Reviewer Responses - egusphere-2025-6066-RC2

Dear Referee #2,

Thank you for taking the time to review our manuscript and for your positive and constructive feedback. We appreciate your comments and agree that incorporating your suggestions will further strengthen the study.

Please see below our responses to in blue.

Best,

David Casson, on behalf of all co-authors

RC2: 'Comment on egusphere-2025-6066', Anonymous Referee #2, 14 Mar 2026

Review of *Propagating Meteorological Uncertainty in Physically Based Mountain Snow Simulations* by Casson and co-authors for EGUSphere

Summary

The authors present a framework to quantify and propagate uncertainty in meteorological forcing through physically based snow modeling in mountainous terrain. It generates probabilistic meteorological forcing ensembles by combining station observations, topographic predictors, and atmospheric reanalysis with both regression and machine-learning methods to estimate spatial fields and their uncertainty. Precipitation observations are corrected for wind-induced gauge undercatch, and spatially correlated random fields are used to produce ensembles of meteorological inputs. These ensembles drive the energy-balance snow model SUMMA in three contrasting basins: the Chena (Alaska), Bow (Canadian Rockies), and Tuolumne (Sierra Nevada); basins that are generally well known in the North American snow hydrology community.

Results show that precip undercatch correction improves winter precipitation estimates, dynamic predictors such as reanalysis enhance regression skill, and Random Forest methods outperform regression in generating meteo fields. Ensemble verification shows positive probabilistic skill. When forced by the ensemble, snow simulations reproduce observed SWE in realistic ways: timing, magnitude, and uncertainty ranges seem reasonable. Overall, the study demonstrates a reproducible and transferable approach for explicitly representing meteorological forcing uncertainty in mountain snow simulations to support hydrologic modeling and water-resource decision making.

I commend the authors on a compelling demonstration of best practices in hydrological modeling, reproducibility, and uncertainty characterization. I believe this paper will serve as a guide for many early-career hydrologists seeking practical examples of rigorous modeling workflows and code development. The use of FAIR principles and Snakemake-based workflows is particularly notable and valuable for the community. The paper is clearly within the scope of the journal and will likely be well cited in the hydrological modeling literature, and potentially beyond in other environmental modeling disciplines.

Thank you for this positive assessment and the time taken to review our manuscript.

Main Comments

Interpretation of Figure 13 – I found these results particularly interesting. Could the authors expand on the implications of the results shown in Figure 13? Specifically, what do the cross-site differences in S-bias & S-pattern relationships reveal about the relative strengths and limitations of the different forcing datasets?

We thank the reviewer for highlighting Figure 13, which could benefit from further interpretation, including what cross site differences reveal about the site climatology and relative forcing datasets.

Across all study basin sites, the lower S-pattern values for the RF based ensemble indicate better temporal evolution of snowpack, including accumulation and melt processes. The ensemble forcing appears to better captures event-scale meteorological variability and temperature dynamics compared to deterministic reanalysis products, which tend to smooth these processes and may miss specific events captured in gauge records.

The cross-site comparison of S-bias is indicative of how forcing datasets contribute to SWE magnitude more specifically. Deterministic reanalysis products exhibit more consistent but often biased behavior across sites, with a pronounced negative S-bias in the Bow basin, likely related to precipitation undercatch and unresolved orographic enhancement in precipitation and temperature. The RF-based ensemble spans both positive and negative S-bias values, indicating that it captures a wider range of plausible accumulation magnitudes. This spread reflects uncertainty in precipitation inputs and phase partitioning, while also demonstrating that many ensemble members reduce magnitude bias relative to deterministic forcings.

The Taylor diagrams demonstrate similar findings, but displayed with different statistical metrics. Ensemble members generally achieve comparable or higher correlation with observations than deterministic forcings, though with a wider spread. Deterministic products occupy a narrower region of the Taylor space, reflecting more consistent but often biased representations of variability, whereas the ensemble provides a distribution of outcomes that includes both improved and degraded simulations. Across basins, these differences highlight how forcing uncertainty propagates differently depending on hydroclimatic regime. The Bow basin shows systematic magnitude bias linked to precipitation underestimation, whereas the Chena exhibits greater spread that may be associated with sensitivity to precipitation phase and accumulation variability. The Tuolumne shows tighter clustering, reflecting stronger topographic control and more consistent storm dynamics.

The results depicted in Figure 13 demonstrate the strength of an ensemble framework to provide a more complete representation of uncertainty and process variability, while also containing realizations that improve upon deterministic forcing, rather than uniformly outperforming them.

The following additions will be made to the manuscript, to improve the discussion and interpretation of the results.

Addition to Figure 13 paragraph (ln 561)

The RF-based ensemble generally exhibits lower S pattern values than deterministic forcings, which indicates better modelling of the temporal evolution of snow accumulation and melt. This suggests that ensemble forcing better captures the temperature driven processes and event variability that are often smoothed in reanalysis products.

In contrast, differences in S bias highlight systematic magnitude errors. Deterministic forcings show more consistent but often biased behavior, with pronounced negative SWE bias in the Bow basin, likely linked to precipitation underestimation and unresolved orographic effects. The ensemble spans both positive and negative bias, reflecting uncertainty in precipitation magnitude and phase partitioning.

Addition to Discussion (ln 601)

The snow simulation results show that the ensemble framework improves representation of snowpack temporal dynamics while explicitly capturing uncertainty in forcing. Rather than uniformly outperforming deterministic products, the ensemble provides a physically meaningful range of outcomes that reflects sensitivity to precipitation magnitude, phase, and variability. This reveals both improved and degraded simulations across sites, highlighting where deterministic forcings are biased and where snow processes are most sensitive to meteorological uncertainty.

Treatment of precipitation phase: There was surprisingly little discussion of how precipitation phase is treated in the model and how the calibration and ensemble framework may influence the impacts of precipitation phase errors on snow simulations. Could the authors expand on this aspect?

I would be interested to know whether there is a metric that could evaluate the relative uncertainty of the ensemble during precipitation events when air temperatures fall within the model's specified snow–rain transition range. In other words, is there a way to quantify how the different forcing-generation methods influence the model's treatment of precipitation phase and the resulting uncertainty in snow accumulation?

We agree that precipitation phase deserves clearer treatment in the manuscript, as it is so implicit to both precipitation and temperature forcing uncertainty translation to snow. In SUMMA, precipitation phase is determined using wet bulb temperature-based partitioning governed by calibrated snow parameters, most notably tempCritRain and tempRangeTimestep. These parameters control the central rain–snow threshold and the width of the transition range over which precipitation is partitioned between liquid and solid forms. As a result, uncertainty in air temperature near freezing directly propagates into uncertainty in precipitation phase, and therefore into uncertainty in SWE.

To better diagnose precipitation-phase uncertainty, we will quantify ensemble disagreement in rain–snow partitioning during precipitation events for which air temperature falls within SUMMA's calibrated transition range. For each ensemble member, we will compute the fraction of precipitation treated as snow (i.e., the ratio of solid precipitation to total precipitation) and use the ensemble spread in this quantity as a measure of phase uncertainty. We will summarize this using the standard deviation of the snow fraction across ensemble members, as well as a normalized uncertainty index based on the ensemble-mean snow fraction.

In addition, to assess the hydrological relevance of this uncertainty, we will quantify the ensemble spread in snowfall amount (i.e., solid precipitation input), which directly controls SWE accumulation. Together, these metrics will allow us to identify periods and regions where forcing-generation methods lead to the greatest uncertainty in precipitation phase, and where this uncertainty most strongly propagates into snow accumulation.

Representation of preferential snow accumulation areas: While the study focuses on GRUs and GRU-averaged snowpack metrics, it would be helpful if the authors could comment on the potential importance of resolving landscape units that exhibit preferential snow accumulation and harbor deep, persistent snowpack. These areas may provide critical surface and groundwater inputs and support important ecological functions. Such regions may also be the least likely to have in situ observations for either forcing or validation.

These landscape units may disproportionately contribute to basin water inputs and may either be more resilient to warming (snowpack refugia) or particularly susceptible to change. The manuscript clearly highlights the need to better resolve slope-scale wind effects from the perspective of gauge collection efficiency, but what about the potential to improve resolution of slope-scale orographic dynamics (e.g., the improvement of high-resolution reanalysis products that can capture precipitation banding from orography and atmospheric rivers, e.g.) in locations where precipitation gauges are absent?

We agree that preferential snow accumulation areas and sub-GRU heterogeneity are critical components of mountain hydrology that are not explicitly resolved in this study. Our use of GRU-averaged representation prioritizes consistency across large domains, but necessarily smooths processes such as wind redistribution, preferential deposition and areas resilient to warming. We will expand the discussion to acknowledge that these landscape units can disproportionately contribute to water storage and runoff, and may be underrepresented both in forcing estimation and validation due to sparse observations.

We also agree that improving representation of slope-scale orographic dynamics is an important direction for future work. While our framework partially accounts for this through dynamic predictors and spatially correlated uncertainty, it does not explicitly resolve fine-scale precipitation variability such as orographic banding or atmospheric river structure. Emerging high-resolution reanalysis and regional atmospheric models offer a promising pathway to better represent these processes, particularly in data-sparse regions. We have added discussion of these limitations and opportunities in the revised manuscript.

The following text with be added to the Discussion:

A key limitation of this framework is the representation of snow processes at the GRU scale, which aggregates sub-grid heterogeneity. Preferential snow accumulation areas, such as wind-sheltered slopes, leeward deposition zones, and forest openings, can sustain deeper and more persistent snowpacks that disproportionately contribute to basin water storage and ecological function. These features are not explicitly resolved and may contribute to residual bias and uncertainty in model evaluations. The challenge is compounded by observational limitations, as these locations are often under-sampled by meteorological and snow monitoring networks. While the ensemble framework partially accounts for unresolved variability through spatially correlated perturbations, it does not explicitly represent slope-scale processes such as wind redistribution or preferential deposition.

Improving representation of these processes will likely require both refined spatial discretization and higher-resolution meteorological inputs. Advances in high-resolution reanalysis and regional atmospheric modeling offer potential to better capture orographic precipitation gradients, banding, and atmospheric river dynamics, particularly in regions lacking dense gauge coverage. Integrating such datasets with distributed snow models represents an important direction for improving representation of snow accumulation in complex terrain.

Detailed Comments

Lines 253-254: Typo: "where data were co-located measurements"

- Agreed, to be updated

Line 410: Replace with "Meteorology”?

- Agreed, to be updated.

Lines 458-459: It may be helpful to mention that snow pillows also record conditions in the absence of overhead forest canopy, which may influence the representativeness of comparisons with spatially averaged model outputs.

- Agree, to be updated

Citation: https://doi.org/10.5194/egusphere-2025-6066-RC2