

Responses to Comments from Reviewer #1

General Comments:

This manuscript provides a detailed evaluation and comparison of runoff and snow simulations using different parameterization schemes of the Noah-MP land surface model in the Missouri River Basin. This work holds scientific value for disaster prevention and mitigation in the basin. However, I have several major concerns, detailed below.

Response: We sincerely thank the reviewer for recognizing the scientific value of our work for disaster prevention and mitigation in the Missouri River Basin and appreciate the thorough review and constructive comments.

To address the reviewer's comments, we have revised all parameterization experiments to use separate, experiment-specific spin-up simulations as recommended, revised Section 6.1 conclusions to reflect the basin-specific context of our findings with supporting references on region-dependency, and expanded Section 6.4 with more in-depth discussions of the identified limitations. We believe these revisions will substantially improve the manuscript. Please see detailed responses below.

Q1. Line 150: I find it unreasonable to use the same initial conditions for the alternative parameterization experiments as for the default simulation. This approach likely affects some of the study's conclusions. I recommend performing a separate spin-up for each experiment. If computational resources are a constraint, the period from January 1, 1995, to December 31, 1999, could be used as the model spin-up phase.

Response: We appreciate this important comment and have fully revised our simulations in response. As recommended, all results presented in the manuscript are now based on experiment-specific spin-up runs. Specifically, each parameterization experiment was independently initialized from January 1, 1980, and run through December 31, 2013, using its respective parameterization scheme settings as a dedicated spin-up setup, ensuring that the initial modeled states entering the 2014–2023 analysis period are fully consistent with each experiment's parameterization configuration. While the overall conclusions and narrative direction of the study remain unchanged, the experiment-specific spin-up produced updated quantitative results throughout the manuscript, including revised KGE statistics, figures, and tables. The updated values better reflect the physical consistency between each parameterization configuration and its

corresponding model state, as the initial conditions now evolve under the same parameter settings used in the analysis period. We therefore consider these revised results to be more physically meaningful than those based on a shared default spin-up. The Methods section (Section 4.1) has been revised as follows.

Revised manuscript (Page 6, Lines 150 – 155) with key changes in bold italic

Initial conditions for all experiments were established through an experiment-specific spin-up procedure. *Each parameterization experiment (default and all four alternatives) was independently initialized from January 1, 1980, and run through December 31, 2013,* using NLDAS-2 forcing data with its respective parameterization scheme settings. The model states at the end of 2013 from individual spin-ups were used as initial conditions for the subsequent 2014–2023 analysis, ensuring that the model state entering the analysis period is fully consistent with each experiment's parameterization configuration.

Q2. Section 6.1: The study concludes that the alternative schemes address three specific limitations of the default scheme. I find these conclusions to be overly strong. Previous research suggests that the performance of land surface process schemes is often region-dependent.

Response: We agree that our conclusions should be more carefully framed to reflect the hydroclimatic conditions specific to the study area. We have revised Section 6.1 to more explicitly acknowledge that the identified improvements may be specific to the Missouri River Basin. As the performance of land surface parameterization schemes is known to vary by region, further investigation would be needed to assess whether these findings generalize to areas with different climates, land cover characteristics, or snow regimes.

A brief note in Section 6.1 directs readers to Section 6.4, where this limitation is discussed in greater depth, including supporting evidence from other studies. We have added the below paragraph in Section 6.1.

Revised manuscript (Page 21, Lines 507 – 512) with key changes in bold italic

It is important to note, however, that the identified limitations of the default scheme and the improvements offered by alternative schemes reflect the specific hydroclimatic conditions of the Missouri River Basin. *These findings may not be directly transferable to regions with*

substantially different climate regimes, land cover characteristics, or snow dynamics, as the performance of land surface parameterization schemes is known to be region-dependent (You et al., 2020; Hu et al., 2023; Chang et al., 2020; Hamitouche et al., 2025; Gan et al., 2019; Li et al., 2022). A more detailed discussion of these limitations is provided in Section 6.4.

Q3. Section 6.4: I agree with the two limitations identified in this section: first, that the scheme recommendations may not be directly applicable to other regions, and second, that the interactions between different land surface physical processes are not considered. I strongly recommend providing a more in-depth discussion of these points rather than merely noting them in passing.

Response: We appreciate this suggestion. Section 6.4 has been expanded to provide a more in-depth discussion of the two identified limitations. For the transferability issue, Section 6.4 now cites evidence from other studies that demonstrate region-dependent scheme performance and discusses how basin-specific characteristics of the Missouri River Basin may constrain the applicability of our recommendations to other regions. The revised manuscript text for this section is provided below.

Revised manuscript (Page 22 - 23, Lines 556 – 567) with key changes as bold & italic

While this study advances understanding of parameterization selection in the Noah-MP model, several limitations should be acknowledged regarding the generalizability of our findings. The current analysis is confined to the Missouri River Basin for the period from 2014 to 2023, which may limit the broader applicability of the recommended parameterization scheme. ***The literature consistently shows that the performance of land surface parameterization schemes varies considerably across different climatic and geographic settings (Gan et al., 2019; Li et al., 2022). For example, Hu et al. (2023) showed that the Koren99 frozen soil scheme (INF experiment) is particularly effective in permafrost regions of the Tibetan Plateau. Chang et al. (2020) found that the Chen97 surface exchange coefficient scheme (SFC experiment) improved ET simulations over subtropical forests in South China. Hamitouche et al. (2025) further documented substantial global variability in the performance of the TOPMODEL equilibrium water table scheme (RUN experiment). Taken together, these findings underscore that scheme recommendations derived from one region should be applied with caution elsewhere. Further***

validation across diverse geographical settings is therefore necessary to establish the robustness of the alternative parameterization approaches identified in this study.

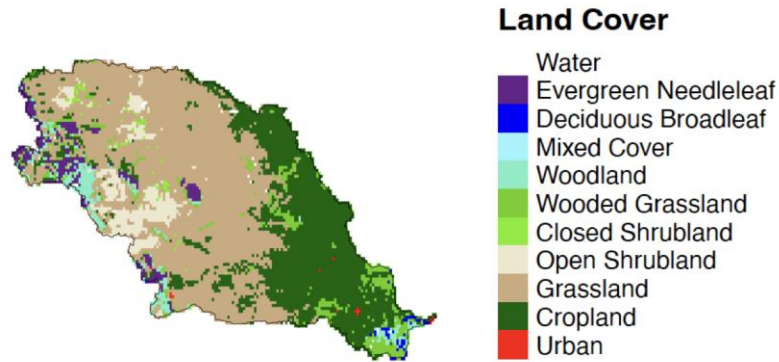
For the process interaction limitation, Section 6.4 now provides a more detailed discussion of how coupling between different parameterization schemes can produce complex, non-linear effects on model performance. Specifically, the revised text illustrates potential interactions among the alternative schemes evaluated in this study, such as how the SFC scheme's reduction in ET may alter soil moisture availability in ways that amplify or dampen the runoff response of the INF scheme, and how SWE improvements from the STC scheme may interact with the RUN scheme's groundwater dynamics through changes in snowmelt timing and recharge rates. The revised text also draws on prior multi-physics ensemble studies showing that certain scheme combinations can produce emergent behaviors not predictable from individual scheme evaluations alone, and that compensating errors between schemes can be disrupted when one scheme is changed (Hong et al., 2014; Gan et al., 2019; Cuntz et al., 2016). The revised manuscript text for the second limitation is provided below.

Revised manuscript (Page 23, Lines 577 – 584) with key changes in bold italic

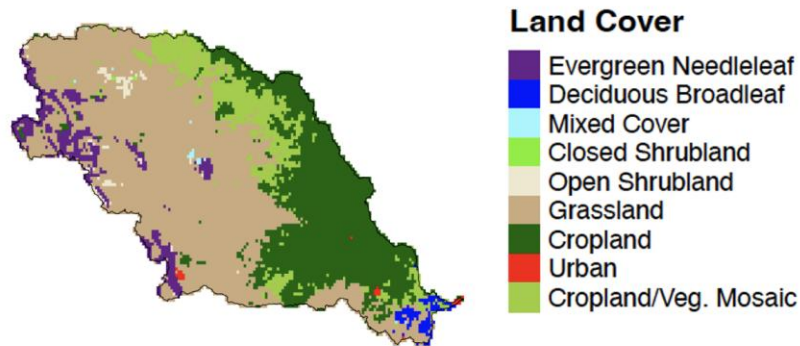
The Noah-MP model offers multiple physics parameterization schemes across various land surface processes, where interactions among different schemes can substantially affect model performance. This study focused on individual parameter sensitivity but did not examine how alternative physics options interact when combined. ***Such interactions can be complex and non-linear: changes in ET, soil moisture, or snowmelt driven by one scheme may amplify or dampen the response of another, and compensating errors between schemes can be disrupted when any one of them is changed (Cuntz et al., 2016). Prior multi-physics ensemble studies have reported that certain scheme combinations produce emergent behaviors not predictable from individual scheme evaluations alone (Hong et al., 2014; Gan et al., 2019).*** Future investigations should implement methodologies that evaluate different combinations of physics schemes, thereby delivering a more comprehensive understanding of model uncertainty and overall performance.

Q4. In Figure 1, does the land cover map represent the actual dataset used in the simulations? Was the IGBP land cover classification scheme employed?

Response: Thank you for your comment. Yes, we used the IGBP land cover classification scheme in our simulations. We have revised Figure 1 and updated the figure caption to explicitly state that the land cover panel (d) uses the IGBP land cover classification as implemented in the NLDAS vegetation dataset. We have also added a clarifying sentence in the data description section (Section 3.1). In addition, all land cover-related figures, tables, and statistics in the manuscript have been updated to ensure consistency with the IGBP classification scheme used in the simulations. The overall findings and interpretations remain unchanged, as the differences between the two land cover classifications are relatively small across the study domain. The updated Figure 1(d) is shown below.



(a) University of Maryland Global Land Cover



(b) IGBP Land Cover

Revised Figure 1(d). Land Cover Classification Revised from University of Maryland to IGBP

Q5. In Figure 4a, the default scheme fails to accurately reproduce the seasonal variation in groundwater storage, whereas it performs well in Figure 4b. Could the authors please attempt to explain this discrepancy?

Response: Thank you for pointing this out. The contrasting groundwater behavior between Fig. 4a and 4b reflects differences in how each subbasin interacts with the default groundwater scheme. In the default TOPMODEL with groundwater storage scheme (Niu et al., 2007), groundwater recharge is governed by the hydraulic conductivity and matric potential of the lowest soil layer, making groundwater storage dynamics sensitive to the moisture state of the deepest soil layer only. When evapotranspiration is overestimated, moisture in the lower layers is depleted, suppressing recharge and dampening groundwater storage variability. In the Battle Creek Keystone Basin (Fig. 4a), this mechanism produces limited seasonal groundwater variability. In contrast, the Gasconade River Basin (Fig. 4b) exhibits stronger seasonal forcing from precipitation and ET that drives sufficient moisture flux even to the lowest soil layer, allowing the default scheme to capture seasonal groundwater dynamics more adequately. A primary reason for this discrepancy is therefore the scheme's sensitivity to lower-layer soil moisture conditions, which varies depending on the hydroclimatic setting of each subbasin.

To address this challenge, we systematically examine climate, land cover, and snow characteristics associated with conditions under which alternative schemes outperform the default in Section 6.2, and provide data-driven recommendations for parameterization scheme selection. Notably, the RUN scheme, which employs the TOPMODEL-based approach with an equilibrium water table assumption, is recommended across multiple environmental classifications, including Cold Semi-Arid, Hot-Summer Humid Continental, Warm-Summer Humid Continental climates, Cropland land cover types, and Ephemeral and Prairie snow regimes (Table 3). This broad effectiveness of the RUN scheme is consistent with its improved representation of shallow groundwater dynamics relative to the default. Nevertheless, the underlying mechanisms of the relationships identified in Section 6.2 warrant further investigation. Additional analysis of the RUN scheme may help disentangle the relative contributions of surface runoff partitioning and groundwater dynamics to overall model performance, and is recommended for future research.

Q6. Is "ROS" in the manuscript an abbreviation for "rain-on-snow"? Please clarify.

Response: Yes, ROS stands for "rain-on-snow." We have defined this abbreviation at its first appearance in the manuscript and ensured its consistent usage throughout the manuscript.

Responses to Comments from Reviewer #2

General Comments:

Although the paper provides valuable implications for operational hydrology over the Missouri River Basin and the methodology section is strong, clearly explaining the differences between the alternative schemes and the default scheme, more effort is needed to make the paper suitable for publication in this journal.

First, the study does not advance the development of model parameterization schemes. Second, there is no thorough analysis explaining why SWE or runoff differ when the alternative parameterization schemes were used by providing changes in key intermediate variables. For example, in the case of the snow/soil temperature parameterization scheme, differences in parameterization may first influence ground temperature, which then impacts snowmelt and ultimately SWE. If the authors could show changes in ground temperature and snowmelt and link them to changes in SWE, the underlying processes would become clearer to readers, and this could also provide new insights. Similarly, in other cases, showing changes in key intermediate variables that influence runoff, either through SWE changes or through other pathways, would substantially improve the quality of the paper.

Most paragraphs in the Discussion section should be removed and rewritten, as, in my opinion, they are not persuasive in explaining the differences in runoff and SWE in the Results section of the paper. More effort should be devoted to explaining how differences in parameterization lead to differences in runoff and SWE by presenting changes in key intermediate variables. For detailed questions and comments, please see the attached file.

Response: We thank the reviewer for recognizing the valuable implications of our work for operational hydrology and for acknowledging the strength of the methodology section. We greatly appreciate the constructive feedback and have substantially revised the manuscript in response to all comments.

Regarding the concern that the study does not advance the development of model parameterization schemes, we respectfully clarify that the primary objective of this study is to provide evidence-based guidance for parameterization scheme selection across diverse hydroclimatic conditions in the Missouri River Basin, rather than to develop new schemes. We believe the evaluation results offer meaningful insights in operational settings where selecting the appropriate model configuration is critical for achieving optimal regional performance.

Regarding the request for more thorough analysis with intermediate variables, we have expanded the analyses of key intermediate variables. Soil moisture has been added to the representative subbasin time series in Figure 4, enabling direct interpretation of how each alternative scheme modifies antecedent conditions and subsequently influences runoff generation. GLEAM-based ET validation has been incorporated into Figure 4, providing observational support for the ET overestimation identified in the SFC experiment. Two additional figures (Figures 7 and 8) have been added to investigate why SWE improvements under the STC experiment do not always translate into runoff improvements. The Discussion section has been substantially rewritten throughout to provide more process-based explanations linking parameterization differences to changes in simulated SWE and runoff.

We believe these revisions collectively address the reviewer's concerns and substantially improve the quality of the manuscript. Detailed responses to each individual comment are provided below.

Q1. (Page 3, Lines 72-74) Does each of the four parameterization schemes govern one of the four processes? Or does each of them govern all four processes?

Building on this context, this study focuses on systematically evaluating four alternative Noah-MP parameterization schemes governing runoff and groundwater, surface exchange coefficients, frozen soil permeability, and snow/soil temperature dynamics across 50 minimally regulated subbasins across the Missouri River Basin.

Response: Each of the four parameterization schemes governs one specific process: (1) runoff and groundwater, (2) surface exchange coefficients, (3) frozen soil permeability, and (4) snow/soil temperature dynamics. Each experiment modifies a single parameterization scheme while keeping the other three at their default settings. To clarify this, we have revised the texts as follows:

Revised manuscript (Page 3, Lines 72 – 76) with key changes in bold italic

Within this context, we focus on systematically evaluating four alternative Noah-MP parameterization schemes, each governing one specific physical process: *(1) runoff and groundwater, (2) surface exchange coefficients, (3) frozen soil permeability, and (4) snow/soil temperature dynamics*. Each experiment modifies a single parameterization scheme while keeping

the other three at their default settings, enabling an isolated assessment of each scheme's contribution to model performance.

Q2. (Page 5, Line 126) I am wondering if you could provide the link from which you downloaded the 1-km UA SWE data. The literature you cite (Broxton et al., 2019) describes the 4-km UA SWE product rather than the 1-km version.

Response: Thank you for catching this error. We use the 4-km UA SWE product, not the 1-km version. We have corrected this in the revised manuscript.

Q3. (Page 10, Lines 278-279) You mean KGE=-0.414 represents the performance of the default parameterization schemes?

Response: No, KGE = -0.414 represents the performance of a benchmark prediction using the observation mean, not the default parameterization schemes.

Q4. (Page 11, Line 310) I wonder whether this is attributable to changes in SWE. Could you please add a component to Table 2 that indicates changes in SWE? The STC experiment produced higher surface, subsurface and total runoff than the default configuration, reflecting altered snow dynamics under its alternative snow/soil temperature time scheme.

Response: Thank you for this suggestion. As recommended, we have added the average annual maximum SWE to Table 2, enabling direct assessment of how differences in simulated snow storage relate to runoff variability across parameterization schemes. The updated Table 2 is provided below.

Revised Table 2. Basin-averaged, 10-year mean water balance components across five parameterization schemes for the entire Missouri River Basin. Components include precipitation (P), evapotranspiration (ET), snow water equivalent (SWE), change in groundwater storage (ΔGW), change in soil moisture (ΔSM), surface runoff (Q_s), subsurface runoff (Q_{sb}), and total runoff (Q_{total}).

Experiment	P (mm)	ET (mm)	SWE (mm)	ΔGW (mm)	ΔSM (mm)	Q_s (mm)	Q_{sb} (mm)	Q_{total} (mm)
Default	534.0	479.5	33.4	-0.3	-6.1	20.6	43.3	63.9
RUN	534.0	473.7	33.4	2.8	-5.6	26.8	40.7	67.5
SFC	534.0	463.6	36.9	0.2	-6.3	24.9	55.4	80.2
INF	534.0	476.9	33.4	-0.1	-6.0	27.9	38.8	66.7

STC	534.0	474.0	48.9	-0.3	-5.8	21.7	47.7	69.4
-----	-------	-------	------	------	------	------	------	------

Notably, the STC experiment produced the highest average annual maximum SWE (48.9 mm) among all schemes, substantially exceeding the default value (33.4 mm). This difference reflects the altered snow accumulation and melt timing under the implicit snow/soil temperature time scheme, which delays snowmelt and prolongs snowpack retention. The enhanced snow storage under the STC scheme is consistent with the higher surface, subsurface, and total runoff relative to the default. This result serves as a critical basis for interpreting the STC experiment's performance throughout the subsequent analysis, and for understanding why its positive impacts on runoff are most evident in snow-dominated subbasins where SWE is systematically underestimated by the default scheme.

Q5. (Page 11, Line 323) Could you please provide the percentage of improved sub-basins in each of the four experiments?

Response: Thank you for this suggestion. As recommended, we have included the percentage of improved subbasins for each experiment in the revised manuscript. In terms of the fraction of subbasins showing improvement, the RUN experiment improved 58% of subbasins (29 of 50), the SFC experiment improved 64% (32 of 50), the INF experiment improved 64% (32 of 50), and the STC experiment improved 50% (25 of 50). This information has been incorporated into the revised manuscript text following Figure 3.

Q6. (Page 12, Figure 2) Could you please provide a figure showing the KGE of the four experiments and the default scheme for each HCDN? For example, a figure with the x-axis representing the HCDN, the y-axis representing the KGE, and five lines representing the four experiments and the default scheme. In this way, readers can more easily see the improvement or degradation of KGE for each HCDN.

Response: Thank you for this constructive comment. As suggested, we added a figure showing the KGE values of all five parameterization schemes (Default, RUN, SFC, INF, and STC) for each HCDN subbasin, with the x-axis representing the HCDN site number and the y-axis representing the KGE value computed against USGS monthly streamflow observations for the 2014-2023 period.

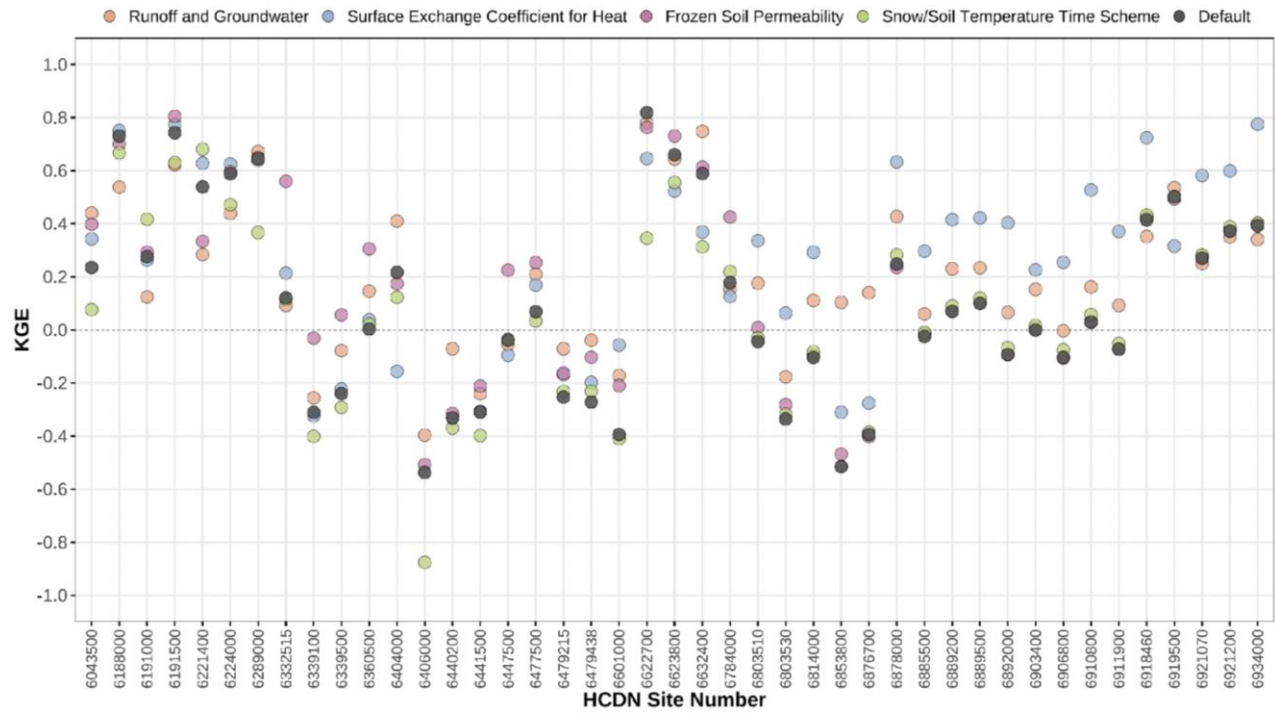


Figure S1. KGE values of the five Noah-MP parameterization schemes (Default, RUN, SFC, INF, and STC) for each of the 50 HCDN subbasins in the Missouri River Basin. The x-axis represents the HCDN site and the y-axis represents the KGE value computed against USGS monthly streamflow for 2014–2023.

This figure has been included as Figure S1 in the supplementary material to provide additional evidence of model performance across subbasins. We believe Figure S1 and Figure 5 complement each other effectively: Figure S1 allows readers to directly assess the absolute KGE performance of each scheme at the individual subbasins, while Figure 5 presents Δ KGE values based on environmental classification, enabling identification of parameterization schemes that consistently outperform the default one under specific hydroclimatic conditions. Together, these two figures provide a comprehensive insight into model performance across all subbasins.

Q7. (Page 12, Line 339) No improvement or deterioration? By how much does KGE decrease?

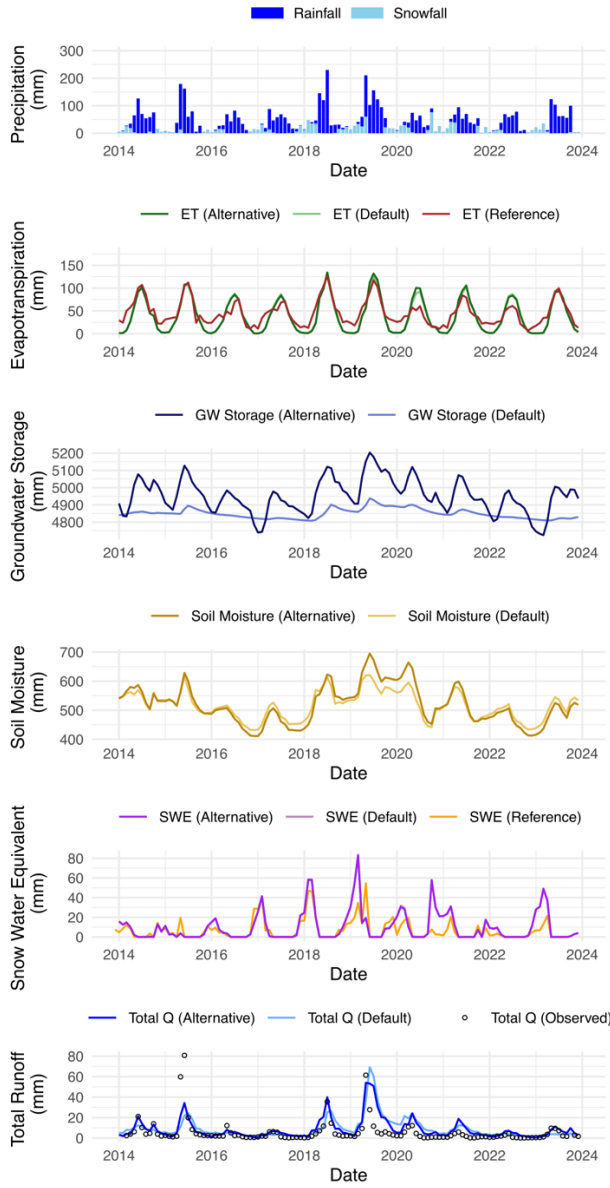
Response: Thank you for this comment. We clarify that "No Improvement" refers to subbasins where Δ KGE is zero or negative. The magnitude of KGE changes, including deterioration in individual subbasins, can be seen through two complementary figures. Figure S1 (added as a supplementary figure in response to Q6) shows the absolute KGE values of all five schemes for each HCDN subbasin, allowing direct comparison between the alternative schemes and the default.

Additionally, Figure 5 presents ΔKGE values organized by environmental classification, from which the degree of degradation can be quantified directly. Together, these two figures provide a comprehensive view of model performance changes across all subbasins.

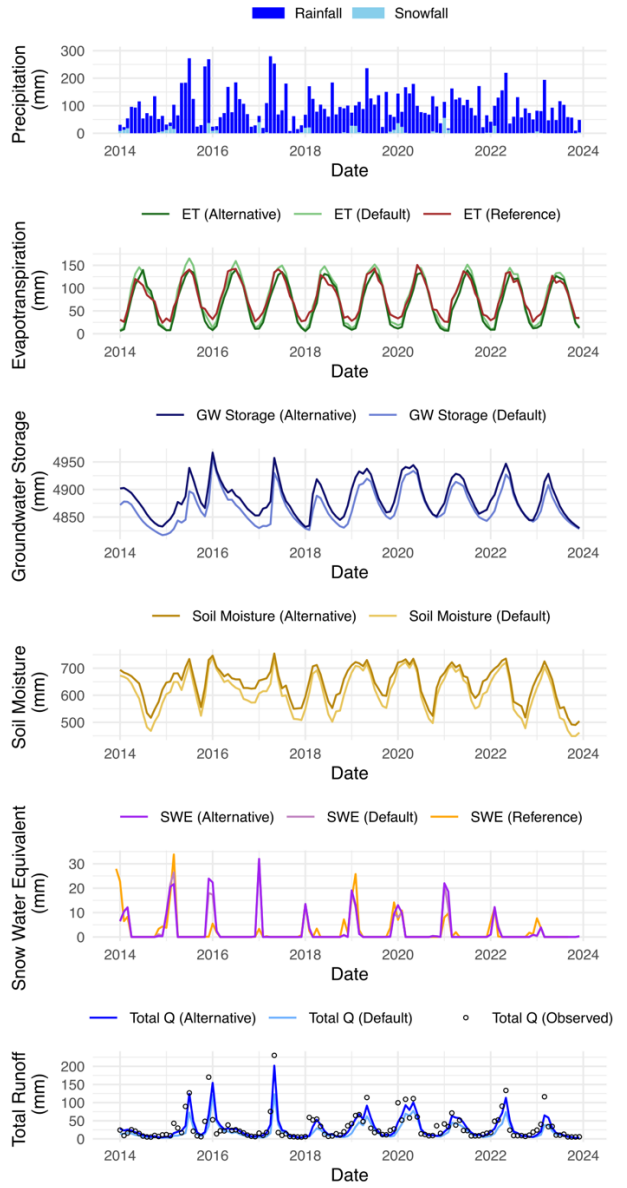
Q8. (Page 14, Figure 4) What does soil moisture look like in each of the four experiments compared to the default?

Response: Thank you for this suggestion. As suggested, we have added soil moisture panels to Figure 4, enabling direct comparison of soil moisture dynamics between the default and each alternative experiment across the four representative subbasins. The updated Figure 4 is provided below.

(a) Battle Creek Keystone Basin (RUN experiment)

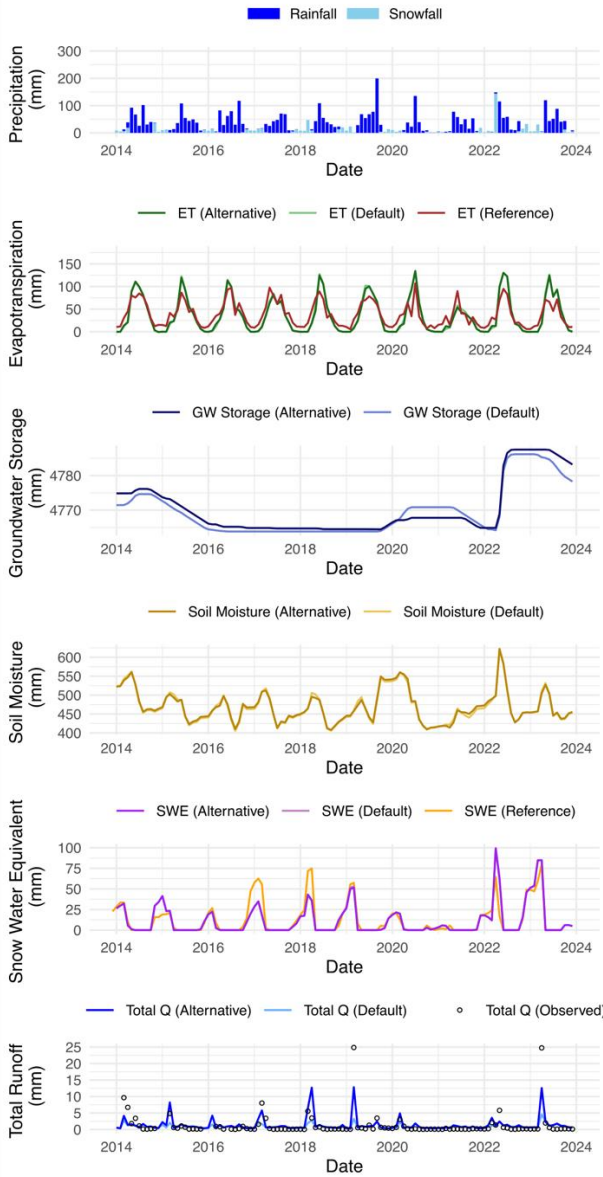


(b) Gasconade River Basin (SFC experiment)

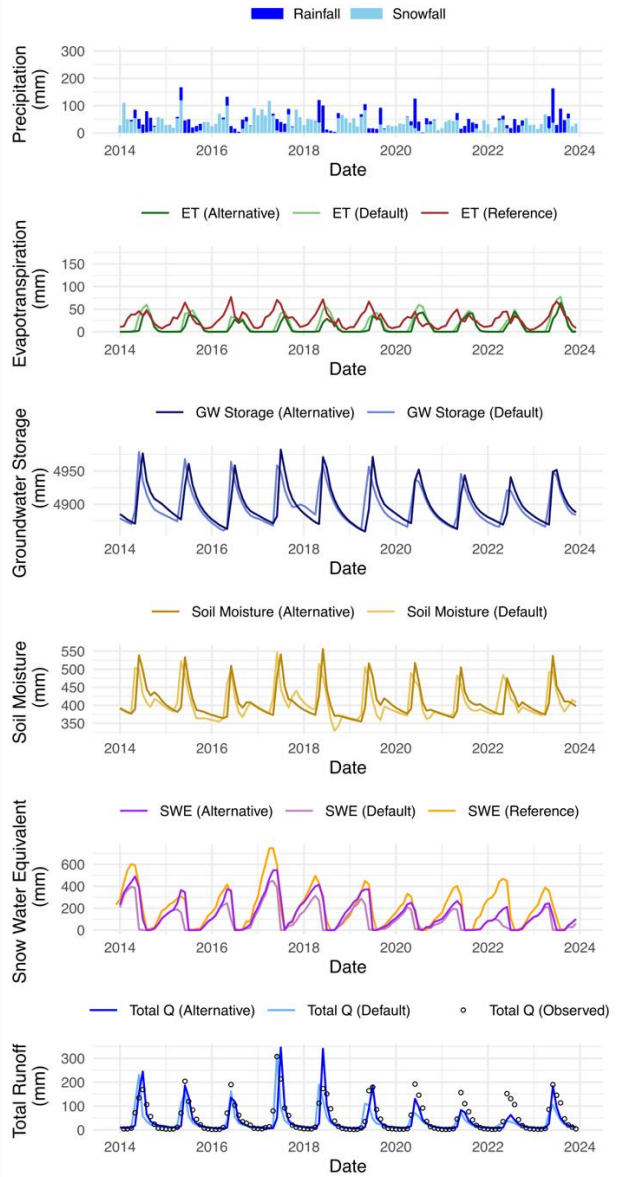


Revised Figure 4. Time series of precipitation, evapotranspiration, groundwater storage, soil moisture, SWE and total runoff of subbasin with notable improvement of KGE by the alternative scheme of (a) runoff and groundwater (RUN), (b) surface exchange coefficient for heat (SFC), (c) frozen soil permeability (INF) and (d) snow/soil temperature time scheme (STC). USGS streamflow is used as observation, with UA SWE and GLEAM evapotranspiration as reference datasets. Default results are indicated by solid lines in lighter colors, and alternative results by solid lines in darker colors.

(c) Bear Den Creek Basin (INF experiment)



(d) Dinwoody Creek Basin (STC experiment)



Revised Figure 4 (continued)

The soil moisture comparisons reveal notable differences among the experiments, particularly for the RUN, SFC, and STC schemes relative to the default. In the RUN experiment (Figure 4a), the altered groundwater dynamics under the RUN experiment amplified the variability in soil moisture, driving wetter conditions during wet periods and drier conditions during dry periods. In the SFC experiment (Figure 4b), the reduced evapotranspiration directly increases soil moisture retention, which contributes to the higher runoff generation observed in this experiment. In the

STC experiment (Figure 4d), delayed snowmelt results in a later but more sustained release of meltwater into the soil, producing distinct seasonal soil moisture patterns compared to the default. These differences in soil moisture states are particularly relevant for flood generation, as antecedent soil moisture is a critical initial condition governing the partitioning of precipitation and snowmelt into runoff.

The updated Figure 4 further reveals an important new insight regarding the relationship between groundwater storage and soil moisture dynamics. Although the default scheme fails to reproduce meaningful seasonal variability in groundwater storage, soil moisture continues to exhibit strong dynamics across all experiments. This suggests that the inability of the default groundwater scheme to capture seasonal groundwater fluctuations does not suppress soil moisture dynamics, as soil moisture is predominantly governed by shorter-term land surface processes such as precipitation and evapotranspiration. Rather, the key limitation of the default scheme appears to lie in its representation of long-memory hydrological behavior, specifically the slow and sustained release of water from deeper groundwater storage that influences baseflow and antecedent wetness conditions over longer timescales. When groundwater storage dynamics are inactive or poorly represented, this long-memory component is effectively absent while shorter-term soil moisture responses remain intact. This distinction has important implications for flood prediction, as accurate representation of both short-term soil moisture and long-term groundwater memory is essential for accurately characterizing antecedent conditions.

Revised manuscript (Pages 13 - 14, Lines 372 – 391) with key changes in bold italic

For the subbasins shown in Fig. 4, the time series comparisons demonstrate how the alternative schemes address limitations of the default setting. The corresponding KGE values indicate improvements based on runoff comparisons with observed streamflow. In Battle Creek Keystone (Fig. 4a), the default shows limited groundwater variability. The RUN experiment produces clear seasonal and interannual groundwater variability which in turn improved simulation of peak runoff in 2014, 2018, and 2023 and increases KGE by 0.19. The 2019 peak was still lower than observed but is more accurately represented than in the default scheme. ***The RUN experiment also shows that the improved groundwater storage dynamics influenced soil moisture variability, with more pronounced wet and dry cycles compared to the default.*** In the Gasconade River (Fig. 4b), the SFC experiment reduces evapotranspiration relative to the default and raises KGE by 0.38. It

corrects the systematic underestimation of flood peaks from 2014–2023, including the largest event in 2017. Comparison with GLEAM evapotranspiration confirms that the default scheme systematically overestimates ET in this subbasin, while the SFC experiment produces more realistic ET estimates. ***The reduced evapotranspiration in the SFC experiment led to increases in soil moisture, contributing to wet conditions prior to flood events.*** In Bear Den Creek (Fig. 4c), the default fails to capture the rapid increases in surface runoff, while the INF experiment resolves this limitation. It improves KGE by 0.44 with better simulation of the 2019 and 2023 peaks. In Dinwoody Creek (Fig. 4d), the STC experiment enhances SWE simulation and produces a KGE improvement of 0.14. The alternative SWE aligns more closely with UA SWE and results in more realistic runoff, particularly during the 2015, 2019, and 2023 floods. ***The delayed snowmelt under the STC experiment produces a later but more sustained infiltration signal, resulting in distinct seasonal soil moisture patterns relative to the default.*** Collectively, these results show how each experiment resolves distinct weaknesses of the default and improves runoff representation across diverse hydroclimatic conditions.

Q9. (Page 19, Line 455) Does the scheme reflect the idea of capillary rise?

Response: Thank you for this insightful question. The TOPMODEL with equilibrium water table scheme (Niu et al., 2005) does not explicitly simulate capillary rise as a dynamic upward flux. However, it implicitly captures the net effect of capillary-driven moisture redistribution through its equilibrium process, whereby the water table depth is determined iteratively from the total water deficit of the entire soil column. When the water table is shallow, this equilibrium condition naturally sustains elevated soil moisture in the upper layers, consistent with the physical influence of capillary forces above a water table. This structural difference distinguishes it from the default TOPMODEL with groundwater storage scheme (Niu et al., 2007), in which recharge is governed by the hydraulic conductivity and matric potential of the lowest soil layer. In the default scheme, overestimation of evapotranspiration depletes soil moisture in the lower layers, suppressing recharge and producing dampened groundwater storage variability with limited seasonal response. The equilibrium scheme bypasses this bottom-layer dependency, allowing more realistic seasonal groundwater variability to emerge. These distinctions have been clarified in the revised Methods section (Section 4.2.1) and the Discussion section (Section 6.1), as shown below.

Revised manuscript (Page 7, Lines 192 – 199) with key changes in bold italic

While both schemes link runoff generation to water table depth, they differ fundamentally in their representation of groundwater dynamics and soil moisture coupling. *In the TOPMODEL with groundwater storage scheme (Niu et al., 2007), groundwater recharge is calculated as a single vertical flux at the bottom of the unsaturated zone, governed by the hydraulic conductivity and matric potential of the lowest soil layer. As a result, groundwater storage dynamics are directly dependent on the moisture state of the deepest soil layer, making the scheme sensitive to conditions in the lower soil layers. In contrast, the TOPMODEL with the equilibrium water table scheme (Niu et al., 2005) determines the water table depth iteratively from the total water deficit of the entire soil profile, allowing the water table to reflect the integrated moisture conditions across all soil layers.*

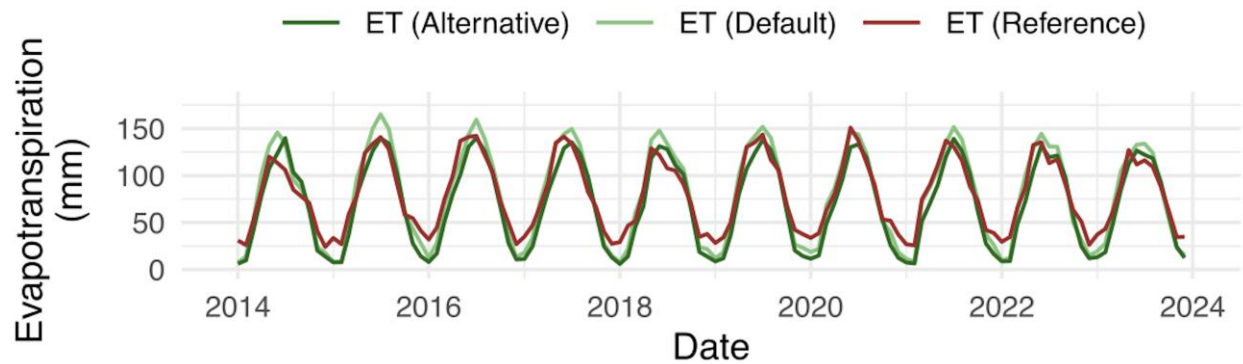
Revised manuscript (Page 20, Lines 475 – 483) with key changes in bold italic

The first limitation identified was the inadequate groundwater dynamics in the default groundwater scheme. Groundwater simulated by the default scheme does not always exhibit strong seasonality, with minimal variability throughout the year (Fig. 4a). *In the default scheme, groundwater recharge is governed by the hydraulic conductivity and matric potential of the lowest soil layer. This makes groundwater storage dynamics sensitive to the moisture state of the deepest soil layer only. Overestimation of evapotranspiration depletes moisture in the lower layers, suppressing recharge and dampening groundwater storage variability (Li et al., 2021; Cai et al., 2014).* Maina et al. (2025) addressed a related issue by integrating the ParFlow model with Noah-MP. *In this study, the alternative TOPMODEL with equilibrium water table scheme overcame this limitation by determining the water table depth from the total water deficit of the entire soil profile.*

Q10. (Page 19, Line 465) No observed ET was shown in Figure 4 or any other figures. You need to provide a figure to support the findings that ET was overestimated in the vegetated areas you mentioned.

Response: Thank you for this important comment. As recommended, we have added ET validation using GLEAM v4.2a (Koppa et al., 2022; Miralles et al., 2025), which provides observationally constrained ET estimates at 0.1° resolution for the period 1980–2023. The GLEAM ET time series

has been incorporated into the updated Figure 4, enabling direct comparison with model outputs across the representative HCDN subbasins.



Revised Figure 4b. Time series of evapotranspiration of subbasin with notable improvement of KGE by the alternative scheme of surface exchange coefficient for heat. GLEAM evapotranspiration is used as reference datasets. Default results are indicated by solid lines in lighter colors, and alternative results by solid lines in darker colors.

The ET comparison reveals that the default scheme systematically overestimates ET relative to GLEAM, particularly in the Gasconade River subbasin (Revised Figure 4b). In contrast, the SFC experiment, which employs the Chen97 surface exchange coefficient scheme to reduce latent heat flux, produces ET estimates that align more closely with GLEAM observations. This finding provides observational support for the claim that the default scheme overestimates ET in vegetated areas, and confirms that the ET reduction achieved by the SFC experiment is physically justified rather than a compensating error. The revised manuscript text in Section 5 has been updated accordingly to incorporate these findings, and the implications for runoff generation in cropland and grassland subbasins are further discussed in Section 6.

Revised manuscript (Page 14, Lines 379 – 384) with key changes in bold italic

In the Gasconade River (Fig. 4b), the SFC experiment reduces evapotranspiration relative to the default and raises KGE by 0.38. It corrects the systematic underestimation of peak runoff from 2014–2023, including the largest flood event in 2017. ***Comparison with GLEAM evapotranspiration confirms that the default scheme systematically overestimates ET in this subbasin, while the SFC experiment produces more realistic ET estimates.*** The reduced

evapotranspiration in the SFC experiment directly increases soil moisture content, contributing to higher antecedent wetness conditions for flood events.

Q11. (Page 19, 473-475) I am wondering how does the alternative scheme improve the representation of rapid runoff generation since it tends to reduce water movement in frozen soils? Could you add more sentences to explain that?

Response: The alternative scheme reduces water movement in frozen soils through a sharper reduction in soil water diffusivity as ice content increases. This decreased water movement effectively limits infiltration capacity, causing more precipitation and snowmelt to be partitioned into surface runoff rather than subsurface flow, thereby improving the representation of rapid runoff generation during frozen soil periods. We have included this explanation in the revised manuscript.

Revised manuscript (Pages 20 - 21, Lines 495 – 505) with key changes in bold italic

A third limitation is the potential underrepresentation of rapid surface runoff generation that can occur during ROS events. The default Noah-MP configuration inadequately captures the physical processes associated with ROS events, including ice lens formation, changes in snowpack thermal conductivity, and alterations in water retention capacity. The default frozen soil permeability scheme tends to predict more permeable frozen soils than typically observed in natural systems, particularly in conditions where ice layers inhibit infiltration (Koren et al., 1999; Niu and Yang, 2006). *The alternative scheme overcomes this limitation through a nonlinear weighting function based on volumetric ice content, which reduces soil water diffusivity quickly as ice content increases. 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. This is particularly important for improving runoff estimates during ROS events, which are often underestimated due to excessive infiltration by the default scheme.*

Q12. (Page 21, Line 500) To improve the quality of the paper, you might consider adding a section that investigates why improvements in SWE do not always translate into improvements in runoff. Selecting two representative sub-basins where SWE is improved but runoff responds differently

(improvement versus degradation), and analyzing changes in intermediate hydrological variables, could help develop a coherent narrative and substantially improve the quality of the paper by addressing the underlying scientific questions.

Response: We appreciate this constructive suggestion. Following the reviewer's recommendation, we have added an analysis investigating why SWE improvements do not always translate into runoff improvements under the STC experiment, and have incorporated two additional figures into the revised manuscript.

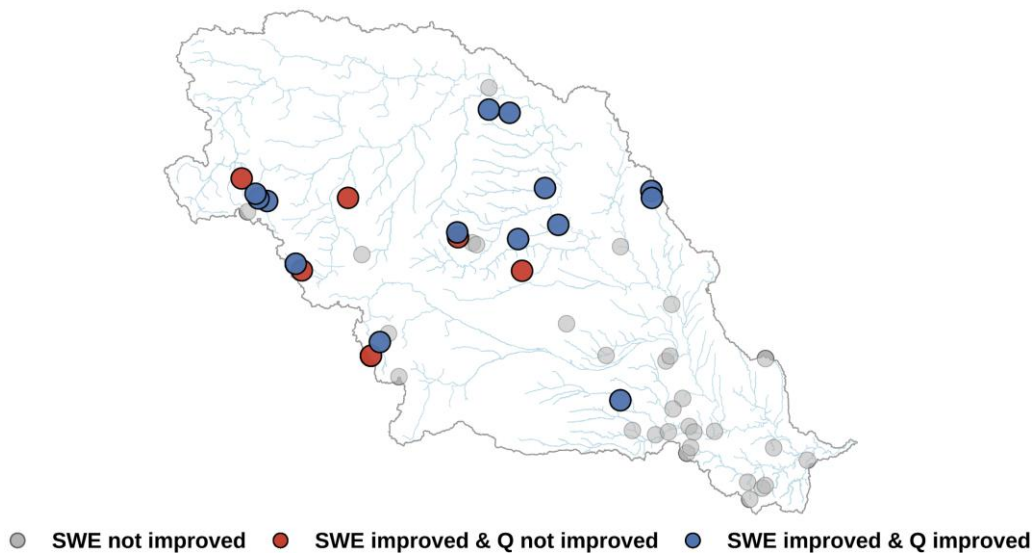


Figure 7. Spatial distribution of STC experiment outcomes across the 50 HCDN subbasins in the Missouri River Basin. Blue circles indicate subbasins where both peak SWE and Apr–Jul runoff were improved relative to the default scheme, red circles indicate subbasins where SWE was improved but runoff was not, and grey circles indicate subbasins where SWE was not improved. Major river flowlines are shown in light blue.

As shown in Figure 7, both improved and non-improved subbasins are spatially distributed across the Missouri River Basin without a clear regional pattern, suggesting that the contrasting runoff responses are driven primarily by subbasin-level hydrological conditions rather than by broad climatic or physiographic gradients. This spatial analysis demands a close examination of the direction of runoff bias under the default scheme.

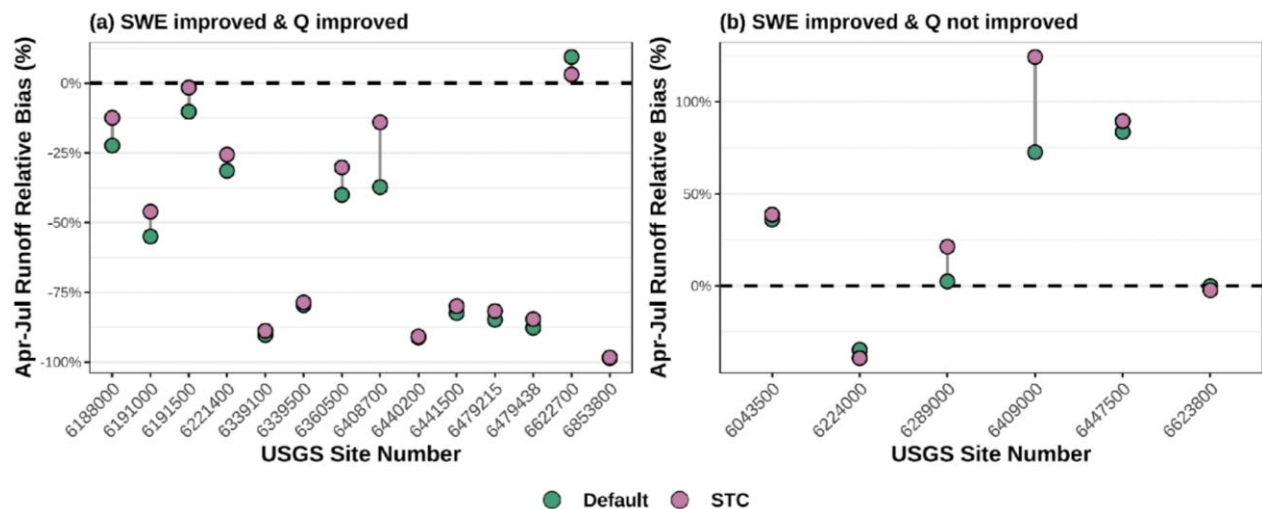


Figure 8. Apr–Jul runoff relative bias (%) for the Default (green) and STC (purple) schemes across SWE-improved subbasins, grouped by runoff improvement status: (a) subbasins where both SWE and Q were improved ($n = 14$) and (b) subbasins where SWE improved but Q was not ($n = 6$). Points are connected by gray lines to highlight the direction and magnitude of change between the two schemes. The dashed line indicates zero bias.

In Figure 8, the STC experiment generally corrected the underestimation of Apr–Jul runoff in subbasins where both SWE and Q were improved. In these subbasins, the default scheme predominantly underestimated runoff, and the STC scheme reduced this negative bias. In contrast, among the 6 subbasins where SWE improved but Q did not, the default scheme overestimated runoff in most cases. These results suggest that the effectiveness of the STC scheme in improving runoff simulation depends on whether the default configuration underestimates or overestimates runoff, underscoring the importance of diagnosing existing model biases before selecting parameterization schemes for operational applications. These discussions have been added to Section 6 of the revised manuscript.

In addition, soil moisture represents a key intermediate variable linking SWE changes to runoff responses. Soil moisture analysis has been incorporated into the representative subbasins in Figure 4 (see response to Q8), providing process-level insight into how altered snowmelt timing under the STC scheme modifies antecedent soil moisture conditions and subsequently influences runoff generation. A more comprehensive basin-scale analysis linking SWE, soil moisture, and runoff across all subbasins, particularly incorporating remotely-sensed soil moisture products as reference data, is identified as a direction for future research which has been added to the revised manuscript accordingly.

Q13. (Page 22, Lines 550-552) If you have the intermediate variables between SWE and runoff that were output from the experiment, could you do some analysis to answer why the snow process representation is improved but runoff is not? You might consider picking a representative subbasin and presenting the analysis.

Response: We thank the reviewer for this suggestion, which is closely related to the analysis added in response to Q12. As described in that response, we have incorporated soil moisture into the representative subbasin analysis in Figure 4, providing process-level insight into how altered snowmelt timing under the STC scheme modifies antecedent soil moisture conditions and subsequently influences runoff generation. The spatial distribution and bias-based analysis presented in Figures 7 and 8 further demonstrate that the effectiveness of the STC scheme in improving runoff simulation is closely associated with the direction of the pre-existing runoff bias in the default scheme, rather than on the degree of SWE improvement alone. Together, these analyses address the underlying question of why snow process improvements do not always translate into runoff improvements. A more comprehensive basin-scale analysis incorporating remotely-sensed soil moisture products as reference data is identified as a direction for future research and has been noted in the revised manuscript accordingly.