General response:

We sincerely thank the Editor and the referees for their careful review and insightful comments. We have provided detailed, point-by-point responses to the comments and suggestions from the Editor and referees. Each response addresses the specific feedback received, with detailed explanations and corresponding modifications to the manuscript where necessary.

Our responses are structured for clarity: Editor and referee comments in black, our responses in blue, and manuscript modifications highlighted in red.

In general, our revisions primarily include the following points:

(1) Improved readability and hydrological framing, including a clearer title/abstract, root-zone implications and limitations, abbreviation definitions, and standardized terminology.

(2) Expanded methodological details on radiometric calibration, footprint aggregation, validation aggregation, Bayesian sampling, baseline comparison, and uncertainty diagnostics.

(3) Updated figure/table captions and text to match the revised figures and tables, including scatter plots, sample sizes, confidence intervals, footprint significance tests, sensitivity diagnostics, and moderated conclusions.

Editor comments:

#1: The editorial board has received very constructive feedback from three reviewers. All of them agree that this is a timely study with the potential to be published in HESS, and I share their view. Please revise the manuscript based on their comments.

Response: Thank you for this helpful comment. We revised the manuscript throughout to address the editor and referee comments, including the hydrological framing, methodological details, updated captions/tables, uncertainty discussion, and moderated conclusions.

#2: HESS is a hydrology journal, and most of its readers are hydrologists who may not be familiar with many of the abbreviations used, particularly in the title. Please enhance readability for a broader audience.

Response: Thank you for this helpful comment. We revised the title and abstract for broader HESS readability, defined key abbreviations at first use, and added a glossary in the Appendix (Lines 556-562; Glossary of abbreviations).

#3: Regarding the implications of this study: surface soil moisture is likely not very critical for plant growth; root zone soil moisture is more important. On page 3, lines 64-65, the authors have listed several methods to infer root zone wetness patterns. However, additional approaches exist to derive root zone moisture conditions, such as remote sensing using the NDII (https://hess.copernicus.org/articles/20/3361/2016/) and a water balance approach driven by observed evaporation (https://hess.copernicus.org/articles/20/1459/2016/; https://hess.copernicus.org/articles/28/4477/2024/). Furthermore, I strongly encourage the authors to discuss the potential implications and/or limitations of your proposed new method for observing root zone moisture, as this would be more useful for agriculture and other water resources management practices.

Response: Thank you for this helpful comment. We agree that root-zone soil moisture is more directly relevant to plant water availability and agricultural water management than surface or near-surface soil moisture. In the revised manuscript, we therefore clarified in the Limitations and Future Directions section that the proposed UAV L-band retrieval targets near-surface, field-scale soil moisture rather than root-zone soil moisture (Lines 494-505).

Referee #1 comments:

#1: General Comments:

The manuscript presents a timely and physically meaningful framework that integrates dual-polarized τ-ω retrieval with RGB-TIR-derived priors and footprint-scale texture information for UAV-based L band soil moisture estimation. The study addresses several important challenges in passive microwave retrieval, particularly those associated with vegetation attenuation, support mismatch, and sub-footprint heterogeneity. One notable strength of the manuscript is that the discussion goes beyond reporting retrieval performance metrics and attempts to provide physical interpretation of retrieval ambiguity and uncertainty behavior.

The primary novelty of the study appears to arise from the integration of RGB-TIR priors, texture descriptors, and Bayesian inversion within a heterogeneity-aware retrieval framework, rather than from the introduction of fundamentally new microwave retrieval physics. This distinction could be stated more explicitly to avoid overstating originality. Nevertheless, the manuscript is generally well motivated, with clear context, rationale, and research objectives that are mostly supported through the subsequent analyses and discussion.

The manuscript contains several promising ideas regarding uncertainty-aware microwave retrieval. However, some interpretations appear somewhat stronger than what is directly supported by the presented analyses. In addition, although acknowledged in the limitations section, the dataset remains relatively limited in both temporal coverage and sample size and it also does not perform formal uncertainty propagation or variance decomposition to explicitly quantify the relative contribution of different uncertainty sources (e.g., radiometer noise, support mismatch, vegetation priors, or structural model error) to the total retrieval uncertainty.

Below are some specific comments that Authors should consider carefully to address and improve the manuscript and making it suitable for publication:

Response: Thank you for this helpful comment. We clarified that the novelty is the integration of RGB-TIR priors, footprint-scale texture descriptors, and Bayesian inversion, rather than a new microwave emission model, and we moderated mechanistic claims in the Discussion and Conclusion.

#2 Specific Comments:

Introduction

The introduction is generally well structured and clearly motivates the importance of vegetation attenuation, support mismatch and sub-footprint heterogeneity in passive microwave retrieval.

However, this section would benefit from a broader discussion of existing retrieval approaches including their pros & cons and clearer distinction among empirical machine learning retrievals, physically constrained Bayesian retrieval frameworks, and operational satellite retrieval methods, particularly in the paragraph beginning around line 84. Paragraph started referring context of recent studies. However, only one study is cited. This paragraph should be supported by more relevant literature to reflect strong rationale of this study objectives.

Response: Thank you for this helpful comment. We expanded the Introduction to distinguish empirical or machine-learning retrievals, operational tau-omega/L-MEB-style retrievals, and UAV or portable L-band retrievals, with additional literature support (Lines 85-93).

#3 Methods

The overall Bayesian τ-ω retrieval framework is physically interpretable and scientifically reasonable.

The validation dataset is relatively small, with 52 calibration points and 57 validation points over a restricted two-day validation period, although the authors acknowledge this limitation.

Response: Thank you for this helpful comment. In response to this concern, we expanded the independent validation dataset from the original 57 validation points to 219 matched validation points covering approximately 12.6 ha. The expanded validation now includes broader field conditions and cross-seasonal observations, which strengthens the evaluation beyond the initial short-period test. Nevertheless, we agree that the validation is still geographically limited to the Pengzhou study area and therefore does not fully demonstrate regional transferability. We have revised the Methods and Results to describe the expanded validation dataset, and we further clarified in the Limitations and Future Directions section that validation across additional regions, soil backgrounds, crop systems, and management conditions is still required before broader application can be claimed (Lines 174-181, 484-493).

#4 The prior mean and uncertainty for vegetation optical depth τ are estimated from calibration data, and τ calibration targets are obtained by inverting the τ-ω model using in situ soil moisture. This creates a two-stage inversion dependency. Please explain how errors in in situ soil moisture, temperature, roughness, and brightness temperature propagate into these τ targets.

Response: Thank you for this helpful comment. We clarified that the τ calibration targets are inferred rather than directly observed, and that errors in TDR soil moisture, brightness temperature, effective temperature, surface roughness, co-registration, and footprint support may be absorbed into the inverted τ targets. We also noted that the prior-uncertainty term partly accounts for this dependency, but does not represent a full formal error-propagation or variance-decomposition analysis (Lines 262-268).

#5 Additional methodological detail is needed regarding the Bayesian implementation. The paper states that Metropolis-Hastings sampling used about 1,000 iterations with 200 burn-in samples and convergence was assessed from trace stability. Please provide acceptance rates, number of chains, proposal distributions, parameter bounds, convergence diagnostics, and sensitivity to chain length.

Response: Thank you for this helpful comment. We added the Metropolis-Hastings sampler settings, state variables, chains, iterations, burn-in, retained samples, proposal distribution, bounds, acceptance-rate information, convergence checks, and chain-length test in Table S1 of Appendix.

#6 Effective temperature Tg is treated as an effective vegetation–soil composite temperature and adjusted using thermal/optical indicators. Please provide the exact correction equation, bounds, and whether thermal calibration/emissivity correction was applied.

Response: Thank you for this helpful comment. We revised the Methods to provide the implemented Tg correction equation, including Tg = T_{base} + ΔT, the optical-based ΔT term, and the ±4 K bound. We also clarified that no independent field emissivity calibration was available, so the thermal data were used as relative scene-level support for Tg and texture rather than as fully traceable absolute radiometric temperature (Lines 215-224).

#7 The study uses standard deviation of optical luminance and thermal temperature as texture descriptors. This is clear and robust, but it may not fully represent real field textures, e.g., row orientation, canopy gaps, or aspect structure. The authors should either justify why standard deviation is sufficient or add more texture metrics as comparison.

Response: Thank you for this helpful comment. We clarified that the standard deviation of optical luminance and thermal temperature was used as a simple and robust first-order proxy for footprint-scale dispersion, rather than as a complete representation of field texture. We now state that these descriptors can partly reflect mixed vegetation cover, exposed soil fraction, and thermal contrast, but do not explicitly capture row orientation, canopy gaps, directional structure, spectral variability, or viewing-geometry effects. We also added this as a limitation and noted that future work could incorporate additional texture metrics, object-based canopy descriptors, and multispectral or hyperspectral information (Lines 244-248 and 506-516).

#8 Results

The manuscript clearly demonstrates improvement in retrieval accuracy and reduction of dry bias. Crop-specific comparisons are found useful and informative.

Current bar plots show RMSE, MAE, and bias well, but scatter plots with 1:1 lines would better reveal whether improvement occurs across the whole moisture range or mainly from bias correction.

Please add scatter plots of retrieved vs observed soil moisture. Scatter plots of retrieved versus observed soil moisture would strengthen evaluation and help assess retrieval behavior across the full moisture range.

Response: Thank you for this helpful comment. We added retrieved-versus-observed scatter plots with 1:1 reference lines to provide a point-level evaluation across the full observed soil-moisture range. The revised text explains that the baseline retrieval is systematically shifted below the 1:1 line, indicating a substantial dry bias, whereas the proposed retrieval is more closely distributed around the 1:1 line with lower RMSE, MAE, and improved correlation (Lines 346-352; Fig. 5).

#9 Some land-cover-specific improvements, especially very large gains for wheat and maize may be based on small class sample sizes, but the number of samples per class is not reported. This makes it difficult to understand its robustness. Sample size for each crop class should be reported to better assess robustness of crop-specific improvements. Please add n, confidence intervals, or bootstrap uncertainty for each land-cover class.

Response: Thank you for this helpful comment. In response to this concern, we expanded the independent validation dataset from 57 to 219 matched validation points and updated the land-cover-specific evaluation accordingly. We also added the sample size n and 95% bootstrap confidence intervals for RMSE, MAE, and bias for each land-cover class in Table 1. The revised Results text now explicitly notes that, although the expanded dataset improves the robustness of the crop-specific comparison, class-specific improvements should still be interpreted with the reported sample sizes and confidence intervals in mind (Table 1).

#10 Thermal texture is significantly related to τ uncertainty but not to soil moisture error, while optical texture has stronger relationships. The discussion should more clearly state that the thermal contribution appears weaker or scene-dependent in this experiment.

Response: Thank you for this helpful comment. We revised the Discussion to more clearly distinguish the roles of optical and thermal texture. The updated analysis shows that optical texture was significantly related to posterior τ uncertainty and weakly related to posterior soil-moisture uncertainty, whereas its relationship with absolute soil-moisture error was not significant. Thermal texture showed weaker associations overall: it was only weakly but significantly related to posterior τ uncertainty, and was not significantly related to posterior soil-moisture uncertainty or absolute soil-moisture error. We therefore now interpret thermal texture as a secondary and more scene-dependent indicator in this experiment, rather than as a direct predictor of soil-moisture retrieval error (Lines 407-440; Fig. 7).

#11 The 4 m x 7 m ellipse performs slightly better than a 5 m x 5 m square. This is a useful result, but the improvement is noticeably small. Please discuss whether this difference is statistically significant and how footprint size/shape uncertainty affects the results.

Response: Thank you for this helpful comment. We added statistical information for the ellipse–square footprint comparison, including bootstrap confidence intervals and p values in Table 2. We also revised the Discussion to clarify that the improvement of the 4 m × 7 m ellipse over the 5 m × 5 m square is modest and class-dependent, rather than universally significant across all land-cover types. Finally, we explicitly noted that footprint size, shape, and the lack of antenna-pattern-weighted aggregation remain important uncertainty sources for future work (Lines 441-448; Table 2).

#12 Discussions

The discussion section is one of the stronger parts of the manuscript because it attempts physical interpretation rather than only reporting retrieval metrics. However, several interpretations become somewhat mechanistic despite being based mainly on regression/correlation analyses. The interpretation should therefore remain more cautious.

Response: Thank you for this helpful comment. We revised the Discussion to make the physical interpretation more cautious and consistent with the correlation-based nature of the analysis. Statements that previously implied mechanistic causality were rephrased as associations or physically plausible interpretations. We also clarified that the texture descriptors should be viewed as proxies for sub-footprint heterogeneity rather than proven causal drivers, and that posterior covariance analysis or formal variance decomposition would be needed for stronger mechanistic attribution (Lines 407-440).

#13 The texture analysis provides useful exploratory insight into retrieval uncertainty under heterogeneous canopies. But, statements suggesting that heterogeneity "primarily enters the posterior through τ-related uncertainty" appear somewhat stronger than what can be rigorously concluded from the presented regression analyses alone.

Response: Thank you for this helpful comment. We revised the Discussion to avoid overinterpreting the correlation results. We now state that footprint-scale texture was more clearly associated with τ-related posterior uncertainty than with soil-moisture retrieval error, rather than claiming a primary mechanistic pathway. We also note that this interpretation remains exploratory and would require posterior covariance analysis or formal variance decomposition for stronger attribution (Lines 421-430).

#14 The observed relationships are physically plausible, but additional posterior covariance analysis or formal uncertainty decomposition would be needed to fully support mechanistic uncertainty propagation claims. Texture descriptors likely act as proxies for broader sub-footprint heterogeneity rather than uniquely causal drivers of retrieval uncertainty. Several discussion statements become somewhat mechanistic despite the absence of formal uncertainty decomposition or covariance analysis.

The sensitivity analysis is useful and demonstrates that the retrieval framework is not excessively fragile to moderate hyperparameter perturbations. However, the current one-at-a-time sensitivity approach does not capture interaction of parameters, nonlinear coupling, or compensating effects among priors, likelihood weighting, and vegetation attenuation constraints. This limitation should be acknowledged more explicitly.

Response: Thank you for this helpful comment. We revised the Discussion to make the uncertainty interpretation more cautious. Texture descriptors are now described as proxies for broader sub-footprint heterogeneity rather than causal drivers, and we explicitly state that posterior covariance analysis or formal variance decomposition would be needed for stronger mechanistic attribution (Lines 421-440).

We also revised the sensitivity-analysis section to clarify that the one-at-a-time tests support local stability within the tested ranges, but do not represent a full global sensitivity analysis and cannot fully capture parameter interactions, nonlinear coupling, or compensating effects among likelihood weighting, τ-prior mean, prior strength, and τ-prior width (Lines 475-482; Fig. 8).

#15 The manuscript should also clarify how the "calibrated" hyperparameter settings were selected and whether there is risk of overfitting to the limited dataset.

Stability conclusions based mainly on RMSE curves may be somewhat overstated without examining posterior uncertainty behavior or parameter identifiability. The manuscript argues that relatively flat RMSE curves indicate robustness. However, flat RMSE behavior may also reflect limited sensitivity of the chosen validation metric, compensating parameter effects, or relatively narrow perturbation ranges. Therefore, it would help to additionally examine posterior uncertainty behavior, parameter identifiability, or posterior distribution shifts, rather than RMSE alone.

Response: Thank you for this helpful comment. We clarified that the calibrated hyperparameters were selected from the calibration scenes using RMSE as the main criterion, with posterior behaviour and physical bounds used as diagnostic checks; the independent validation dataset was used only for evaluation. We also acknowledged the potential risk of overfitting to the field-scale calibration dataset (Lines 299-304).

We revised the sensitivity analysis to include posterior diagnostics in addition to RMSE.

Figure 8 now reports the mean 95% credible-interval width of Ws and the posterior correlation between Ws and τ, and we clarify that the one-at-a-time tests indicate only local stability, not global identifiability or absence of parameter compensation (Lines 469-482; Fig. 8).

#16 Limitations and Uncertainty

The limitations section is generally honest and well presented. The authors acknowledge issues related to limited sample size, restricted observation dates, lack of full seasonal coverage, and the absence of detailed decomposition of different error sources. However, the discussion could be strengthened by more clearly separating and discussing the major uncertainty sources at least, such as radiometer measurement noise, TDR support mismatch, geolocation/co-registration uncertainty, thermal calibration uncertainty, vegetation prior uncertainty, and structural uncertainty associated with the τ-ω model.

Response: Thank you for this helpful comment. We revised the Limitations and Future Directions section to separate the major uncertainty sources more clearly. We now distinguish measurement errors, support-matching errors, and model-structure errors, including radiometer brightness-temperature noise, TDR support mismatch, geolocation/co-registration uncertainty, footprint assumptions, thermal calibration/emissivity uncertainty, vegetation-prior uncertainty, and τ–ω model simplifications. We also added that future work could use hierarchical error models, antenna-pattern-weighted aggregation, posterior covariance analysis, or formal variance decomposition to better partition these uncertainties (Lines 517-525).

#17 Conclusions

The conclusions are generally consistent with the presented results. The importance of heterogeneity-aware retrieval is clearly articulated. The conclusion section would further benefit from more clear separation of demonstrated findings, physically plausible interpretation, and future research directions.

Future studies should place greater emphasis on overcoming data scarcity issues, and performing formal uncertainty propagation through variance decomposition, which would further strengthen the findings of this study and improve the interpretability of the retrieval framework.

Response: Thank you for this helpful comment. We revised the Conclusion to more clearly separate demonstrated findings, physically plausible interpretation, and future research directions. Specifically, we distinguish the observed improvements in retrieval accuracy and bias reduction from the correlation-based interpretation of texture-related uncertainty. We also added future directions emphasizing broader cross-regional validation and formal uncertainty propagation, including posterior covariance analysis or variance decomposition (Lines 536-553).

#18 Minor/Editorial Comments:

Several remote sensing discipline-oriented terminologies are introduced throughout the article without any context or broad overview of them for readers beyond remote sensing discipline or for general readers. Please add their brief description either in footnote or in a Glossary as part of Supplementary materials with properly citing it to see those from Supplementary materials.

Double check first appearance of abbreviations/expansion and then abbreviations in subsequent sections.

Please check reference style as per HESS guidelines.

Response: Thank you for this helpful comment. We checked the first-use definitions of abbreviations and added a glossary of key remote-sensing and statistical terms in the Appendix. We also reviewed the manuscript for consistent terminology and revised the reference list according to the HESS style requirements (Appendix, Glossary of abbreviations).

Referee #2 comments:

#1 The authors developed a Bayesian retrieval framework for soil moisture estimation. The novelty of the manuscript relies on the utilization of UAV L-band brightness temperature and the RGB and thermal infrared information. The topic is interesting for different applications and for the community.

My main concern about the methodology is somehow not completely clear about the application of different algorithms, and then the evaluation of the presented framework is only through the RMSE and the standard deviation. Moreover, the application of this framework seems to be designed for local conditions. Even though the author recognised this limitation in the discussion section, it could be interesting to add how it can be transferred to other regions or areas.

Response: Thank you for this helpful comment. We expanded the evaluation beyond RMSE and standard deviation by distinguishing accuracy metrics from posterior uncertainty diagnostics and added transferability limitations for regional application.

#2 Major comments:

The framework has been developed to address the uncertainty in the retrieval information; however, the author limited the evaluation to computing the standard deviation. Other metrics or indicators could provide more detailed information about the uncertainty and their contributions to the final output.

Response: Thank you for this helpful comment. We expanded the uncertainty evaluation beyond posterior standard deviation. The revised manuscript now distinguishes accuracy metrics from Bayesian uncertainty diagnostics, including posterior standard deviation and 95% posterior credible intervals. We also added analyses linking texture descriptors to posterior uncertainty and τ uncertainty, and revised the sensitivity analysis to include the mean 95% credible-interval width of Ws and the posterior correlation between Ws and τ. Finally, we clarified in the limitations that a full decomposition of uncertainty contributions from radiometer noise, TDR support mismatch, co-registration, thermal calibration, vegetation priors, and τ–ω model structure remains an important direction for future work (Lines 407-416, 469-474, 517-525; Figs. 7 and 8).

#3 In subsection 2.3. The PoLRa flight provided information at a very high resolution (8 mm and 10cm), but subsequently, for validation purposes, the data were spatially aggregated to approximately 3 m. This raises a few questions: firstly, during spatial aggregation, the data may have been smoothed, particularly as the order of magnitude is so different; could you explain this in more detail? Secondly, how is spatial aggregation carried out? Could you provide further details on this step? Finally, if the data is aggregated, what is the main advantage of using very high-resolution data? Additionally, I suggest adding a scheme of the Validation procedure, because the text is sometimes confusing.

Response: Thank you for this helpful comment. We revised Section 2.3 to clarify that the validation labels were generated by footprint-consistent aggregation rather than by an arbitrary 3 m smoothing step. Repeated TDR readings were first averaged at each sampling location, and ground measurements within the assumed radiometric footprint were then averaged to form footprint-scale validation labels. The primary validation used the heading-oriented 4 m × 7 m ellipse, while the 5 m × 5 m square was used for the footprint-shape sensitivity test. We also clarified that this aggregation reduces support mismatch but inevitably smooths some point-scale variability. The high-resolution RGB-TIR data remain important because they preserve sub-footprint vegetation, temperature, and texture information for footprint-scale summarization. The revised workflow figure now includes footprint-consistent feature extraction, validation aggregation, baseline comparison, and uncertainty evaluation (Lines 153-181).

#4 Lines 165 to 173, this paragraph mentions that the aggregation step is central to the methodology. However, a more detailed explanation of this procedure is required, as the current text limits itself to merely mentioning it in that paragraph. This makes the methodology confusing and difficult to follow.

Response: Thank you for this helpful comment. We expanded the Methods to describe the aggregation procedure more explicitly. We now clarify that both validation labels and RGB-TIR auxiliary variables were generated through footprint-consistent aggregation rather than single-point or single-pixel sampling. Specifically, repeated TDR readings were first averaged at each sampling location, and ground measurements falling within the assumed radiometric footprint were then averaged to form footprint-scale validation labels. High-resolution RGB and TIR variables, including vegetation fraction, thermal state, and texture descriptors, were also summarized within the same microwave support area. The primary aggregation used the heading-oriented 4 m × 7 m ellipse, while the 5 m × 5 m square was used for the footprint-shape sensitivity test. We also updated Figure 3 to show the aggregation and validation steps more clearly (Lines 153-173; Fig. 3).

#5 Figure 3 presents the workflow of the paper. However, some steps are missing or could be included to expand the workflow. This would make the diagram easier to understand for the reader. For example, adding the validation process. And the comparison with the baseline is an important part of the paper.

Response: Thank you for this helpful comment. We updated the Figure 3 caption and Methods text to include preprocessing, validation, baseline comparison, and uncertainty evaluation in the workflow (Fig. 3).

#6 Lines 196 to 200. In this paragraph, mentioned that the single scattered albedo was parametrised and adjusted. This paragraph is somewhat confusing to understand; moreover, mentioned a physically reasonable range. How is this range defined? Is it based on the literature, or is it more a subjective range?

Response: Thank you for this helpful comment. We revised the Methods to clarify the parameterization of the single-scattering albedo and the τ prior. Specifically, we now define $\omega = \mathrm{clip}(\omega_0 + 0.03H, 0.00, 0.20)$, where ω_0 is the baseline single-scattering albedo and H is the normalized heterogeneity index. We also define the τ prior as $\tau = \mathrm{clip}(\tau_0 + 0.18\mathrm{FVC} + 0.10H, 0.00, 0.80)$, where τ_0 is the calibration-derived intercept term. The bounded ranges were used as conservative L-band crop-field constraints to prevent nonphysical solutions and to keep the inversion within the range supported by the calibration scenes and sensitivity tests, rather than as purely subjective or universal crop-independent constants (Lines 225-230).

#7 Line 246: Posterior inference was performed through Metropolis-Hastings sampling in the two-dimensional state space. Could you elaborate more about this step? What does this algorithm involve, and, above all, what is meant by the term ‘posterior inference’?

Response: Thank you for this helpful comment. We revised the Methods to explain posterior inference in more accessible terms. We now clarify that posterior inference means estimating plausible pairs of soil moisture Ws and vegetation optical depth τ that are consistent with the observed TBH/TBV, the τ–ω forward model, and the RGB-TIR-informed prior. We also added a brief explanation of the Metropolis-Hastings MCMC procedure and reported the sampler settings, including chains, iterations, burn-in, proposal distribution, bounds, acceptance rates, and convergence checks in Table S1 (Table S1).

#8 Please could you elaborate more about the calibration process and how the optimal parameters were defined?

Response: Thank you for this helpful comment. We revised the Methods to clarify the calibration process and the definition of the calibrated parameter settings. The hyperparameters were selected from the calibration scenes using RMSE as the main criterion, with posterior behaviour and physical bounds used as diagnostic checks; the independent validation dataset was used only for evaluation. We also expanded the sensitivity analysis to show how the calibrated settings relate to validation RMSE, posterior credible-interval width, and Ws–τ posterior correlation, while noting that these settings remain field-scale calibrated rather than universally transferable parameters (Lines 298-304; Fig. 8).

#9 In the result section, some interpretations are not directly supported by the presented analyses, e.g., lines 302-306, 336- 339, 368-371

Response: Thank you for this helpful comment. We revised the Results section to avoid interpretations that were not directly supported by the presented analyses. Statements that previously implied mechanistic attribution were either removed, rephrased as descriptive observations, or moved to the Discussion with more cautious wording. In particular, the land-cover-specific results are now presented as class-dependent performance differences rather than definitive attribution of error sources, and the texture-related interpretation is explicitly described as correlation-based rather than mechanistic. We also clarified that remaining explanations require further support from posterior covariance analysis or formal uncertainty decomposition (Results and Discussion).

#10 Minor comments:

Many abbreviations are not defined in the text; some of them are widely used in the remote-sensing discipline. However, the text would improve by adding a glossary.

Response: Thank you for this helpful comment. We added an Appendix glossary and checked first-use definitions for UAV, RGB, TIR, L-band, TDR, RMSE, MAE, CI, TBH/TBV, FVC, MCMC, and Metropolis-Hastings (Appendix).

#11 Figure 2 could be improved by including the location of the areas.

Response: Thank you for this suggestion. We revised Figure 2 by adding the study-area location map in panel (a-b), making the geographic location of the experimental areas clearer.

#12 Throughout the manuscript, the authors use the terms: metrics, statistics, and statistical metrics. Do these terms refer to similar metrics? If so, could the terms be standardised, or, if not, could the differences between them be specified?

Response: Thank you for this helpful comment. We standardized the terminology throughout the manuscript. RMSE, MAE, and bias are now referred to as accuracy metrics, posterior standard deviation and credible intervals as uncertainty diagnostics, and optical/thermal local standard-deviation variables as texture descriptors. We also removed the ambiguous term "statistical metrics" to avoid confusion.

#13 Figure 5, could you add proper labels to help the reader better understand the results?

Response: Thank you for this suggestion. We revised the figure, now shown as Figure 6, by adding land-cover labels to the representative fields and using a consistent color bar for the baseline and proposed retrieval maps, making the spatial comparison easier to interpret.

#14 In the discussion section, some paragraphs are particularly long and hard to understand. I suggest reorganizing and splitting some of them.

Response: Thank you for this helpful suggestion. We reorganized the Discussion section by splitting long paragraphs into shorter, more focused paragraphs. The revised text now separates statistical results, physical interpretation, methodological limitations, and future directions more clearly.

Referee #3 comments:

#1 This manuscript presents an interesting study on field-scale soil moisture retrieval using UAV-borne L-band passive microwave radiometry integrated with RGB and thermal infrared information. Overall, this research is timely and potentially valuable for hydrological remote sensing, precision agriculture, and field-scale soil moisture monitoring. However, several important methodological and validation issues need to be addressed before the manuscript can be considered for publication.

Response: Thank you for this helpful comment. We revised the manuscript to address the methodological and validation issues.

#2 Major comments:

Section 2.2 and Section 2.3: As described in these sections, the microwave observations were acquired by a DJI Matrice 600 platform carrying the PoLRa v3.1 L-band radiometer, and the effective microwave support area was estimated to be approximately 4 m * 7 m. However, the manuscript does not provide sufficient information on radiometric calibration, brightness-temperature uncertainty, incidence-angle control, antenna pattern, or footprint determination. Since the proposed method relies heavily on footprint-consistent aggregation of RGB-TIR features, these missing details may directly affect the credibility and reproducibility of the retrieval framework. Therefore, it is advised to provide the calibration procedure, TBH/TBV quality-control strategy, antenna footprint calculation, and uncertainty associated with the assumed footprint in the revised manuscript.

Response: Thank you for this helpful comment. We revised Sections 2.2 and 2.3 to clarify that TBH/TBV were obtained from the system-level radiometric processing of the PoLRa instrument and were further screened using physical plausibility, attitude stability, and spatial matching checks. We also added details on incidence-angle consistency, footprint determination, and the treatment of residual radiometric and footprint uncertainty through the microwave likelihood spread and footprint-shape sensitivity test (Lines 141-162, 182-193).

#3 Section 2.3 and Section 4.3: The calibration dataset includes only four scenes, approximately 1.3 ha, and 52 calibration points, while the independent validation dataset includes six scenes, approximately 3.3 ha, and 57 matched validation points collected over two consecutive days. These two days mainly represent pre- and post-rainfall conditions associated with a small rainfall event. Although this design is useful for a field-scale demonstration, it is not sufficient to demonstrate the general transferability of the proposed framework. The authors should moderate the related conclusions or provide additional validation across different crop growth stages, irrigation conditions, surface roughness states, and seasonal moisture regimes.

Response: Thank you for this helpful comment. We expanded the independent validation dataset from 57 to 219 matched validation points, covering approximately 12.6 ha across 13 scenes and five observation days. We also moderated the conclusions and clarified in Section 4.3 that, although the expanded dataset strengthens the field-scale evaluation, all validation data are still from the Pengzhou study area and therefore should not be interpreted as evidence of general regional transferability (Lines 174-181, 484-493).

#4 Section 2.5 and Section 3.1: The baseline comparison is not sufficient to demonstrate the necessity of all components of the proposed framework. The principal baseline is a Dobson plus dual-parameter τ-ω inversion without the RGB-TIR Bayesian prior. This comparison shows that the full proposed method performs better than the unconstrained baseline, but it does not reveal which component mainly contributes to the improvement. It is advised to conduct additional experiments to help readers better understand whether all proposed components are necessary.

Response: Thank you for this helpful comment. We added a component-wise comparison to evaluate the contribution of each part of the proposed framework. The revised Methods now define the baseline, +FVC, +FVC+TIR, and full +FVC+TIR+Texture configurations, and Figure 4 reports their validation performance under the same calibration-validation split. The results show that FVC provides the largest first-order improvement, while TIR and texture descriptors provide additional gains (Lines 305-311, 320-345; Fig. 4).

#5 Section 3.2 and Table 1: The manuscript reports very large improvements for several land-cover types. For example, wheat RMSE decreases from 0.0921 to 0.0088 $m^3 m^{-3}$, and maize RMSE decreases from 0.0857 to 0.0187 $m^3 m^{-3}$. These results are promising but may be strongly affected by the number of validation samples in each crop category. The manuscript should report the sample size for each land-cover type and provide confidence intervals for class-specific RMSE, MAE, and bias.

Response: Thank you for this helpful comment. We expanded the independent validation dataset and updated the land-cover-specific evaluation accordingly. The revised Table 1 now reports the sample size n and 95% bootstrap confidence intervals for RMSE, MAE, and bias for each land-cover class. We also revised Section 3.2 to interpret class-specific improvements more cautiously, noting that land-cover-specific results should be considered together with the reported sample sizes and confidence intervals (Table 1).

#6 Section 4.1 and Table 2: The manuscript reports that the ellipse footprint slightly outperforms the square footprint, with RMSE decreasing from 0.0443 to 0.0418 $m^3 m^{-3}$. The physical explanation is reasonable, but the numerical difference is small, only 0.0025 $m^3 m^{-3}$. The authors should report whether this difference is statistically significant and whether it holds across individual scenes, land-cover types, and moisture conditions. Additional tests using different footprint sizes or antenna-pattern-weighted aggregation would provide stronger evidence that footprint consistency is a major source of improvement.

Response: Thank you for this helpful comment. We revised Table 2 to report bootstrap confidence intervals and p values for the ellipse–square footprint comparison by different land-cover class. We also revised Section 4.1 to state more cautiously that the 4 m × 7 m ellipse generally outperformed the 5 m × 5 m square, but that the improvement was small and class-dependent rather than universally significant. We further clarified that additional tests using different footprint sizes or antenna-pattern-weighted aggregation remain important future work (Lines 441-448; Table 2).

#7 Minor comments:

Section 2.3: The date format “03/04”, “04/04”, and “05/04” is confusing. Please use an unambiguous format, such as “2026-04-03”, “2026-04-04”, and “2026-04-05”.

Response: Thank you for this helpful comment. We changed the ambiguous date format to unambiguous ISO dates in the manuscript text and captions, including 2026-04-03, 2026-04-04, and 2026-04-05 (Lines 174-179).

#8 Figure 4: The authors should clarify whether the bootstrap confidence intervals were calculated by resampling validation points, scenes, or both.

Response: Thank you for this helpful comment. We revised the Figure 4 caption and Methods text to state that bootstrap confidence intervals were computed by resampling matched validation points (Fig. 4).

#9 Table 1: Please add the number of samples for each land-cover class. This is essential for interpreting the reported crop-specific accuracy.

Response: Thank you for this helpful comment. We added land-cover sample size n in Table 1 and referred to it explicitly in the Results text (Table 1).

#10 Figure 5: It is advised to use a common color scale between the baseline and proposed retrievals to allow direct visual comparison in the soil moisture maps. In addition, the font size of the labels, color-bar tick values, and numerical annotations is too small, which reduces the readability of the figure.

Response: Thank you for this helpful suggestion. We revised the soil-moisture map figure, now shown as Figure 6, by using a common color scale for the baseline and proposed retrievals to support direct visual comparison. We also enlarged the field labels, color-bar tick values, and numerical annotations to improve readability.