail. The insensitive parameters are assigned with the default values suggested in the model user manual or adopted from the literature. We will add the results of the suggested sensitivity analysis on the key parameters and strengthen the discussion of the adopted approach in the revised version.

The sequential calibration of module by module is a commonly adopted logical procedure to calibrate such a fully-distributed physics-based hydrological model. The module by module calibration offers a good diagnostic power, as it isolates which modules (e.g., snowmelt, glacier melt) are causing discrepancies between observation and simulation. This allows an incremental validation, as each module can be tested and validated before integrating with the next. By doing so, errors in specific processes (e.g., snow or glacier melt) can be addressed without compromising other well-performing modules. Through the simplification of the calibration strategy, the parameter interactions are reduced, which leads to more stable model results.

Despite the calibration is sequential, we rerun the whole model (including all modules) each time when a parameter in a module is perturbed, and we focus on the model performance to the observed variables of that module. For example, when we calibrate a parameter in the snow module, we run the whole hydrological model and all temporal and spatial hydroclimatic outputs are produced, but we focus on the model performance compared to the snow water equivalent at the observed stations and spatial snow coverage. In this way, the hydrological processes between the modules are interconnected and the consistency is ensured. We will improve the articulation of the model calibration in the revised manuscript.

Comment #3: The finding that subsurface lateral flow plays a minor role in streamflow generation is intriguing but requires further scrutiny. The assumption of homogeneous subsurface properties (e.g., hydraulic conductivity, storage coefficient) across the catchment may oversimplify the complex geology of the Martell Valley, which is noted to be heterogeneous (Section 2). This assumption could bias the model toward underestimating lateral flow. The authors should explore whether spatially variable subsurface parameters, informed by available geological data, could alter this conclusion.

Answer: The subsurface of our model is not fully homogeneous. We applied essential heterogeneity in the subsurface by calibrating the hydraulic conductivity in each soil layer and for each soil type. More detailed heterogeneity of the subsurface characteristics in finer spatial resolution is challenging to apply, as this information is vastly unavailable in the high-elevation catchments with glaciers. 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. Assigning heterogeneous subsurface properties horizontally and vertically in the ungauged areas involves numerous assumptions. We therefore applied a conservative approach to ensure the essential subsurface heterogeneity. Exploring the impact of spatially variable subsurface parameters is out of scope of this study.

Comment #4: The model's inability to reproduce the strong winter recession at borehole ID 4478 (Section 5.3) suggests limitations in capturing preferential flow paths or other subsurface processes. The manuscript would benefit from a deeper discussion of alternative mechanisms (e.g., macropore flow, fractured bedrock) that could explain this discrepancy, potentially supported by literature or additional field observations.

Answer: We will strengthen the discussion on these likely alternative mechanisms in the revised manuscript.

Comment #5: The challenges of integrating point-scale groundwater observations into a distributed model are well-articulated, but the proposed solutions (e.g., using TWI to guide piezometer placement, comparing neighboring cells) need more rigorous evaluation. For instance, how representative are the TWI-based recommendations for other high-alpine catchments with different topographic or geologic characteristics? A sensitivity analysis of TWI resolution or comparison with other topographic indices could enhance the generalizability of these recommendations.

Answer: We suggest calculating TWI as an additional information to support the decision on where to install the piezometers, besides the expert knowledge and field information. The TWI is mainly derived based on the Digital Elevation Model data and does not relate to the geological data. The spatial resolution of the TWI (or the hydrological model resolution) should be determined by the individual site characteristics such as topography. In our study, we tested the spatial resolution of 25x25m, 50x50m, and 100x100m, and we finally adopted 25x25m by balancing the trade-off between the computational intensity and the site characteristics. Future studies are welcome to test this approach and adopt a reasonable spatial resolution based on their site condition and research aims.

Comment #6: The manuscript highlights the mismatch between observed and modeled river networks due to DEM uncertainties (Section 5.6.3). This issue could significantly affect groundwater-surface water interactions, yet it is only briefly addressed. A quantitative assessment of DEM uncertainty (e.g., comparing simulations with different DEM resolutions) would strengthen the discussion and provide more concrete guidance for future studies.

Answer: We did test different spatial resolutions of 25x25m, 50x50m, and 100x100m in the hydrological model beforehand and finally adopted the 25x25m, by balancing the trade-off between the computational resources and the details of the topography, which is reasonably captured by this DEM resolution. More detailed DEM leads to significantly higher computational demand, which is not plausible and unlikely to significantly alter the conclusions.

Comment #7: The authors note that the Martell Valley is relatively dry compared to other Alpine catchments (Section 5.6.2), which may limit the applicability of findings to wetter environments. Similarly, the lithology (crystalline bedrocks, shallow soils) may not be representative of other high-alpine settings. The discussion should more explicitly address the conditions under which the key findings (e.g., minor role of lateral flow, rapid groundwater response) are likely to hold, potentially by comparing with studies in contrasting catchments.

Answer: We will strengthen the discussion and articulation of the generalisation of the research findings in terms of climatic and hydrogeologic conditions.

Comment #8: The manuscript claims that the rapid groundwater response is rarely simulated by hydrological models (Section 6), but this statement requires more substantiation. A brief review of other physics-based models (e.g., HydroGeoSphere, ParFlow) and their ability to capture such dynamics would contextualize the novelty of WaSiM's performance and clarify the need for improved subsurface parameterization.

Answer: We agree with this point and will strengthen the discussion by comparing this study with similar modeling efforts in the high alpine studies.

Comment #9: The underestimation of winter baseflow (Sections 5.4, 6) is attributed to shallow river channels and homogeneous subsurface parameterization, but observational uncertainties in low-flow measurements (e.g., sensor limitations in freezing conditions) are also significant (Section 5.6.4). The manuscript should more clearly disentangle model limitations from observational uncertainties, possibly by discussing the reliability of winter discharge data or exploring alternative data sources (e.g., tracer studies) to validate baseflow contributions.

Comment #10: The claim that baseflow contributes significantly to winter streamflow (up to 40% in some subcatchments, Section 5.4) is compelling but relies on model simulations rather than direct observations. Additional evidence, such as isotopic or chemical tracers, could corroborate this finding and enhance confidence in the model's representation of baseflow dynamics.

Answer to #9 and #10 on additional evidence: We will strengthen the discussion regarding the winter baseflow from shallow groundwater by considering the suggested points. Our water stable isotope data of groundwater and river sampled in winter low-flow condition (2022/23, 2023/24) do show consistent values and thus provide another evidence that shallow groundwater provides winter baseflow. We, however, did not show these data in the manuscript as they were not sampled regularly in high frequency. We agree that this aspect will be better articulated in the revised version.

Comment #11: The introduction is comprehensive but lengthy, with some repetition (e.g., challenges of alpine hydrology are mentioned multiple times). Streamlining the introduction to focus on key gaps and the study's objectives would improve readability.

Answer: We will improve the readability of the introduction by considering the given suggestion.

Comment #12: Section 5.6 is titled "Challenges and opportunities for modeling high-alpine glaciated environment," but it primarily discusses challenges. Explicitly addressing opportunities (e.g., leveraging remote sensing, integrating machine learning for parameter estimation) would balance the narrative and highlight future research directions.

Answer: Thanks for the suggestion. We will address this aspect by enhancing the discussion on the future opportunities, such as the newly emerged approach on integrating machine learning technique into physics-based modelling.

Comment #13: The use of abbreviations (e.g., PEQ, TWI, DTW) is frequent, and a glossary or table defining these terms would aid readers unfamiliar with the terminology.

Answer: Thanks for the suggestion, but there are actually only a few abbreviations in this study, and we feel that a glossary table might be unnecessary in this case.

Comment #14: Figure 8 is visually rich but overwhelming due to the number of panels. Consider splitting it into two figures (e.g., one for percolation/recharge, another for groundwater level/exfiltration/infiltration) or using a subset of months to improve clarity.

Answer: We acknowledge this suggestion but would prefer to keep them in a panel to enable to intercompare the results between the months and the inter-related variables.

Comment #15: Table 3 and Table 4 list calibrated parameters but lack units for some parameters (e.g., "Scaling for max.deposition" in Table 3). Ensuring consistency in units and providing brief explanations for less intuitive parameters would enhance accessibility.

Answer: The unit of "Scaling for max.deposition" in Table 3 is already given right after this term; it is in mm. We ensured to provide all units for all parameters in the tables. For the ones that do not have a unit, we marked them as [-]. A brief explanation is provided after each parameter as well by following the model user manual to allow readers for cross-checking them in detail in the manual.

Comment #16: The caption for Figure 5 could clarify that panels (c-d) show simulations for multiple grid cells, as this is not immediately obvious from the figure alone.

Answer: We will revise this information in the caption for improving readability.

Comment #17: I highly recommend to discuss the SWAT-MODFLOW papers which integrates SW-GW and cite below paper:

Assimilation of sentinel-based leaf area index for modeling surface-ground water interactions in irrigation districts

Answer: Thanks for the suggested literature and we will consider it. However, the suggested literature focuses on coupled modeling in the irrigation fields, while our study focuses on the high alpine hydrology in the glaciated environment. Despite the topic of modeling, the research approach, focus, and site conditions are largely different.