Reviewer 2

General assessment

The manuscript tackles a relevant and timely problem in hydrological modelling and contains a solid technical foundation. Nevertheless, major revisions are required before it can be considered for publication. In particular, clearer justification of methodological choices, improved structuring of the Results and Discussion, and a more cautious interpretation of findings—especially regarding hydrological realism and applicability to ungauged basins—are needed to support the conclusions drawn.

RESPONSE: Thank you for your comments and for acknowledging our topic as relevant, timely, technically sound. Below, we have provided our responses for your comments in blue text.

Major issues

1. Scale consistency and representativeness

A central issue throughout the manuscript concerns **consistency across spatial scales and data representativeness**. The study combines watershed-scale calibration, field-scale in-situ soil moisture observations, downscaled satellite products, and HRU-based model outputs, **but the implications of these scale mismatches are not sufficiently addressed**. In particular, the use of a very coarse soil dataset (FAO–UNESCO digital soil map) alongside high-resolution DEM and land-use data raises questions about **how soil spatial variability is represented and how soil parameters should be interpreted**. These issues condition the interpretation of sensitivity analysis, parameter uncertainty, and claims of improved hydrological representation and should be discussed more explicitly.

RESPONSE: Specifically, we now clarify that although the FAO–UNESCO soil map is coarse, TopoSWAT uses the high-resolution DEM to redistribute soil morphologic and hydrologic properties along terrain-derived similarity classes, improving the spatial realism of soil properties compared to relying on the raw soil map alone. This approach is well supported by previous findings showing that terrain-informed redistribution improves model representation of soil variability better than even SSURGO-scale datasets when local pedon data are limited (Buell, 2022; Fuka et al., 2016). We further explain that soil-moisture comparisons at the field scale rely on TI-based aggregation, which ensures that both in-situ observations and HRU-level model outputs reflect the same hydrologic similarity structure. These revisions make clear how we mitigate scale mismatches and how soil parameters and soil-moisture patterns should be interpreted within the SWAT-VSA framework:

"Although this study integrates datasets spanning multiple spatial scales—from a coarse FAO–UNESCO soil map to a 1-m DEM and field-scale soil-moisture measurements—several features of the SWAT-VSA framework allow these scales to be harmonized in hydrologically meaningful ways. First, soil property limitations associated with the coarse FAO–UNESCO map were mitigated through the use of TopoSWAT, which redistributes soil morphologic and hydrologic properties along topographic index (TI) gradients. This topography-informed redistribution has been shown to provide more realistic spatial patterns of soil storage and hydraulic conductivity than either raw coarse soil maps or even SSURGO-resolution datasets when local pedon data are unavailable (Buell, 2022; Fuka et al., 2016). As a result, although the original soil map is coarse, the soil parameters used in the model reflect a finer-scale, terrain-consistent structure.

Second, SWAT-VSA's modified HRU definition, which incorporates TI classes together with land use and soil type, provides a natural bridge between field-scale and watershed-scale processes in saturation-excess systems. Because soil-moisture patterns in such systems are strongly governed by topographic controls (Easton et al., 2008; Lyon et al., 2004), the TI classes represent hydrologically similar landscape units. In our study, all three TI classes present in the watershed are also represented within the 4.2-ha monitoring field, ensuring that field-scale in-situ measurements span the same hydrologic response types represented by the HRUs.

Finally, both field-scale and watershed-scale soil-moisture estimates were generated using TI-weighted aggregation of HRU-level outputs, matching the approach used to compute TI-weighted in-situ field averages (Asfaw et al., 2025). This alignment ensures that comparisons between modeled and observed soil moisture occur at consistent hydrologic similarity units rather than purely geometric scales. Together, these elements justify the representativeness of field observations and support the interpretation of model parameters and soil-moisture dynamics across the different spatial scales used in this study."

We have also added a section scale consistency:

"2.6 Scale Consistency and Representativeness

This study integrates datasets with differing spatial resolutions, including 1-m DEM, 30-m land cover, 500-m downscaled soil moisture, and coarse FAO–UNESCO soils. While these mismatches are unavoidable, several factors ensure that they do not compromise model interpretation.

First, the watershed is small (14.5 km²) and strongly governed by topographic controls typical of saturation-excess systems, meaning that relative wetness patterns are driven primarily by terrain rather than fine-scale soil heterogeneity. Second, SWAT-VSA's TI-based HRU structure redistributes soil and hydrologic properties along topographic gradients, providing hydrologically meaningful spatial organization even when source soil maps are coarse. Third, satellite soil moisture is used only in its watershed-averaged temporal form, not for resolving spatial variability, reducing sensitivity to its native resolution. Finally, field-scale evaluation uses TI-weighted aggregation of both in-situ and HRU-level modeled soil moisture, harmonizing scales between observations and simulation."

2. Sensitivity analysis and parameter interpretation

The sensitivity analysis underpinning parameter selection lacks sufficient clarity. It is unclear whether **sensitivity was assessed with respect to streamflow, soil moisture**, or both, despite the use of multiple calibration objectives. Because sensitivity rankings can differ substantially depending on the target variable, presenting a single ranking without clarification limits interpretability and affects later discussion of parameter uncertainty. In addition, some parameter **ranges (e.g., AWC) appear unrealistically wide** given their dependence on soil texture, raising concerns about physical plausibility and parameter compensation during calibration.

RESPONSE: We have edited this section in the methods, adding:

"The sensitivity analysis was performed using streamflow as the target variable because streamflow integrates the effects of runoff generation, soil moisture storage, evapotranspiration, and baseflow processes and therefore provides a comprehensive indicator of the watershed's hydrologic response. Although soil moisture was not used directly as a sensitivity target, all parameters with theoretical relevance to soil moisture dynamics, including AWC, ESCO, EPCO, Bulk Density, and

Ksat, were included to ensure that calibration on soil moisture or multi-objective metrics drew from a parameter set already demonstrated to influence hydrologic behavior.

Parameter ranges were intentionally broad for soil-related properties (e.g., AWC, Bulk Density, Ksat) to reflect known limitations of regional soil maps when applied to small, topographically complex watersheds. Previous studies demonstrated that allowing wider parameter bounds and redistributing soil properties along terrain-derived similarity classes, results in more physically realistic estimates of soil moisture storage and runoff generation in saturation-excess landscapes (Buell, 2022; Fuka et al) (please also refer our response to comment 1). Accordingly, wide ranges were necessary to enable SWAT-VSA to explore soil parameter values consistent with the local pedon-scale variability characteristic of the study site.

These ranges therefore do not imply physically unrealistic conditions; instead, they acknowledge uncertainty in the initial soil dataset while allowing the calibration algorithms to identify parameter combinations that are hydrologically plausible for a small watershed in which topographic convergence, shallow soils, and spatially variable wetness exert strong control on runoff processes."

3. Interpretation of results: statistical improvement vs. hydrological realism

While the Results clearly demonstrate improved statistical agreement of soil moisture estimates under soil moisture and multi-objective calibration, the manuscript does not sufficiently distinguish between **improved fit to processed calibration targets and improved hydrological process representation**. This distinction is critical given the reliance on downscaled and bias-corrected satellite soil moisture data. The **Results section** also mixes quantitative findings with interpretation, contributing to **redundancy with the Discussion** and obscuring the main messages. Either **additional diagnostics demonstrating improved hydrological realism or a more cautious framing** of the reported improvements is needed.

RESPONSE: Thank you for highlighting the need to more clearly distinguish between statistical improvements and genuine gains in hydrologic realism. In the revised manuscript, we reorganized the Results to remove interpretive statements and present only quantitative findings. We then expanded the Discussion to explicitly differentiate (i) improved agreement with the processed calibration target from (ii) improved representation of underlying hydrologic processes. This includes a critical assessment of water-balance partitioning, soil-moisture dynamics, and recession behavior, and acknowledges structural constraints of SWAT-VSA that limit the extent to which soil-moisture calibration alone can improve process realism. The revised Discussion now clearly identifies where the DSM calibration achieves statistical gains but sacrifices hydrologic consistency, and where the MO calibration yields more physically meaningful improvements: We have made an effort to incorporate suggested improvements to elucidate the performance gains in process-level improvements:

"Despite improvements in statistical performance, important aspects of the soil-moisture and streamflow dynamics remain imperfectly captured, and these patterns highlight structural limitations within SWAT-VSA as well as trade-offs among the different calibration strategies. Across all models, peak flows were underestimated, and the soil-moisture-calibrated (DSM) model showed a noticeably slower recession after storm events relative to the streamflow-calibrated (SF) and multi-objective-calibrated (MO) models. The MO and SF models captured the recession limb more realistically, whereas the DSM model tended to overestimate low-flow conditions, as also illustrated by the flow-duration curve (Figure X). These patterns indicate that while soil-moisture-only calibration can improve the timing of near-surface wetting, it does not necessarily improve simulation of lateral redistribution or drainage processes that govern low-flow and recession behavior.

Water-balance diagnostics reinforce this distinction. The DSM model produced 13–16% lower ET and higher total streamflow than the SF and MO models, primarily because the DSM calibration converged on significantly lower AWC values. Lower AWC increases the proportion of precipitation available for quick runoff and lateral flow rather than storage and evapotranspiration. This reflects a compensation pattern rather than improved representation of the true hydrologic processes: the model reduces AWC to match the temporal variability of the satellite-informed soil-moisture target, but at the cost of producing unrealistic moisture storage and ET dynamics. By contrast, the SF and MO calibrations selected higher AWC and more moderate ESCO/EPCO values, resulting in ET fractions that are more consistent with regional expectations for humid, temperate watersheds (approximately 60% of annual water balance; Sanford et al., 2012).

These results underscore a key point: not all apparent improvements in statistical fit correspond to improved hydrological realism. The DSM calibration improves the variance structure and temporal alignment of surface soil moisture, a statistical improvement, but the associated ET underestimation, unrealistic dryness biases, and excessive partitioning into quickflow indicate that the internal process representation is less physically plausible. In contrast, the MO model balances soil-moisture fit with realistic streamflow recession behavior and produces water-balance components closer to independent regional estimates, suggesting that its improvements reflect both statistical and process-level gains.

Some performance limitations reflect structural constraints of the SWAT model itself. SWAT's unsaturated zone representation is vertically lumped within layers and does not explicitly simulate lateral soil-moisture redistribution among HRUs; excess water leaves an HRU and is routed directly to stream reaches. This structural feature restricts the model's ability to represent cross-slope moisture exchange, a key mechanism in saturation-excess systems, and likely contributes to the DSM model's unrealistic ET–runoff partitioning. Similarly, SWAT's simplified treatment of evaporation (ESCO, EPCO) limits the model's ability to adjust vertical moisture redistribution without inducing compensatory parameter behavior. These constraints help explain why soil-moisture-only calibration does not automatically produce hydrologically consistent behavior, even when surface soil-moisture fit is improved.

Finally, the added complexity introduced by satellite-soil moisture downscaling, bias correction, and multi-objective calibration is justified only to the extent that it produces demonstrable process-level improvements rather than solely statistical gains. The present results suggest that downscaling + bias correction substantially improves the temporal quality of the soil-moisture calibration target, that DSM calibration alone yields statistical improvements but reduced physical realism, and that MO calibration provides the strongest balance between statistical performance and process fidelity.

Therefore, while the added complexity of downscaling and multi-objective calibration is warranted in this small, saturation-excess watershed, caution is needed in generalizing these findings to other settings or assuming that improved satellite-soil-moisture fit automatically implies improved hydrologic realism."

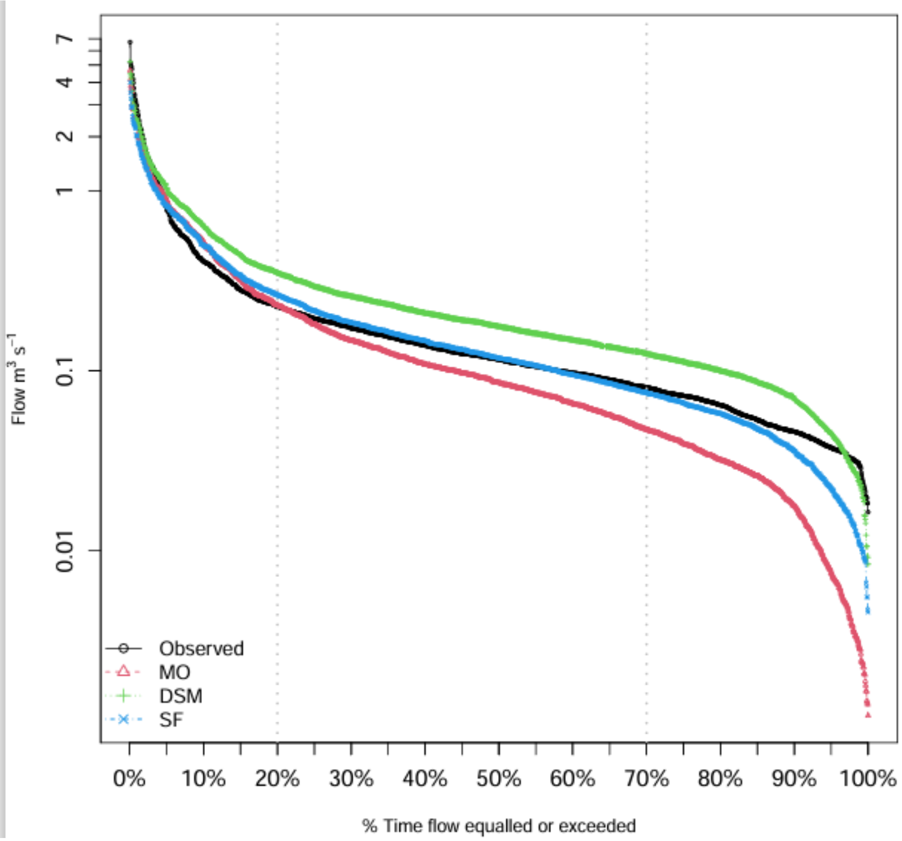

Figure : Flow-duration curves for observed streamflow and the three calibrated SWAT-VSA models: soil-moisture-only calibration (DSM), streamflow-only calibration (SF), and multi-objective calibration (MO). The figure illustrates how each calibration strategy affects the distribution of simulated flows, with notable differences in high-flow exceedance probabilities and low-flow recession characteristics.

4. Generality of conclusions and ungauged basin claims

Several conclusions—particularly those related to applicability in ungauged basins—extend beyond what is directly demonstrated. Although satellite soil moisture data are used in calibration, the approach relies on bias-corrected products, and the bias correction itself is informed by local in-situ measurements. As a result, no truly ungauged scenario is evaluated. These **limitations should be clearly articulated and reflected in the framing of the Conclusions**, with broader implications presented as potential rather than demonstrated.

RESPONSE: Thank you for this important comment. We agree that the conclusions in the earlier draft implied stronger applicability to ungauged basins than was directly demonstrated. We have revised the manuscript to clearly state that (i) our study did not test a fully ungauged scenario, (ii) in-situ soil-moisture measurements were not used to compute the bias-correction parameters and therefore do not introduce circularity, and (iii) any statements regarding ungauged applications should be interpreted as potential rather than established capability. We now emphasize that further testing is needed in settings where no local data are available to guide bias correction or evaluate model performance. We have revised the conclusions section as:

"Although this study demonstrates that downscaled and empirically bias-corrected satellite soil-moisture data can improve model performance in a small, saturation-excess watershed, these findings do not represent a fully ungauged scenario. The bias-correction procedure was developed using effective precipitation estimates and the uncertainty bounds produced by the downscaling algorithm; in-situ soil-moisture observations were not used to compute the bias-correction parameters. Nonetheless, in-situ measurements provided qualitative confirmation of the limited dynamic range of the downscaled product and helped motivate the need for bias correction. Because this auxiliary information would not be available in a truly ungauged basin, the broader applicability of the approach should be viewed as potential rather than demonstrated.

Thus, while the results indicate that satellite-soil-moisture–based calibration may expand modeling capability in data-sparse or minimally gauged settings, further evaluation is required before the approach can be confidently applied in fully ungauged watersheds where no local soil-moisture or streamflow data are available for validation or qualitative assessment. The framework presented here should therefore be considered a promising step toward enabling soil-moisture–driven calibration in ungauged basins."

Section-specific and technical comments

Introduction

The opening sentence is overly method-focused. Consider starting with the **hydrological relevance of soil moisture as a key state variable.**

The literature review is extensive but largely descriptive; **the specific conceptual gap addressed by this study should be articulated more clearly and earlier**.

RESPONSE: We have substantially revised the introduction, please see response to previous comments for details.

Methods

Move the Study Area subsection to the beginning of the Methods to better contextualize data choices, model structure, and calibration strategy.

RESPONSE: We have made these changes

Section 2.1.2: Clarify the soil moisture measurement method and sampling depths.

RESPONSE: We have made these changes

Section 2.5.1: Specify which observed variable(s) were used in the sensitivity analysis and describe the **sensitivity method and ranking metric** more clearly.

RESPONSE: We have made these changes, see response to reviwer 2 comment 2.

Line 225: The FAO–UNESCO digital soil map (1:5,000,000) is extremely coarse relative to watershed size and the study's focus on soil processes. **Clarify how many soil classes are represented, how soil parameters were derived, and why finer-resolution datasets were not used.**

RESPONSE: We have made these changes, see response to previous comments

Results

Results are organized by calibration approach, making comparison difficult. Reorganizing or explicitly synthesizing results around key outcome variables (e.g., streamflow, soil moisture, water balance, parameter behavior) would improve clarity.

RESPONSE: We have reorganized the results section following this approach

Interpretive statements appear throughout the Results, contributing to redundancy with the Discussion.

RESPONSE: We have removed interpretive statements from the results and moved them to the discussion

Discussion

Much of the Discussion reiterates Results or compares performance metrics with previous studies, but **offers limited deeper synthesis.**

Key limitations (scale mismatch, reliance on processed soil moisture products, single watershed and model structure) should **be explicitly framed as constraints** on interpretation.

RESPONSE: We have made these changes, see response to previous comments.

Conclusions

The Conclusions would benefit from a more cautious and conditioned framing. Improved statistical performance should not be equated directly with improved hydrological realism.

Claims regarding applicability to ungauged basins should be tempered, given that bias correction relies on in-situ data and no ungauged scenario is tested.

RESPONSE: We have made these changes, see response to previous comments.

Figures and tables

Table 1: Clarify the sensitivity target variable or present separate rankings for streamflow and soil moisture; justify the AWC parameter ranges in terms of physical plausibility.

RESPONSE: We have made these changes, see response to reviewer 2 comment 2.

Table 2 should be moved to the Results section or Appendix.

RESPONSE: We have made these changes.

Figure A1: (1) The workflow diagram does not fully or accurately reflect the methodology described in the text. (1) It is unclear why the "Pre-processing" box is connected to "SWAT Initialization," as the downscaled SMAP data are used for calibration rather than model setup. (2) **The link between in-situ soil moisture and satellite soil moisture data—via the bias correction step—is missing.** (3) The sensitivity and uncertainty analysis steps are not represented in the workflow, despite their importance in parameter selection and interpretation. (4) Objective functions are shown only for the model evaluation stage, even though they are central to model calibration; these should either be shown consistently for both stages or removed. After addressing these issues and improving the overall visualization, this figure would be more appropriate for inclusion in the main text rather than the Appendix.

RESPONSE: (1) The incorrect connection between “Pre-processing” and “SWAT Initialization” has been removed. (2) We clarified in the revised workflow that in-situ soil-moisture data are not used for bias correction; however, the updated diagram now explicitly shows where in-situ data are used for independent evaluation to avoid implying a missing linkage. (3) The workflow figure has been updated to include both the sensitivity analysis and parameter-uncertainty analysis steps, reflecting their roles in parameter screening and interpretation. 4) Objective functions are now shown consistently for both calibration and evaluation stages.

After incorporating these revisions and improving the overall clarity of the schematic, the workflow diagram has been moved from the Appendix to the main text as recommended.

Figure A2: The map layout is too low-resolution and visually unattractive. Both the legend and the two scale bars overlap with the maps. It is not clear where the study area is located on the right-hand side of the map. Adding a land-use map would be a useful improvement.

RESPONSE: We have edited this figure per the comment

Figure 3: Simplify by removing the legend and axis labels, retaining only values as pie labels.

RESPONSE: We have edited this figure per the comment

Figure 5: Specify the soil depth represented in the caption.

RESPONSE: We have added the soil depth

Line-specific comments

Lines 144–146: A reference is missing for the downscaling method; clarify why this approach was selected.

RESPONSE: We have added the missing reference.

Line 154: Specify the soil depth.

RESPONSE: We have added the soil depth

Line 161: Clarify what is meant by “field scale” (likely the 4.2 ha pasture).

RESPONSE: We have added a definitiontion of field scale. The reviwer is correct, in this context field scale refers to the 4.2 ha pasture.

Line 180: Clarify why river stage data are of particular interest and how they are used. I think only discharge data were used in this study.

RESPONSE: We have edited this section and removed reference to stage

Lines 257–259: Specify the values of parameters *d* and *k* in Eq. (1).

RESPONSE: We have edited this section, adding parameter values.

Line 380: Remove the reference to “Figure 1”.

RESPONSE: We removed this reference

Code availability

The provided GitHub link does not work and should be corrected.

RESPONSE: GitHub link updated. The previous repo was set to private, the link is correct and is updated to public.

Citation: https://doi.org/10.5194/egusphere-2025-5813-RC2