Dear authors,

The revised version of your manuscript still suffers from some issues that should be addressed. Specifically, the core idea of the study, i.e. that satellite soil moisture data can improve hydrological modelling at watershed scale, remains not clearly supported by the presented results. This is probably due to imprecision in the presentation of the results, but it may also be related to calibration strategies not completely consistent with the objectives (and conclusions drawn), as pointed out by Referee #1.

Please, carefully consider all the comments received before resubmitting a further revision.

Best regards
Roberto Greco

Response: We appreciate the thoughtful comments provided by the Editor and Reviewers, which have significantly improved the manuscript. In this revision, we addressed all reviewer comments through substantial revisions to the methodology description, results interpretation, discussion, and limitations sections. We clarified the intended scope of the study and the role of the downscaled and bias-corrected satellite soil-moisture product, improved transparency regarding the use of in-situ observations during model selection, resolved inconsistencies in soil-depth definitions, refined interpretations of calibration performance, expanded discussion of hydrologic realism, parameter behavior, and equifinality, improved several figures and their accompanying discussion, corrected technical errors, and carefully reviewed the entire manuscript for consistency, language, formatting, and references. We believe these revisions strengthen the manuscript and address the concerns raised during review

Referee #1

I appreciate the substantial efforts the authors have made to revise and improve the manuscript in response to previous comments. The revised version provides a more detailed explanation of the study motivation, the role of downscaled and bias-corrected satellite soil moisture, and the applicability of SWAT-VSA in a small, saturation-excess-dominated watershed. However, I still have several concerns, and I therefore suggest that the following issues be addressed before the manuscript is considered for publication.

Comment: 1. The manuscript aims to evaluate whether downscaled and bias-corrected satellite soil moisture data can improve hydrological model performance. However, the current experimental design does not clearly isolate the effects of the original satellite product, downscaling, and bias correction. The calibration strategies compare streamflow-only calibration, soil-moisture-only calibration using the downscaled and bias-corrected soil moisture product, and multi-objective calibration. This design demonstrates that using the final DSM product can improve soil moisture estimation relative to streamflow-only calibration, but it does not show whether the improvement is due to the use of satellite soil moisture in general, the downscaling step, or the bias-correction procedure. To support the stated contribution more rigorously, the authors should consider adding calibration experiments using: (i) the original or Enhanced SMAP product, (ii) the downscaled but uncorrected soil moisture product, and (iii) the downscaled and bias-corrected DSM product. Such a comparison would allow the authors to explicitly quantify the added value of downscaling and bias correction for both soil moisture and streamflow performance.

Response: The contribution is the entire workflow, not the individual components. We have edited text to reflect this in the discussion:

“The experimental design was intended to evaluate the utility of the final downscaled and bias-corrected soil-moisture product for calibration and was not designed to isolate the relative contributions of the native satellite retrieval, downscaling procedure, and bias-correction method. Consequently, the reported performance improvements should be interpreted as the combined effect of the preprocessing and calibration framework.”

And in the Limitations:

“The relative contributions of the Enhanced SMAP product, machine-learning downscaling, and event-based bias correction were not evaluated independently. Future work should compare calibration performance across these intermediate products to better quantify their individual contributions.”

Comment: 2. The role of the in-situ soil moisture measurements in model evaluation and model selection needs clarification. The manuscript states that in-situ soil moisture observations were used exclusively for independent model evaluation and were not involved in preprocessing or calibration. However, the calibration strategy section indicates that final model selection among Pareto-optimal solutions additionally considered skill in reproducing field-scale in-situ soil moisture observations. If the in-situ data were used to select the final model, then they are not fully independent validation data. This potential circularity should be explicitly addressed. The authors should clarify whether in-situ soil moisture was used only after final model selection, or whether it influenced selection of the final multi-objective solution. If the latter is true, the manuscript should avoid describing the in-situ evaluation as fully independent.

Response: We have edited the sections describing the use of the in-situ soil moisture, clarifying that they were not used in the formal calibration, but were used as a comparator in selecting the final model:

"In-situ soil moisture observations were not used in the derivation of the bias-correction parameters. However, they were considered when selecting among Pareto-equivalent solutions and therefore should be viewed as an auxiliary model-selection criterion rather than a fully independent validation dataset.

Although calibration uses the downscaled and bias corrected satellite soil moisture data, all model evaluation is performed using independent observations not involved in preprocessing or calibration. 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, not for calibration, providing an independent benchmark for field scale soil moisture dynamics. "

Comment: 3. Figure 3 is used to support the statement that the bias-corrected DSM product shows improved temporal agreement with in-situ measurements. However, Figure 3 only shows the downscaled and bias-corrected/downscaled soil moisture time series and does not include the in-situ observations. Therefore, the figure can demonstrate that the bias correction increases peak responses after precipitation events, but it cannot directly demonstrate improved agreement with in-situ soil moisture. The authors should either add in-situ soil moisture observations to Figure 3, preferably for the same period, or provide quantitative statistics comparing the original downscaled product, the bias-corrected product, and in-situ observations. Without such evidence, the claim of improved agreement with in-situ observations is not fully supported by the figure.

Response: We have edited statements referring to this figure:

"Following bias correction, the soil moisture product exhibits increased peak responses immediately following precipitation events, and improved temporal agreement with in-situ measurements, and a greater dynamic range (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. "

Comment: 4. The description of the bias-correction threshold appears inconsistent. In section 2.4.2, the threshold precipitation depth is described as the equivalent value required to bring the top 120 mm of soil from field capacity to saturation. In section 2.4.3, the threshold is described as representing the depth required to fill the top 0–50 mm soil layer from field capacity to saturation. This inconsistency is important because the satellite-calibrated soil moisture layer is treated as 0–50 mm, whereas the in-situ evaluation uses a 0–120 mm composite. The authors should clarify which soil depth was used to derive the threshold and ensure that the bias-correction rationale is consistent with the soil moisture layer used for calibration.

Response: The reviewer is correct that the manuscript inadvertently referred to two different soil depths when describing the precipitation threshold used in the bias-correction procedure. The threshold was derived using the soil-water storage deficit of the upper 0–50 mm soil layer, which is the layer represented by the satellite soil-moisture product and used during calibration. The reference to 0–120 mm was incorrect and reflected the depth of the independent field-scale soil-moisture observations used for model evaluation rather than calibration. We have revised the manuscript to consistently describe the threshold as the precipitation depth required to bring the upper 0–50 mm soil layer from field capacity to saturation and clarified that the 0–120 mm composite layer is used only for comparison with in-situ measurements.

Comment: 5. The parameter sensitivity analysis is based on streamflow NSE, even though the study focuses on both streamflow and soil moisture calibration. This may bias the sensitivity ranking toward parameters that control streamflow rather than those that control soil moisture dynamics. Since soil moisture is a central calibration target in this manuscript, the authors should consider reporting sensitivity results using soil moisture metrics as well, or at least discuss how relying only on streamflow-based sensitivity may influence parameter selection and interpretation.

Response: We have added text to the methods to address this point:

“Because the sensitivity analysis was based on streamflow NSE, parameters that primarily influence soil-moisture dynamics may be underrepresented in the resulting rankings. However, all parameters with known relevance to soil-water storage and redistribution were retained for calibration.”

Comment: 6. The manuscript states that multi-objective calibration improves overall model performance for both streamflow and soil moisture. However, the streamflow results suggest that multi-objective calibration mainly preserves streamflow NSE and RMSE rather than clearly improving them relative to streamflow-only calibration. In addition, the multi-objective model shows a larger negative streamflow bias during the evaluation period. The authors should moderate this claim and more clearly distinguish between maintaining streamflow performance and improving streamflow performance. A more accurate interpretation may be that multi-objective calibration substantially improves soil moisture performance while maintaining streamflow skill at a level comparable to streamflow-only calibration.

Response: We have edited the relevant text to reflect this point. In the Abstract:

"The multi-objective calibration using streamflow and satellite soil moisture improved soil-moisture performance while maintaining streamflow performance comparable to streamflow-only calibration."

And Conclusions:

"Multi-objective calibration using both streamflow and soil moisture data outperformed single-objective approaches by improving estimation of both variables while reducing parameter uncertainty improved soil-moisture performance and maintained streamflow performance comparable to streamflow-only calibration."

Comment: 7. The workflow shown in Figure 2 is difficult to follow and should be redesigned. The authors should simplify the workflow by separating it into clearer blocks.

Response: We have edited the figure caption as:

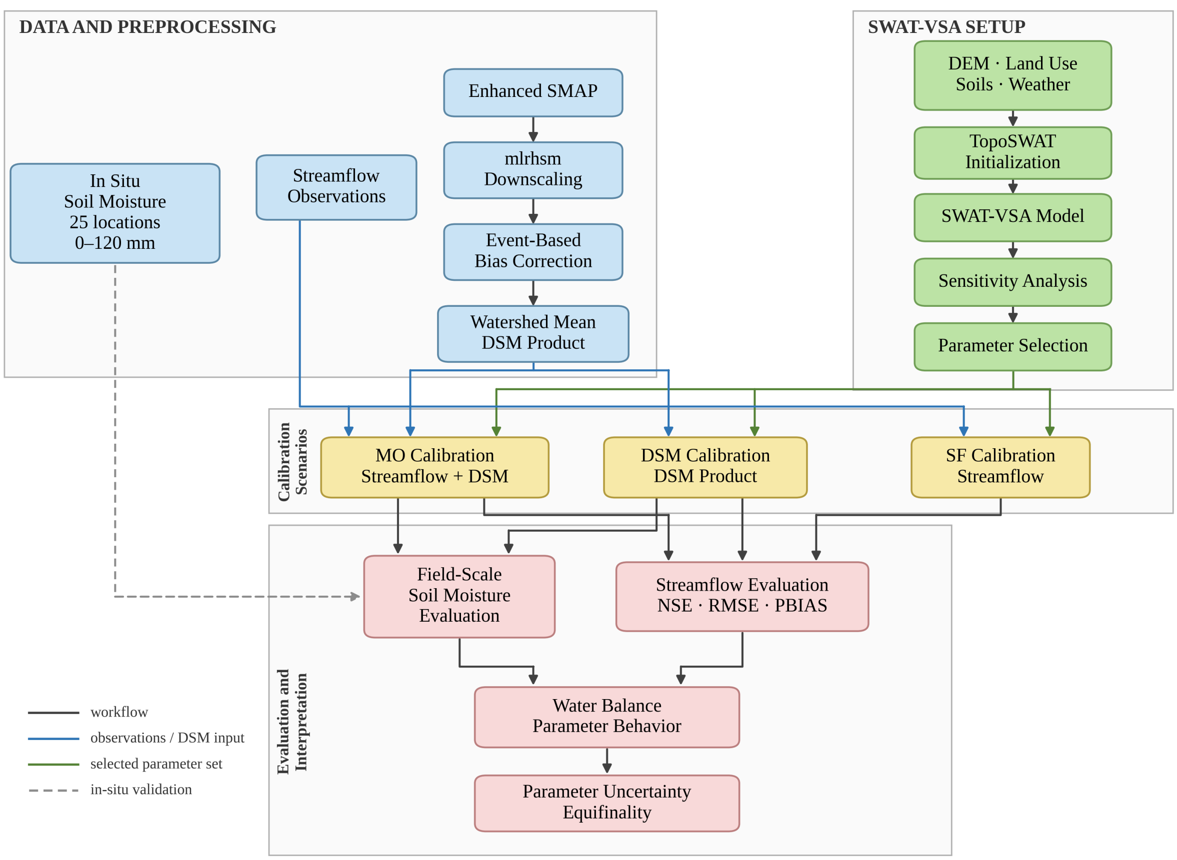

Figure 2: Study workflow showing generation of the downscaled and bias-corrected soil-moisture product (DSM), SWAT-VSA setup, calibration, evaluation, and uncertainty analysis. Three calibration approaches were compared: streamflow-only (SF), soil-moisture-only (DSM), and multi-objective (MO). Streamflow and field-scale in-situ soil-moisture observations were used to assess model performance and interpret parameter behavior and equifinality.

Comment: 8. The water-balance and parameter-behavior analysis is useful, but its purpose should be stated more explicitly. This section appears intended to evaluate whether improvements in statistical model performance correspond to physically realistic hydrologic behavior, or whether they result from compensatory parameter changes. The authors should make this purpose clearer at the beginning of the section.

Response: We have added a sentence to 3.4 to address this:

"This section examines whether improvements in model performance metrics correspond to physically realistic hydrologic behavior or arise primarily from compensatory parameter adjustments."

Comment: 9. There appears to be a unit error in Line 474. The manuscript reports an RMSE value for streamflow during the evaluation period using units of $m^3\ m^{-3}$, which are soil moisture units. This should be corrected.

Response: We have fixed this unit error as $m^3\ m^{-s}$

Referee #2: Mikolaj Piniewski, mikolaj_piniewski@sggw.edu.pl

The authors have addressed the major concerns raised in the initial review, and the manuscript has been substantially improved. I recommend acceptance subject to a few minor technical corrections:

Comment: 1. Figure 2 provides a useful overview of the workflow and is now appropriate for inclusion in the main text. However, the visual presentation could be further improved by reducing arrow crossings and simplifying the layout to enhance readability.

Response: We have edited the figure and caption as:

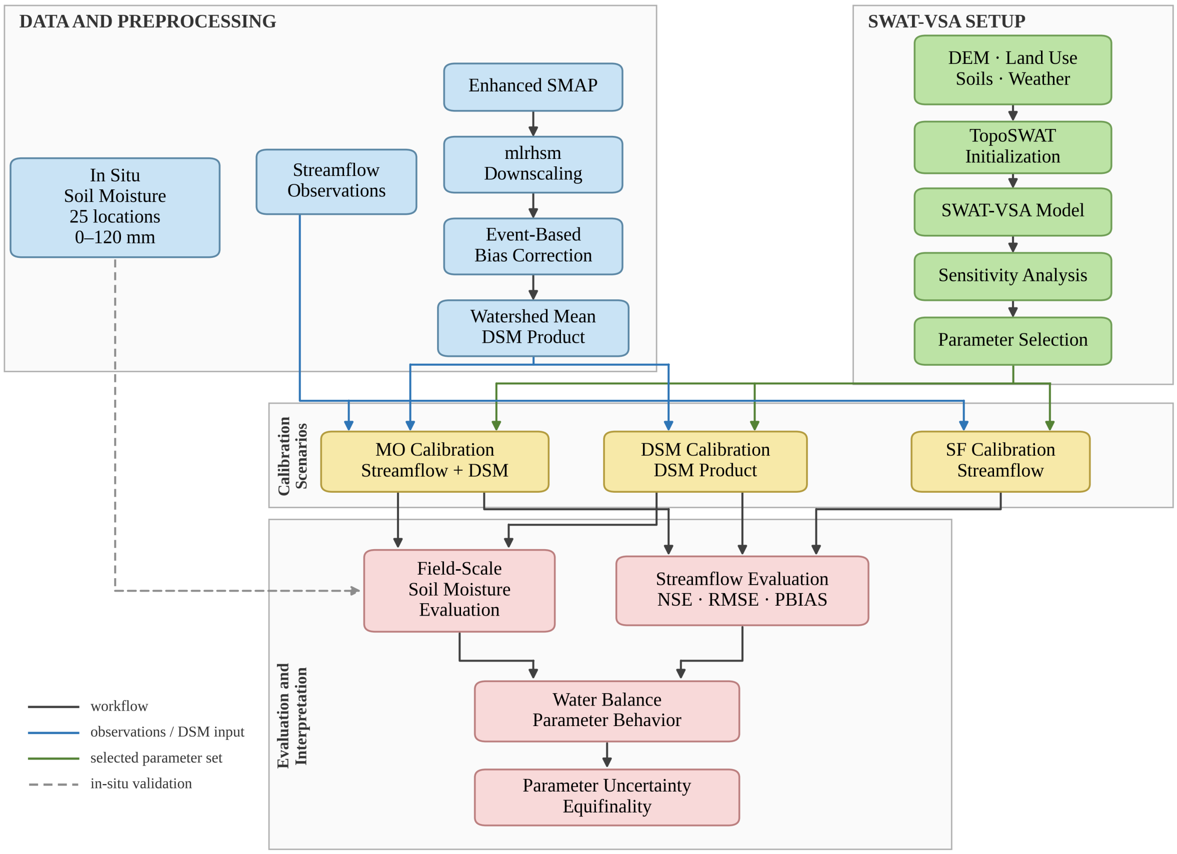

Figure 2: Study workflow showing generation of the downscaled and bias-corrected soil-moisture product (DSM), SWAT-VSA setup, calibration, evaluation, and uncertainty analysis. Three calibration approaches were compared: streamflow-only (SF), soil-moisture-only (DSM), and multi-objective (MO). Streamflow and field-scale in-situ soil-moisture observations were used to assess model performance and interpret parameter behavior and equifinality.

Comment: 2. Figure 10 is included in Section 4.5 but is not sufficiently integrated into the discussion. The authors should explicitly refer to the figure in the text and briefly interpret the uncertainty patterns shown, explaining how they support the conclusions regarding parameter uncertainty and equifinality.

Response: We have edited section 4.5 as:

“Consistent with previous studies (Rajib & Merwade, 2016; Silvestro et al., 2015), parameter uncertainty was lowest under multi-objective calibration (Fig. 10), particularly for soil-water- and evapotranspiration-related parameters. Figure 10 shows that the range of acceptable values for AWC, Ksat, ESCO, and EPCO was substantially narrower under multi-objective calibration than under either streamflow-only or soil-moisture-only calibration. This reduced parameter spread indicates fewer equally acceptable parameter combinations and therefore reduced equifinality (Beven, 1993). By simultaneously constraining streamflow and soil-moisture dynamics, the multi-objective calibration more effectively restricted the feasible parameter space and improved confidence in parameters governing soil-water storage and vertical moisture redistribution. Together with the water-balance analysis, these results suggest that multi-objective calibration not only improved model performance but also reduced compensatory parameter behavior."

Comment: These are minor presentation issues and do not affect the scientific conclusions of the study.

Response: We have fixed these presentation errors.