Preface from the Authors:

We thank the editors and reviewers for their careful and constructive evaluation of our manuscript. Consideration of the comments improved the clarity, rigor, and framing of the study. We have addressed the comments in full and revised the manuscript accordingly, including restructuring the Results and Discussion, clarifying methodological choices, strengthening the distinction between statistical performance and hydrologic realism, and explicitly outlining the scope and limitations of the approach. The revised manuscript now more clearly articulates the conditions under which downscaled satellite soil-moisture data can meaningfully inform hydrologic model calibration in small, terrain-controlled watersheds, while appropriately tempering broader generalizations.

Below, we have provided our detailed responses to the comments including text from the updated manuscript, where appropriate.

Reviewer 1 Comments

General

This manuscript examines the use of streamflow and downscaled and bias-corrected satellite soil moisture data for calibrating a SWAT-VSA model in a small, saturation-excess dominated watershed. The topic is relevant and timely; however, some issues need to be addressed before the manuscript can be considered for publication.

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

1. The Introduction does not sufficiently clarify the specific novelty of this study relative to existing literature. If the main contribution lies in calibrating hydrological models using downscaled and bias-corrected satellite soil moisture, the authors should more clearly review and position their work within existing studies on satellite soil moisture downscaling, bias correction, and reported performance improvements. Similarly, if the key contribution is the application of SWAT-VSA in saturation-excess dominated watersheds, the importance and added value of this choice should be more explicitly articulated. At present, the objectives and innovation of the study are not clearly distinguished from prior work.

RESPONSE: Thank you for this comment. We substantially revised the Introduction to explicitly articulate the novelty of this study relative to previous work. Specifically, we now clarify (i) how our approach differs from prior studies that used satellite soil moisture for calibration, (ii) why the small, saturation-excess watershed context is unique, and (iii) how the SWAT-VSA structure allows us to evaluate field-scale soil moisture performance, something not demonstrated in previous studies. The revised text also explicitly situates our downscaling and bias-correction approach within the current literature, highlighting the lack of studies using downscaled satellite soil moisture in watersheds < 100 km², and especially the absence of studies evaluating performance at field scale using TI-based HRU aggregation. We now clearly state the conceptual and methodological contributions and end the Introduction with explicit, concise research objectives:

"Reported outcomes, however, are mixed. Some studies show improved streamflow or drought metrics (Azimi et al., 2020; De Santis et al., 2021; Nanda et al., 2023; Patil & Ramsankaran, 2017; Sun, 2016; Wakigari & Leconte, 2023), while others report improved soil moisture simulation with little or no degradation streamflow performance (Rajib et al., 2016; Dangol et al., 2023; Duethmann et al., 2022; López López et al. 2017). These trade-offs reflect differences in watershed scale, dominant runoff mechanisms, satellite sensing depth, preprocessing approaches, and model structure.

" Uncertainty in soil moisture observation/estimation is particularly pronounced in small, saturation-excess–dominated watersheds, where runoff generation depends strongly on antecedent root zone soil moisture storage, topographic convergence, and hydrologic connectivity (Beven et al. 2021; Easton et al., 2008; Lyon et al., 2004). In such settings, coarse-resolution satellite soil moisture products, both native or downscaled, may capture large scale temporal dynamics but fail to represent event-scale wetness organization critical for runoff initiation (Gomis-Cebolla et al., 2022; Duethmann et al., 2022). Moreover, many previous studies are conducted in large basins (>700 km²), with comparatively few demonstrations in smaller basins (Duethmann et al., 2022; Dangol et al., 2023) and most lack independent evaluation of internal state estimates against field observations, limiting insight into whether calibration improvements reflect physical realism or compensatory parameter behavior. In addition, fewer studies report the effect of incorporating satellite soil moisture products for model calibration on both soil moisture and streamflow estimation performance (Dangol et al., 2023; Eini et al., 2023; Rajib et al., 2016). And while there is agreement that remotely sensed soil moisture can be used to calibrate a watershed model for improved soil moisture estimation, the scale of soil moisture outputs is still coarse (> $2km^2$), and additional effort is required to inform field-scale management applications."

Objectives

"The objectives were to test whether satellite soil-moisture-informed calibration improves temporal soil moisture variability without degrading streamflow performance, and whether multi-objective calibration can balance predictive accuracy with hydrologic realism. By combining downscaled satellite soil moisture, terrain-informed model structure, and independent field evaluation, this work advances understanding of when and how satellite soil moisture can meaningfully constrain watershed model parameters in small, terrain-controlled catchments."

2. The study is conducted in a single small watershed (~14 km²) dominated by saturation-excess runoff and relies on the SWAT-VSA model. While this choice is physically appropriate for the study site, it substantially limits the generality of the conclusions. The manuscript should more explicitly discuss the limits of applicability of the proposed approach, and the conclusions should be clearly framed as primarily applicable to small, saturation-excess runoff dominated catchments.

RESPONSE: Thank you for noting the need to more clearly delimit the scope of our conclusions. We have added a Limitations section to explicitly emphasize that (i) the study was conducted in a single, small (<15 km²) watershed, (ii) the watershed is strongly dominated by saturation-excess runoff, and (iii) SWAT-VSA is specifically designed for such hydrologic settings. We now state clearly that the proposed downscaling and bias-correction approach, as well as the soil-moisture–based calibration findings, should be interpreted primarily within the context of small, humid, saturation-excess systems. We also highlight challenges that may arise in larger or more heterogeneous basins, including increased spatial variability in soils, land cover, and precipitation, and the potentially weaker topographic control on soil moisture patterns. We have added:

"4.7 Limitations

This study was conducted in a small (14.5 km²), saturation-excess-dominated watershed. While appropriate for evaluating soil-moisture-driven calibration in terrain controlled hydrologic systems, this limits the generalizability of the findings. In larger or more heterogeneous watersheds, greater spatial variability in precipitation, land use, soils, and hydrologic connectivity may influence both the downscaled soil-moisture behavior and the calibration performance. In particular, the event based bias correction approach applied here relies on watershed average effective precipitation and may not capture spatially variable wetting responses in more complex basins.

The applicability of the SWAT-VSA model is restricted to saturation-excess runoff environments, where shallow soil storage deficits and topographic convergence largely control runoff generation. Watersheds dominated by infiltration-excess processes, weak terrain control, or arid and semi-arid conditions may not benefit from the TI-based structure. Nonetheless, saturation-excess processes are widespread in humid-temperate regions of the United States (Buchanan et al., 2018) as well as globally, suggesting that this approach remains relevant to a broad class of headwater catchments.

The downscaled soil moisture product underestimates peak wetness during high-moisture periods. Although the bias corrected improves event-scale responsiveness while preserving long-term temporal structure, it does not resolve underlying spatial or physical limitations of the satellite product. Consequently, the performance improvements should be interpreted as enhanced temporal constraint rather than complete correction of soil-moisture uncertainty.

Overall, the findings should be interpreted within these constraints. Application of this framework beyond small, saturation-excess watersheds will require careful consideration of watershed characteristics and evaluation across additional sites before broader inferences are made."

3. The study relies on a pretrained mlhrsm downscaling model and an empirical, precipitation-threshold-based bias correction scheme. However, the regional applicability of the pretrained downscaling model is not evaluated, and no comparison is shown between original, downscaled, and bias-corrected soil moisture. The authors should explicitly show these differences and discuss

how the downscaling and bias correction steps influence the soil moisture signal used for calibration and, consequently, the model results.

RESPONSE: Thank you for highlighting the need for comparing the soil moisture products. We have added the following in the results and discussions sections:

“ 3.1 Comparison of Downscaled and Bias-Corrected Soil Moisture Products

To evaluate how bias-correction alters the soil moisture signal used for calibration, two products were compared at the watershed scale: (1) the 500-m downscaled soil moisture series generated using the mlhrsm pretrained quantile random forest model, and (2) the bias-corrected downscaled product used in model calibration. The uncorrected downscaled soil moisture product exhibits a reduced dynamic range and underrepresents peak wetness relative to in-situ observations. Following bias correction, the soil moisture product exhibits increased peak responses immediately following precipitation events, and improved temporal agreement with in-situ measurements (Fig. 3). The bias correction primarily affects wet-period responses while the lower envelope of the soil moisture signal remains unchanged, confirming that the procedure enhances event responsiveness without artificially shifting long-term trends.”

Correspondingly we add the following secition (4.1) in the discussion “4.1 Effect of Downscaling and Bias Correction on the Soil-Moisture Signal

The downscaled satellite soil-moisture product captured the broad temporal patterns of wetting and drying at the watershed scale but exhibited a reduced dynamic range and attenuated peak wetness relative to in-situ observations. This behavior is consistent with previous evaluations of satellite and downscaled soil-moisture products, which often reproduce large-scale temporal variability but smooth event- and fine-scale extremes, particularly in topographically complex catchments (Brocca et al., 2017; Duethmann et al., 2022).

The event-based bias-correction approach applied here selectively enhanced soil-moisture responses during precipitation events while preserving dry-period behavior and long-term seasonal structure. By conditioning the adjustment on effective precipitation, the correction addressed the commonly reported attenuation of peak wetness in pretrained downscaling products without altering the lower envelope of the soil-moisture signal (Duethmann et al., 2022; Xu et al., 2025). Importantly, the correction parameters were not derived from in-situ soil-moisture measurements, ensuring that the adjusted product remained independent of the evaluation data and avoids circularity. events, .

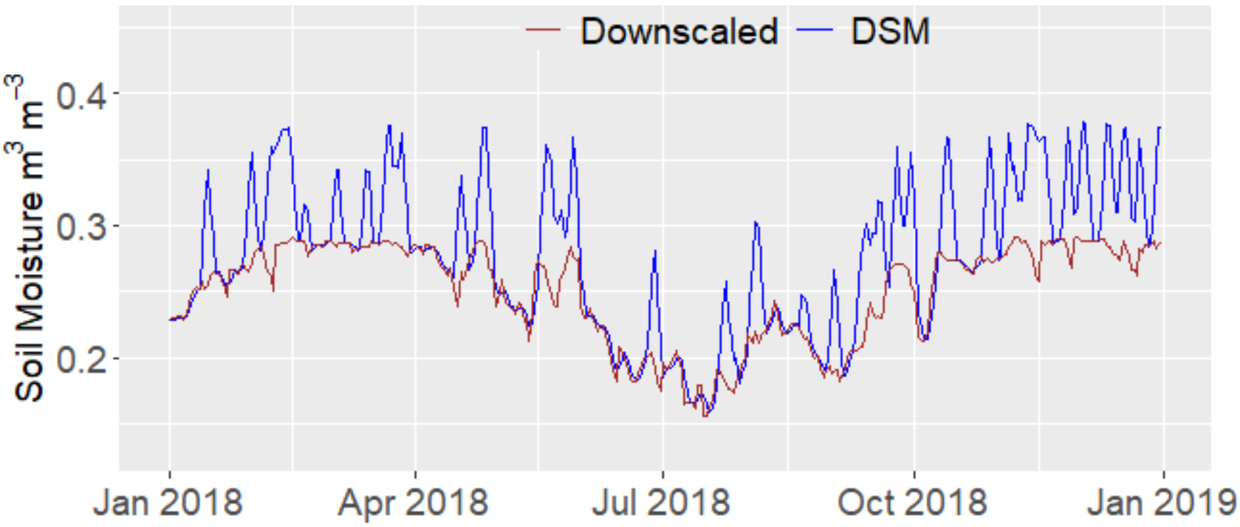

Figure 3: Time series of downscaled (mlrhsm product) and downscaled-bias-corrected soil moisture (DSM) at the watershed scale. Only data from one year are shown for brevity and clarity.

4. A key concern relates to the representativeness of the soil moisture observations. The 25 soil moisture sensors are located within a 4.2 ha pasture, whereas the modeled watershed covers approximately 14 km². It is unclear whether such a limited area is sufficient for evaluating model performance at the watershed scale.

RESPONSE: Thank you for raising this important point regarding the representativeness of the field-scale soil-moisture observations. We have expanded the manuscript to clearly explain why the 4.2-ha monitoring field provides a hydrologically meaningful basis for evaluating model estimates in a 14.5-km² saturation-excess watershed. Specifically, we now emphasize that (i) soil moisture patterns in saturation-excess systems are governed primarily by topographic controls, as demonstrated in previous work (Easton et al., 2008; Lyon et al., 2004), (ii) SWAT-VSA uses topographic index (TI)–based HRUs, ensuring consistent representation of topographic wetness classes across the watershed and the monitored field, and (iii) all three TI classes present in the watershed are also represented within the field, allowing field-scale observations to capture the full range of hydrologic response types. We also clarify that model outputs were spatially extracted to the exact field boundary and aggregated using TI-weighted HRU area fractions, ensuring that evaluation is carried out at comparable functional units. Lastly, the small watershed size supports the assumption of relatively uniform precipitation and evapotranspiration inputs. These clarifications are now included in the manuscript:

“Field-scale mean soil moisture was calculated using TI-class-weighted aggregation, consistent with the SWAT-VSA HRU structure. Point measurements were first grouped by TI class and then aggregated to a field-scale mean using the relative areal coverage of each TI class within the monitoring field. Although the monitored field represents a small fraction of the watershed, previous studies in similar saturation-excess system show that spatial soil-moisture variability is primarily governed by topographic controls rather than localized soil heterogeneity (Easton et al., 2008; Lyon et al., 2004; Asfaw et al., 2025). While land-use and soil variability contribute to watershed-scale moisture patterns, the small size of the watershed supports the assumption of near-uniform precipitation and evapotranspiration inputs, reducing the likelihood that climatic gradients confound field-to-watershed comparisons. Furthermore, previous work in the study area shows that spatial soil-moisture differences are dominated by topographic effects rather than local soil variation effects (Asfaw et al., 2025)."

5. An additional concern is that the evaluation of soil moisture performance throughout the manuscript is based primarily on soil moisture data that have undergone both downscaling and bias correction by the authors, without independent validation. While this preprocessing may be necessary, it raises the question of how the reliability of the calibrated model can be independently assessed.

RESPONSE: Thank you for this important comment. We agree that calibration against a processed soil-moisture product requires independent validation to ensure that model performance is not an artifact of the downscaling or bias-correction procedure. We now clarify in the manuscript that all model evaluation was carried out using independent observations that were not used in any stage of the downscaling or bias-correction process. Specifically:

RESPONSE: In-situ soil moisture measurements (0–120 mm, 25 points across a 4.2-ha field) were used solely for independent evaluation, not for calibration or for informing σ, k, or d in the bias-correction method.

Daily streamflow observations were likewise completely independent of the satellite soil-moisture data and serve as an additional validation for the soil-moisture-calibrated model.

The combination of these two independent datasets provides two orthogonal validation pathways, allowing us to determine whether statistical improvements during calibration reflect true hydrologic skill rather than agreement with the processed soil-moisture product.

These clarifications have been added to the manuscript:

Although calibration uses the downscaled and bias-corrected satellite soil-moisture series, all model evaluation is performed using independent observations not involved in the preprocessing or calibration steps. In particular, in-situ soil moisture measurements collected at 25 locations within the 4.2-ha field (0–120 mm depth) were used exclusively for model evaluation, providing an independent benchmark for field-scale soil-moisture dynamics."

These in-situ data were not used to compute the downscaling uncertainty (σ), the bias-correction factor (k), or the threshold precipitation depth (d), ensuring that no circularity is introduced into the calibration process.

In addition, independent streamflow observations were used to evaluate the hydrologic performance of the soil-moisture-calibrated models, providing a second, process-level assessment of model reliability that does not depend on the satellite product. Together, these two independent validation datasets allow us to determine whether improvements in soil-moisture calibration reflect genuine gains in hydrologic realism rather than improved agreement with a processed calibration target.

We have added to the manuscript:

“In-situ soil-moisture observations were not used to derive or calibrate the bias-correction parameters (σ, k, or d). Instead, σ is provided directly by the downscaling model, d is estimated from soil hydraulic properties and target depth, and k is set a priori for parsimony. In-situ measurements are used only for independent model evaluation and to qualitatively confirm that the corrected soil-moisture series exhibits a physically realistic range of wetness.”

The manuscript would benefit from a clearer explanation of how the authors ensure that the improved soil moisture performance reflects genuine model skill rather than agreement with a processed target dataset. In particular, the role of independent in-situ soil moisture observations in validating the calibration results should be more explicitly discussed, and the potential circularity introduced by calibrating and evaluating against processed soil moisture data should be acknowledged.

RESPONSE: Although it is a common practice to use split sample calibration and evaluation in hydrology, i.e. a portion of a dataset is used for model calibration and the other portion for model evaluation. As rightly noted, since the soil moisture data is not a direct observation for the spatial scale in this study and that it had undergone several forms and rounds of processing, we explicitly used independently collected in-situ soil moisture data for model evaluation.

The primary model performance evaluation comes from the use of independent measurements, i.e. independent streamflow and in-situ soil moisture measurements. We provide a clearer explanation in a revised Methods subsection (“Bias-Correction Rationale, Procedure, and Non-circularity”) detailing exactly which data informed the correction, the physical basis for the precipitation threshold d why scaling by the downscaler-reported σ is appropriate, and an explicit non-circularity statement clarifying that in-situ measurements were not used to compute σ, k, or d or to tune the trigger:

“2.4.3 Bias-Correction Rationale, Procedure, and Non-circularity

A parsimonious event-based bias correction was applied to the watershed-average downscaled soil moisture series to enhance short-term wetting responses while preserving the long-term temporal structure. The correction (Eq. 1), is triggered when daily effective precipitation exceeds a threshold of 5 mm, representing the depth, d, required to fill the upper soil storage (from field capacity to saturation) over the top 0–50 mm. When triggered, the downscaled soil-moisture estimate is increased by a multiple (k) of the model-reported uncertainty (σ), where σ is the standard deviation around the downscaled mean returned by the mlhrsm quantile random forest model.

This approach selectively expands peak soil-moisture responses during hydrologically events, addressing the observed attenuation of wetness extremes in the uncorrected downscaled product while leaving dry-period behavior and long-term means unchanged. The precipitation threshold reflects the shallow storage deficit that must be overcome before saturation-excess runoff responses emerge in topographically organized landscapes.

In-situ soil-moisture observations were not used to derive or calibrate the bias-correction parameters (σ, k, or d). Instead, σ is provided directly by the downscaling model, d is estimated from soil hydraulic properties and target depth, and k is set a priori for parsimony. In-situ measurements are used only for independent model evaluation and to qualitatively confirm that the corrected soil-moisture series exhibits a physically realistic range of wetness."

We have also added a section on soil moisture layer compatibility among products:

"To ensure consistent comparison across data sources, soil moisture depths were aligned between satellite estimates, model outputs, and in-situ measurements. The 0–50 mm SWAT soil layer was used for calibration against the satellite soil-moisture series, while a 0–120 mm composite layer was reserved for evaluation against field measurements (Table A1). All calibration analyses hereafter refer to the watershed-average, bias-corrected soil-moisture time series unless otherwise noted."

We added this table to the Appendix.

Source	Nominal / effective depth represented	How we aggregate to watershed/field	Model layer used for comparison	Role in this study
Downscaled satellite SM (mlhrsm from Enhanced SMAP)	Surface layer (0–50 mm)	Average all 500 m pixels intersecting the watershed to a daily watershed mean	0–50 mm SWAT layer	Calibration target in DSM and MO scenarios
SWAT-VSA model output	Discrete layers: 0–50 mm and 50–120 mm (plus standard deeper layers)	HRU outputs aggregated to watershed or field boundaries using TI-class/HRU area weights	0–50 mm (for satellite comparison); 0–50 mm + 50–120 mm composite = 0–120 mm (for field comparison)	Calibration (0–50 mm) and Evaluation (0–120 mm)
In-situ TDR measurements (25 points)	0–120 mm (field sampling depth)	TI-class-weighted average across the 4.2 ha pasture	0–120 mm model composite	Independent evaluation only

		to obtain a field-scale mean		

Note: The 0–50 mm equivalent designation for the satellite product is an operational alignment for comparability; it does not imply that the satellite directly senses 50 mm uniformly. The purpose is to avoid mismatched depths during calibration and to keep field-scale evaluation at the measured 0–120 mm depth.

Finally, we have added more detail to the limitation section:

"The downscaled soil moisture product underestimates peak wetness during high-moisture periods. Although the applied bias-correctedbias correction product improves event-scale responsiveness while preserving long-term temporal structure, it does not resolve underlying spatial or physical limitations of the satellite product. Consequently, the performance improvements should be interpreted as enhanced temporal constraint rather than complete correction of soil-moisture uncertainty.

6. In addition, the authors should clarify why downscaling to 500 m resolution is necessary if the soil moisture data are ultimately averaged to the watershed scale for calibration. It should be discussed whether this averaging undermines the advantages of using a spatially explicit, VSA-based distributed model.

RESPONSE: Thank you for raising this question. We have clarified in the manuscript why downscaling to 500 m remains valuable even though the soil-moisture series used for calibration is ultimately aggregated to the watershed scale. First, the native SMAP product has a 10-km footprint, close to an order of magnitude larger than our 14.5-km² watershed, and therefore contains additional spatial variability from outside the target watershed. Without downscaling, the watershed-average series would inherit the spatial aggregation over the 10-km product and might not be representative of the specific conditions in the watershed.

Second, although the calibration uses the watershed-mean soil-moisture time series, SWAT-VSA uses TI-based HRUs internally to simulate spatially organized saturation patterns. The downscaled product does not need to match spatial patterns at 500 m resolution to be hydrologically useful; what matters is that the temporal signal driving calibration has realistic variability and amplitude. Finally, the 500 m resolution reflects the options available in the mlhrsm tool, which does not currently allow user-specified target resolutions.

These points are now clearly articulated in the revised manuscript:

"Downscaling was required because the native 10-km Enhanced SMAP resolution is far larger than the 14.5-km² watershed. The 500 m downscaled product constrains the satellite signal to the watershed boundary and improves temporal representativeness at the scale of interest.

These spatio-temporal improvements are essential for constraining soil-moisture-related model parameters at the watershed scale. The choice of 500 m resolution was also constrained by the available pretrained models in the mlhrsm downscaling package, which does not currently provide alternative target resolutions.

Importantly, spatial averaging of the downscaled soil-moisture fields does not undermine the value of using the SWAT-VSA model. SWAT-VSA relies on TI-based HRUs to represent spatial patterns of saturation and runoff generation. While calibration uses a watershed-scale soil-moisture time series, the model internally maintains spatially explicit hydrologic processes using TI classes. HRU-weighted aggregation ensures that model-predicted soil moisture is scaled to the watershed or field boundary in a manner consistent with the model's internal hydrologic structure."

7. The Discussion section could be streamlined by reducing repetitive literature comparisons and would benefit from a more critical assessment of which aspects of soil moisture and streamflow dynamics remain poorly captured. The authors should discuss whether these limitations reflect structural constraints of the SWAT model (e.g., simplified representation of unsaturated flow processes), and whether the observed performance gains justify the additional complexity introduced by downscaling, bias correction, and multi-objective calibration. A clearer distinction between statistical improvement and process-level improvement would strengthen the interpretation of the results.

RESPONSE: We have made an effort to incorporate suggested improvements to elucidate the performance gains in process-level improvements:

"..... Nevertheless, the calibration strategy influenced bias patterns. The soil-moisture-only calibrated model consistently overestimated low flows, whereas the streamflow- and multi-objective-calibrated models exhibited underestimation bias (Fig. A1)...."

and

"4.3 Statistical Improvements verses Hydrologic Realism

Although soil-moisture-only calibration produced measurable improvements in statistical agreement with soil-moisture observations, these gains did not uniformly translate into improved behavior across all hydrologic processes. The soil-moisture-calibrated model better reproduced soil surface wetting dynamics but exhibited biased low-flow behavior and altered water-balance partitioning. This pattern illustrates a central concern raised in recent multi-objective calibration studies: improvements in fit to a particular variable may reflect parameter compensation rather than improved process representation (Gupta et al., 2009; Duethmann et al., 2022; Mei et al., 2023).

The multi-objective calibration provided a more balanced outcome by simultaneously constraining streamflow and soil moisture responses. Independent field-scale evaluation showed that this approach reduced both RMSE and bias across a wider range of soil-moisture conditions while preserving streamflow performance comparable to streamflow-only calibration. These findings align with previous work showing that multi-objective approaches can reduce equifinality and improve internal consistency, even if individual hydrologic performance metrics do not increase dramatically (Rajib et al., 2016; Silvestro et al., 2015).

Crucially, this study reinforces the distinction between improved statistical fit and improved hydrologic realism. Gains in soil-moisture metrics alone should not be interpreted as evidence of improved system representation unless supported by independent evaluation and water-balance diagnostics.”

“4.6 Implications for Small Saturation-Excess Watersheds

The results demonstrate that downscaled, empirically bias-corrected satellite soil-moisture data can improve internal hydrologic state estimation in small, saturation-excess-dominated watersheds when combined with a terrain-informed model structure. The SWAT-VSA framework, using topographic-index-based HRUs and TI-weighted field-scale evaluation, enables direct assessment of whether calibration improvements extend to management- relevant scales (Easton et al., 2008; Asfaw et al., 2025).

At the same time, the findings underscore the need for caution. Soil-moisture-only calibration can yield misleading improvements if evaluated solely on statistical metrics, particularly in models with simplified representations of unsaturated and lateral soil-moisture processes. Multi-objective calibration offers a more robust balance between surface wetness dynamics and integrated flow behavior, although its effectiveness remains dependent on watershed characteristics and model structure (Duethmann et al., 2022; Mei et al., 2023). 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.”

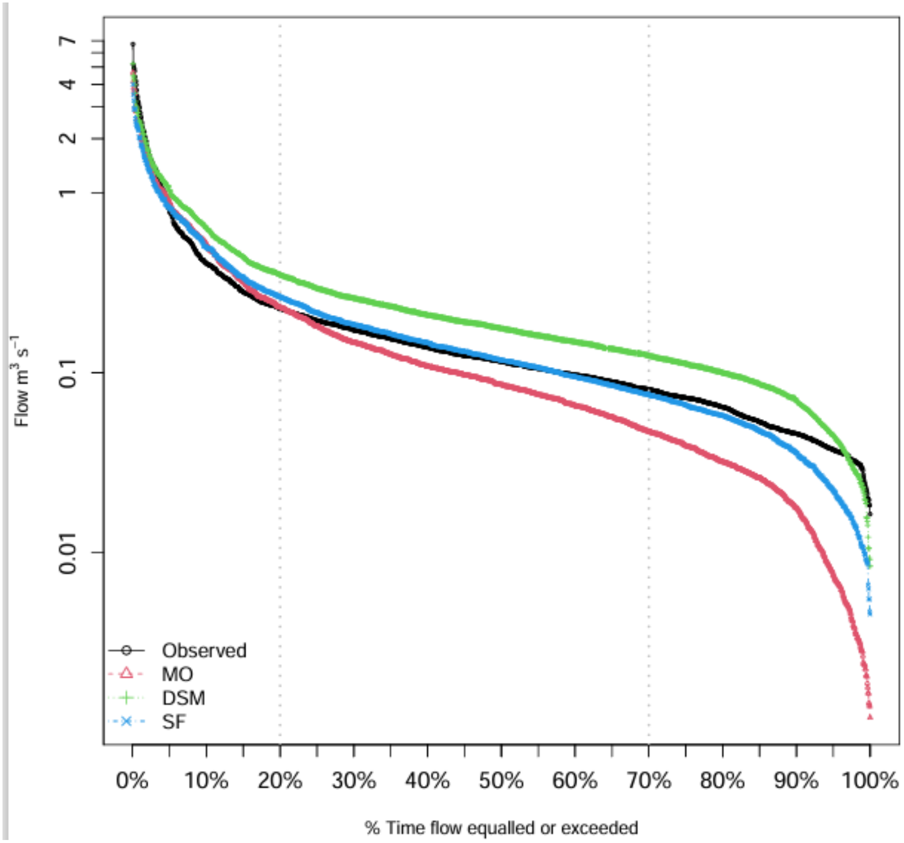

Figure A1: .Flow-duration curves comparing observed daily streamflow with simulations from the soil-moisture-only (DSM), streamflow-only (SF), and multi-objective (MO) calibration strategies. Differences among curves illustrate how calibration targets influence hydrologic realism: DSM captures surface wetting dynamics, but underestimates recession flows and high-flow peaks, whereas MO and SF better reproduce the observed distribution across both high-frequency (low flows) and low-frequency (peak flow) events.

Minor comments

1. Figures in the Appendix (e.g., Fig. A1 and Fig. A2) are essential for understanding the workflow and study area and should be moved to the main text.

RESPONSE: Figures have been moved.

2. The description in Section 2.2 and Fig. A2 may give the misleading impression that the study area is the full 57 km² Stroubles Creek watershed rather than the upper ~14.5 km² watershed.

RESPONSE: The description is updated as:

“Stroubles Creek is a 19.5 km long perennial stream in the New River Valley, VA, Fig. A2. In this study, the watershed outlet, 800 26’ 47’’ W and 370 12’ 10’’ N, is selected to define the study area. This outlet is about 1 km downstream of the flow monitoring station where the stream drains a 17.1 km^2 area.

The Virginia Tech StREAM Lab monitoring station, located at 80°26’ 42’’ W and 37°12’ 37’’ N, was the stream flow data source; at this monitoring station the creek drains the upper 14.5 km^2 watershed.”

3. The number of calibration iterations is reported in Section 2.5.4 (Parameter Uncertainty) but should be clearly stated in the Calibration Strategy section.

RESPONSE: We have edited as:

“The algorithm was implemented using the ‘DEoptim’ R package (Mullen et al., 2011). DEoptim was set up to evaluate 13 parameter vectors in each iteration for a total of 50 iterations per calibration. The number of iterations was decided sufficient after monitoring the successive diminishing returns in objective function minimization.”

4. The sensitivity analysis period (2015–2019) differs from the calibration and evaluation periods without sufficient justification. In addition, results in Fig. A3 are shown only for 2018, which requires explanation.

RESPONSE: We have restricted the results shown in Fig. A3 to 2018 for visual clarity. Similar results were obtained for the other simulated years, although the results were not shown for brevity.

The relative sensitivity values were calculated for model simulation from January 1, 2015, through December 31, 2019. This period adequately covers the range of streamflow variability and watershed responses, including during low-flow and high-flow periods (Figure 1), as well as the average soil moisture variability across the seasons.

5. Figure 1 and Figure 4 do not clearly convey differences among the three calibration strategies and could be improved for clarity.

RESPONSE: To improve clarity, the time span on the x-axis is now limited to one year. The full results are presented in the appendix for reference.

Citation: https://doi.org/10.5194/egusphere-2025-5813-RC1

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 to 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 (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:

"Soils data were collected from the FAO-UNESCO Digital Soil Map (FAO, 2007). TopoSWAT downscales soil texture, soil depth, available water capacity, and hydraulic conductivity across TI classes. This was demonstrated to improve spatial representation of soil properties in VSA dominated watersheds (Fuka et al., 2016; Collick et al., 2015).

......Importantly, spatial averaging of the downscaled soil-moisture fields does not undermine the value of using the SWAT-VSA model. SWAT-VSA relies on TI-based HRUs to represent spatial patterns of saturation and runoff generation. While calibration uses a watershed-scale soil-moisture time series, the model internally maintains spatially explicit hydrologic processes using TI classes. HRU-weighted aggregation ensures that model-predicted soil moisture is scaled to the watershed or field boundary in a manner consistent with the model's internal hydrologic structure. ......

..... Field-scale mean soil moisture was calculated using TI-class-weighted aggregation, consistent with the SWAT-VSA HRU structure. Point measurements were first grouped by TI class and then aggregated to a field-scale mean using the relative areal coverage of each TI class within the monitoring field. Although the monitored field represents a small fraction of the watershed, previous studies in similar saturation-excess system show that spatial soil-moisture variability is primarily governed by topographic controls rather than localized soil heterogeneity (Easton et al., 2008; Lyon et al., 2004; Asfaw et al., 2025). While land-use and soil variability contribute to watershed-scale moisture patterns, the small size of the watershed supports the assumption of near-uniform precipitation and evapotranspiration inputs, reducing the likelihood that climatic gradients confound field-to-watershed comparisons. Furthermore, previous work in the study area shows that spatial soil-moisture differences are dominated by topographic effects rather than local soil variation effects (Asfaw et al., 2025)."

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.

We add the following the manuscript

" ... A total of 20 parameters relevant to streamflow generation and soil-moisture dynamics were evaluated using relative sensitivity based on changes in Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) ..... "

And add to Table 1

" Sensitivity Ranking (SR), using streamflow as the target variable "

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:

“4.3 Statistical Improvements verses Hydrologic Realism

Although soil-moisture-only calibration produced measurable improvements in statistical agreement with soil-moisture observations, these gains did not uniformly translate into improved behavior across all hydrologic processes. The soil-moisture-calibrated model better reproduced soil surface wetting dynamics but exhibited biased low-flow behavior and altered water-balance partitioning. This pattern illustrates a central concern raised in recent multi-objective calibration studies: improvements in fit to a particular variable may reflect parameter compensation rather than improved process representation (Gupta et al., 2009; Duethmann et al., 2022; Mei et al., 2023).

The multi-objective calibration provided a more balanced outcome by simultaneously constraining streamflow and soil moisture responses. Independent field-scale evaluation showed that this approach reduced both RMSE and bias across a wider range of soil-moisture conditions while preserving streamflow performance comparable to streamflow-only calibration. These findings align with previous work showing that multi-objective approaches can reduce equifinality and improve internal consistency, even if individual hydrologic performance metrics do not increase dramatically (Rajib et al., 2016; Silvestro et al., 2015).

Crucially, this study reinforces the distinction between improved statistical fit and improved hydrologic realism. Gains in soil-moisture metrics alone should not be interpreted as evidence of improved system representation unless supported by independent evaluation and water-balance diagnostics.”

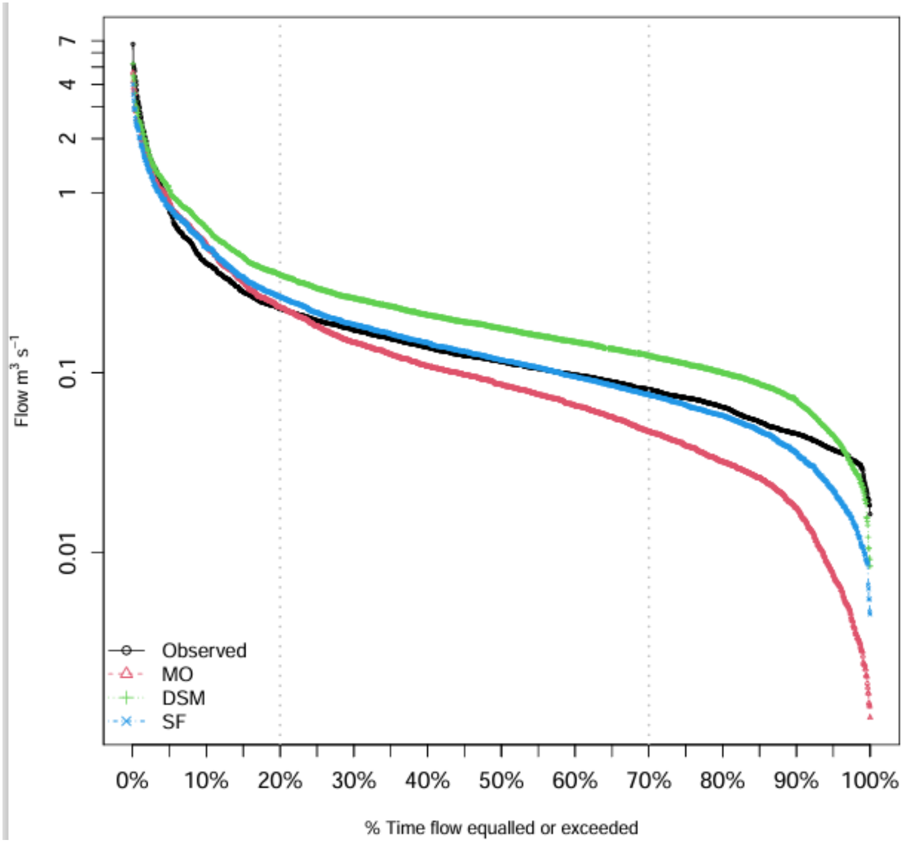

Figure A1: Flow-duration curves for observed streamflow and the three calibrated SWAT-VSA models: soil-moisture-only calibration (DSM), streamflow-only calibration (SF), and multi-objective calibration (MO). The figure illustrates how each calibration strategy affects the distribution of simulated flows, 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:

"In-situ soil-moisture observations were not used to derive or calibrate the bias-correction parameters (σ, k, or d). Instead, σ is provided directly by the downscaling model, d is estimated from soil hydraulic properties and target depth, and k is set a priori for parsimony. In-situ measurements are used only for independent model evaluation and to qualitatively confirm that the corrected soil-moisture series exhibits a physically realistic range of wetness"

And in the Conclusion section

"....While further evaluation is required in fully ungauged watersheds where no local soil-moisture or streamflow data are available for validation or qualitative assessment. "

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 reviewer 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 definition of field scale. The reviewer 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