

Response to Comments by Reviewer #1

General Comments:

The revised manuscript has improved substantially, and the authors have addressed the concerns raised by the previous reviewers, particularly with respect to the SWE-focused analysis. I also share some of the concerns noted by Reviewer #2 in the previous round, namely that minor revisions could further strengthen the discussion section. Overall, I find this manuscript close to being suitable for publication.

Response: We thank the reviewer for the positive assessment of our revised manuscript and for the constructive comments in this round of review. We have carefully addressed each of the specific points raised below, with particular attention to strengthening the discussion section (Section 6) through additional analyses of rain-on-snow-related peak runoff behavior (Q1), expanded contextual information on the subbasin classifications underlying Table 3 (Q2 and Q3), and an extended discussion of frozen-soil process representation, including preferential flow through macropores (Q4). We also revised Figure S1 for improved clarity (Q5). We believe these revisions have further strengthened the manuscript and thank the reviewer again for their time and valuable feedback throughout the review process.

Q1. Comment 1 (Lines 495–505): Could the author comment on the ROS events observed in the study region? For example, information on changes in peak runoff magnitude or timing at sites where INF improved performance would provide valuable insight.”

Response: This is a valuable suggestion, and we conducted an additional analysis of annual peak runoff at the subbasins where the INF experiment improved overall performance. For each of these subbasins, we computed the annual peak flow (the maximum monthly total runoff within each water year, Oct–Sep) for both the default and INF simulations, and compared each to the observed annual peak. Years in which the simulated peak timing differed from the observed peak timing by more than one month were excluded, since the resulting bias would reflect a mismatch in timing rather than in magnitude at a common flood event. We then calculated the relative peak-flow bias $(\text{peak_sim} - \text{peak_obs})/\text{peak_obs} \times 100$ for each remaining year and averaged it across the 10-year study period (2014–2023) for each subbasin.

This analysis showed that peak runoff was improved (i.e., bias magnitude was reduced relative to the default) at 28 of the 32 INF-improved subbasins. The 10-year average improvement ranged

from 1.0% to 72.0%, with a mean of 25.1%. We added a sentence summarizing this result to Section 6.1, immediately following the discussion of the INF scheme's effect on frozen soil permeability and surface runoff partitioning (see revised text below).

Revised Paragraph (Pages 20 – 21, Lines 501 – 507) with key changes in bold italic

This reduced diffusivity effectively limits infiltration capacity under frozen soil conditions, causing a greater fraction of precipitation and snowmelt to be partitioned into surface runoff rather than subsurface flow. *Among the subbasins where INF improved overall performance, peak runoff estimates also improved at 28 sites, with an average improvement of 25.1% over the 10-year period, ranging from 1.0% to 71.0%.* This is particularly important for improving runoff estimates during ROS events, which are often underestimated due to excessive infiltration by the default scheme.

Q2 (Table 3): A brief description of the dominant climate, land cover, and snow classifications in the basin (i.e., the relative area covered) would be helpful contextual information to accompany this table.

Response: We appreciate this suggestion. Rather than reporting basin-wide area percentages, we chose to summarize the distribution of the 50 HCDN subbasins themselves across each classification, since this directly reflects the sample underlying the recommendations in Table 3. We have added a new supplementary figure (Figure S3) showing the number and proportion of subbasins in each climate, land cover, and snow category, along with a sentence in Section 6.2 noting that classifications with relatively few subbasins warrant cautious interpretation. The added text and figure are shown below.

Revised Paragraph (Page 21, Lines 516 – 521) with key changes in bold italic

Two key criteria were used to recommend alternative parameterization schemes (Table 3). First, at least 50 percent of the subbasins within each environmental classification had to show improvement relative to the default to ensure sufficient representation. Second, the experiment needed to demonstrate an average KGE improvement of at least 0.1 within that classification. *The distribution of the 50 HCDN subbasins across these classifications is summarized in Figure S3. Classifications represented by relatively few subbasins warrant cautious interpretation.*

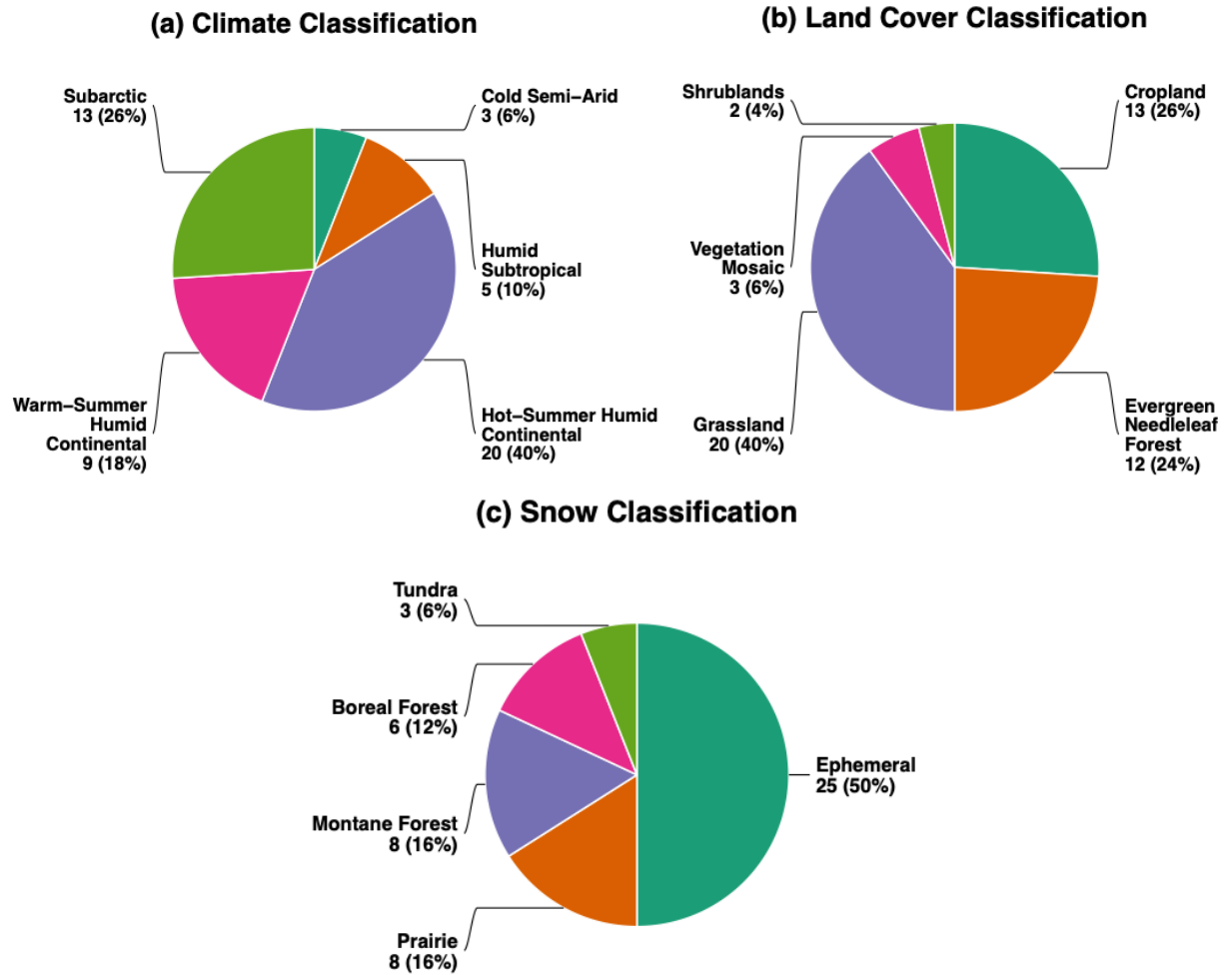


Figure S3. Number and proportion (%) of each classification for (a) climate, (b) land cover, and (c) snow among the 50 HCDN subbasins.

Q3 (Section 6.2): It appears that the default configuration performed better for colder or forested locations. For colder environments, could this be because the STC configuration does not yield significant improvements in runoff, as discussed in Section 5.4? For forested environments, might this suggest that other alternative parameterizations (e.g., canopy-related processes) could further improve model performance?

Response: We thank the reviewer for this insightful comment. We agree with both points raised. For colder classifications, the retention of the default scheme is indeed consistent with the STC experiment's limited overall runoff improvement discussed in Section 5.4. For forested classifications, we agree that canopy-related processes are a promising avenue for further improvement, as none of the four experiments evaluated in this study specifically targeted canopy

interception or dynamic vegetation parameterizations. We have revised the text in Section 6.2 to clarify both points, as shown below.

Revised Paragraph (Page 22, Lines 528 – 533) with key changes in bold italic

In contrast, the STC experiment does not meet the criteria in any classification, reflecting its limited overall runoff improvement discussed in Section 5.4. The default is retained for Subarctic climates as well as Evergreen Needleleaf Forest, Shrublands, Montane Forest, Boreal Forest, and Tundra environments. ***Among these, the forest-related classifications, Evergreen Needleleaf, Montane, and Boreal Forest, may benefit from alternative canopy-related parameterizations such as interception or dynamic vegetation, which none of the four experiments targeted.***

Q4 (Section 6.4): An additional limitation related to frozen soil processes that could be discussed is the representation of preferential flow through macropores (e.g., <https://doi.org/10.1029/2022WR033075>). Could the authors also comment on the finding from this study that increased permeability parameterization for frozen soils improves Noah-MP performance across the continental United States?”

Response: We appreciate the reviewer pointing us to this reference. We agree that preferential flow through soil macropores is an important limitation of the current frozen-soil parameterization in Noah-MP that was not addressed in the original manuscript. We have revised Section 6.4 to explicitly discuss this limitation and to comment on the findings of Agnihotri et al. (2023), who demonstrated that higher frozen-soil permeability to better account for preferential flow significantly improved Noah-MP Spring streamflow simulations across the Mississippi River Basin. We have incorporated this discussion within our broader recommendation for a process-oriented evaluation framework, in which frozen soil permeability parameterization serves as a concrete example of how process-focused scheme comparisons can identify substantial performance improvements within a single physical process.

Added Paragraph (Pages 23 - 24, Lines 594 – 606)

As an alternative strategy to reduce the complexity arising from interactions among multiple physical processes, future studies could adopt a process-oriented evaluation framework that

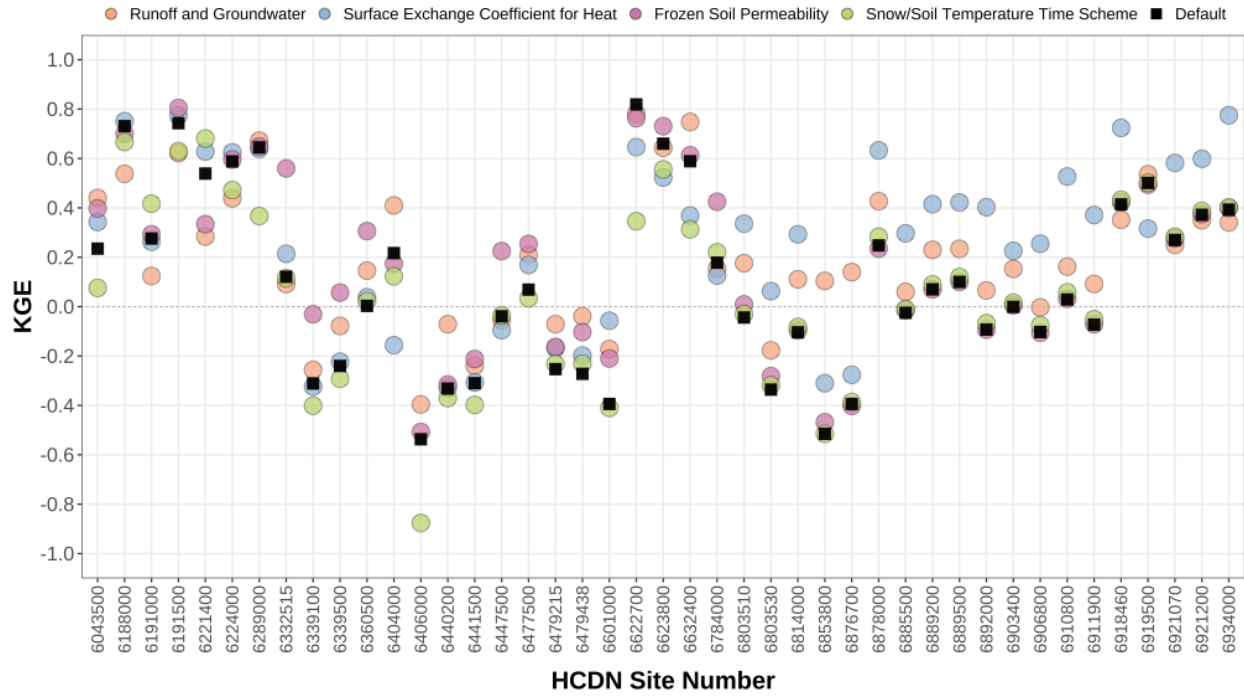
systematically compares multiple parameterization schemes within a single physical process. This approach would help isolate process-specific uncertainties while identifying parameterizations that provide the greatest performance improvements. For example, the current frozen-soil parameterization in Noah-MP does not fully account for preferential flow through soil macropores, which may contribute to biases in infiltration and runoff generation during frozen conditions. Agnihotri et al. (2023) evaluated combinations of supercooled liquid water and soil hydraulic property schemes in Noah-MP over the Mississippi River Basin and demonstrated that representing higher frozen-soil permeability to better account for preferential flow significantly improved spring streamflow simulations. In addition, Abolafia-Rosenzweig et al. (2026) tested 256 combinations of Noah-MP physics options across 790 SNOTEL stations in the western United States and identified an updated configuration that reduced both snow accumulation and ablation rate errors relative to the default. Such process-focused evaluations could be extended to other land surface processes to further improve the physical realism and predictive capability of Noah-MP.

Added Reference

- Agnihotri, J., Behrangi, A., Tavakoly, A., Geheran, M., Farmani, M. A., and Niu, G.-Y.: Higher frozen soil permeability represented in a hydrological model improves spring streamflow prediction from river basin to continental scales, *Water Resour. Res.*, 59, e2022WR033075, <https://doi.org/10.1029/2022WR033075>, 2023.
- Abolafia-Rosenzweig, R., He, C., Ikeda, K., Liu, C., Lin, T.-S., Barlage, M., Xue, L., Rasouli, K., Zhang, Y., Newman, A., and Dudhia, J.: Revisiting snow water equivalent bias in the Noah-MP land surface model in the western U.S., *Water Resour. Res.*, 62, e2025WR042586, <https://doi.org/10.1029/2025WR042586>, 2026.

Q5 (Figure S1): I suggest using a different symbol for the default to distinguish it from the experimental configurations.

Response: We agree with this suggestion. In the revised Figure S1, the default configuration is now shown as a black square, clearly distinguishing it from the experimental configurations. We thank the reviewer for this helpful suggestion, which has improved the clarity of the figure.



Revised Figure S1. KGE values of the five Noah-MP parameterization schemes (Default, RUN, SFC, INF, and STC) for each of the 50 HCDN subbasins in the Missouri River Basin. The x-axis represents the HCDN site and the y-axis represents the KGE value computed against USGS monthly streamflow for 2014–2023.