We thank both reviewers for their careful and constructive assessments. The comments have prompted substantial revisions that, we believe, materially strengthen the manuscript. The principal changes are summarised below, followed by the detailed point-by-point responses to each reviewer.

## Summary of changes

1. **Abstract** — rewritten: the motivation is condensed to a single sentence, the specific innovations are stated up front, and quantitative accuracy and performance metrics are added (RMSE $\leq 2$ cm in 11 of 12 dry/transition cases; $\sim 4$ and $\sim 8$ cm in the wet phase; 6–12$\times$ faster than HYDRUS-1D; $\beta \approx 0.92$–$0.93$).
2. **Introduction** — a new concise paragraph names the closest hybrid GW–UZ–SW couplers and forward-references the new capability-comparison Table 1 (`tab:model-comparison`; GWSWEX vs. HYDRUS-1D, MODFLOW–UZF1, MIKE SHE, ParFlow, and conceptual bucket models).
3. **Section 2 (model description)** — opens with an explicit design-philosophy statement and frames GWSWEX as an intermediate-complexity alternative rather than a replacement for fully integrated 3-D Richards models.
4. **Methods** — each experimental setup now carries an explicit evaluation objective (*basic*: slow-recharge, deep-water-table, capillary-fringe stress test; *intensive*: saturation-excess stress test raising a shallow water table to the surface); a verification-versus-validation scope statement is added; and the justification of HYDRUS-1D as a like-for-like 1-D community-standard benchmark is expanded.
5. **New parameter-classification Table 6** (`tab:param-class`) — every model parameter is classified as physically based, empirical/semi-physical, or purely numerical, with typical ranges and calibration guidance; the table is preceded by explanatory text.
6. **Figures** — fully revised for accessibility and clarity: a colour-vision-deficiency-safe palette with redundant line-style coding (greyscale-separable); transparency and draw-order on overlapping traces; a perceptually-uniform heatmap colour map; enlarged sans-serif type with in-figure legends; vector PDFs at the journal widths. The figure-generation pipeline is updated in `scripts/build_figures.py`.
7. **Figure captions (Figs 3–7) and Results text** — rewritten to interpret the physics: the transmission-zone lag underlying the basic dry-phase recession; the saturation-excess mechanism by which clay floods fastest despite the lowest $K_{\mathrm{sat}}$; and the localisation of the intensive-clay GWSWEX–HYDRUS residual to the saturated–unsaturated capillary coupling zone.
8. **Results** — a new per-step water-balance analysis of the intensive-loam saturated window: GWSWEX routes the residual $P - E_{\mathrm{act}} - T_{\mathrm{act}}$ to surface-water storage at prescribed-potential ET (the only mass-conservative outcome at $\theta_s$); HYDRUS-1D root uptake is $\approx 8.4\times$ the prescribed $T_p$ (0.84 vs. 0.10 mm h$^{-1}$), giving peak ponding $\sim 16$ vs. $\sim 75$ mm. The HYDRUS-1D manual, `SINK.FOR`, and the Feddes literature were examined and the excess reported as an empirical observation.
9. **Discussion** — a new "innovation-to-outcome" opening synthesis links each design choice to a measured outcome; the ponding-mechanism analysis is expanded with the routing/runoff coupling caveat; and a sentence on the nonlinear, mutually interacting nature of the parameter responses cites Morris (1991), Sobol' (2001), Saltelli et al. (2010), and Pianosi et al. (2016).

10. **Limitations** — expanded to name validation against measured groundwater levels, soil moisture, runoff, and evapotranspiration as the immediate successor experiment (underway), and to state the code-to-code verification scope boundary explicitly.

---

# Response to Reviewer 1

We thank the reviewer for their thorough and constructive assessment of our manuscript. Their valuable insights regarding the structure of the Introduction, the formatting and accessibility of our figures, the need for a model-comparison matrix, the role of field validation, the classification and calibration of model parameters, and the non-linear nature of parameter sensitivity have helped us significantly improve the quality and clarity of the manuscript. Below we address each comment in detail.

## 0.1 Comment 1

> *The Introduction should be condensed and reorganized to clearly highlight the importance of GW-UZ-SW coupling, the limitations of existing 3-D Richards-equation-based models and bucket-type models, and the specific novelty of GWSWEX. When presenting the results, the authors should improve the readability of figures and tables.*

**Response:** We appreciate the reviewer highlighting this. Making the novelty of GWSWEX immediately apparent to the reader is important, and the original manuscript could have done more to achieve this. While the Introduction was already designed to follow a progressive structure: motivating GW–UZ–SW coupling (paragraph 1), detailing the limitations of 3-D Richards and bucket-type models (paragraph 2), identifying the remaining literature gap (paragraph 3), and presenting GWSWEX's specific contributions (paragraph 4), we recognize that GWSWEX's position relative to existing models was not made sufficiently explicit within this flow. To address this, we have added a new concise paragraph to the Introduction that identifies the closest hybrid GW–UZ–SW couplers and directs the reader to a newly added capability-comparison table (Table 1). This provides a clearer basis for comparing GWSWEX with existing frameworks and makes the specific novelty more immediately apparent to the reader.

The readability of figures and tables is addressed below under Comment 2.

## 0.2 Comment 2

> *Some figures contain multiple soil profiles, forcing scenarios, and solver comparisons, which makes them difficult to interpret. The layout, font size, legends, and color scheme should be improved to better highlight the key findings. When emphasizing the novelty of GWSWEX, the authors should provide a clearer comparison with existing models.*

**Response:** We have comprehensively updated all figures to improve their clarity, visual consistency, and accessibility. The revised figures employ a color-vision-deficiency-friendly palette combined with distinct line styles, ensuring that individual series are fully distinguishable in grayscale. We have also optimized line transparency and draw-order to prevent overlapping curves from obscuring data, enlarged font sizes, and refined legend placement. The sensitivity heatmap has been updated with a perceptually uniform color map. All figures are now provided as high-resolution vector graphics formatted according to the journal's size recommendations.

To address the suggestion for a clearer comparison with existing models, we have introduced a comprehensive capability-comparison table (Table 1), which is detailed in the response to Comment 3.

## 0.3 Comment 3

*A comparative table is suggested to compare GWSWEX with HYDRUS-1D, MODFLOW-UZF, MIKE SHE, ParFlow, and conventional bucket models in terms of vertical resolution, groundwater-head tracking, capillary rise, mass conservation, computational efficiency, and coupling capability. When validating the proposed model, the authors should consider using field observations.*

**Response:** We agree that a structured comparison table is an excellent way to clarify GWSWEX's relative position. We have added Table 1 to the manuscript, which evaluates GWSWEX against HYDRUS-1D, MODFLOW–UZF1, MIKE SHE, ParFlow, and conceptual bucket models. The comparison covers subsurface dimensionality, representation of matric-potential gradients, prognostic groundwater-head dynamics, capillary rise, constitutive formulation, per-step mass conservation, suitability for ensemble applications, two-way GW–UZ–SW coupling, and open-source availability. The table is introduced in the revised Introduction, and the practical implications are discussed in the Discussion section.

Our response regarding validation against field observations is detailed under Comment 4.

## 0.4 Comment 4

*The current verification mainly relies on comparisons with HYDRUS-1D under idealized numerical scenarios. Field data, such as observed groundwater levels, soil moisture, runoff, or evapotranspiration, would help demonstrate the practical applicability of GWSWEX.*

**Response:** We agree entirely that field validation is the essential next step and is the focus of our ongoing follow-up work. The present paper is, by design, a model description paper. Its primary objective is to evaluate the numerical behavior, accuracy, and mass-conservation properties of GWSWEX's dual solvers under perfectly controlled conditions where forcing, parameters, and boundary conditions are known without uncertainty. This isolation of numerical performance is necessary to separate the model's mathematical fidelity from the significant uncertainties introduced by field-scale parameter estimation and atmospheric data.

To clarify this scope, we have revised the manuscript to explicitly distinguish between the controlled numerical verification presented here and subsequent field validation. Additionally, we have expanded the Limitations section to state clearly that the current work does not constitute a full field validation and to identify field-based evaluation as the immediate next phase of model assessment.

## 0.5 Comment 5

*The manuscript should clarify which parameters are physically based, empirical, or numerical. Recommended parameter ranges and calibration guidance for different soil types and hydrological conditions would improve the practical value of the model.*

**Response:** We thank the reviewer for this valuable suggestion, which significantly improves the practical utility of the model. We have added Table 4 to the manuscript, which systematically classifies every GWSWEX parameter into one of three categories: physically based, empirical or semi-physical, or numerical. For each parameter, the table provides typical physical ranges, along with practical guidance on calibration and interpretation across different soil and vegetation types. This table is preceded by explanatory text outlining how parameters should be constrained or calibrated.

### 0.6 Comment 6

*When discussing the nonlinear responses of soil hydraulic parameters, ET-stress thresholds, infiltration parameters, and inter-layer connectivity parameters, the authors should consider relevant recent studies on nonlinear correlations and predictive modelling, including: https://doi.org/10.1029/2024JB029018 https://doi.org/10.1029/2023WR036494 https://doi.org/10.1029/2023WR036203*

**Response:** The parameter responses in GWSWEX are indeed highly non-linear and mutually interacting, and we agree that emphasizing this is important. We have prepared the sensitivity-analysis discussion to make these non-linear parameter interactions explicit, referencing established global sensitivity analysis literature (Morris, 1991; Sobol', 2001; Saltelli et al., 2010; Pianosi et al., 2016) to support the discussion.

Regarding the three suggested references (10.1029/2024JB029018; 10.1029/2023WR036494; 10.1029/2023WR036203): we have reviewed these studies and found that they focus on non-Darcy and inertial flows in rough, partially filled rock fractures at the pore/fracture scale. GWSWEX, by contrast, operates macroscopically within the Darcy regime of unsaturated porous-media flow under sub-saturated forcing, which excludes the pore-scale fracture dynamics and inertial regimes discussed in those papers. Consequently, we have not included these specific citations, but we hope the reviewer finds the strengthened discussion of non-linearity and parameter interaction, grounded in directly relevant sensitivity-analysis literature, an appropriate response to the underlying concern.

---

## Response to Reviewer 2

We thank the reviewer for their comprehensive, rigorous, and highly constructive review of our manuscript. Their insights regarding the connection between GWSWEX's design innovations and its numerical performance, the formatting of the abstract, the description of model design philosophy, the physical representation and real-world applicability of our experimental scenarios, the physical interpretation of the surface-ponding discrepancies, and the justification of our benchmarking approach have significantly strengthened the paper's scientific framing and clarity. Below we address each of the comments in detail (numbered for clarity).

### 0.7 Comment 1 (General Comment)

*This paper presents a new hybrid model named GWSWEX, which is developed to address the limitations of sophisticated three-dimensional models and simple bucket-type models, enabling groundwater and runoff simulations with satisfactory accuracy and low computational cost. The authors also attempt to evaluate the accuracy and computational efficiency of this newly developed hybrid model by comparing it with a* *one-dimensional model under specific scenarios. The results reveal that the proposed hybrid model features remarkably less computation time while achieving accuracy comparable to the reference model. Nevertheless, the authors fail to clearly elaborate the rationale behind the comparative experiments, and their review of existing research progress is overly general and merely conceptual without detailed benchmarking against previously developed hybrid models. Even in the comparison with the selected 1D model, the discussion of results inadequately interprets the link between the model's innovations and its numerical performance, which greatly hinders readers' understanding of the novelty of this study. It is suggested that the authors elaborate on the innovations of the proposed hybrid model in greater detail, especially clarifying how these innovations contribute to improvements in simulation performance.*

**Response:** We appreciate the reviewer raising this as a central concern; it identifies what was, in retrospect, the most consequential weakness of the original manuscript. The connection between GWSWEX's design choices and the numerical behavior they produce was present in our results but was not drawn out explicitly enough for the reader, and the rationale behind the experimental design was presented without sufficient context to make its purpose self-evident. We have made substantial revisions to address each of these points:

1. *Rationale behind the comparative experiments:* We have added explicit statements of the physical evaluation objectives for each experimental setup in the Methodology section (Sect. 2.1). The *basic* scenario is designed to isolate gravity-driven drainage, ET-induced recession, and capillary exchange under a slow-recharge, deep-water-table regime representative of dominant field-scale seasonal water budgets. The *intensive* scenario stress-tests the transition to full-column saturation and dynamic top-boundary condition switching under sustained high-intensity forcing. We have also expanded the justification for selecting HYDRUS-1D as the benchmark: rather than serving as a proxy for a full 3-D integrated model, it provides a rigorous, like-for-like 1-D numerical reference for evaluating the fundamental physics and numerical behavior of GWSWEX's core unsaturated-zone component in isolation from lateral routing complexities — a necessary first step before coupling introduces additional confounding factors.
2. *Benchmarking against existing hybrid models:* The new Table 1 provides a structured capability comparison against representative integrated, intermediate-complexity, and bucket models, giving readers a concrete basis for situating GWSWEX's position. We fully agree, however, that a detailed numerical benchmarking of GWSWEX as a fully coupled integrated hydrological model against other hybrid frameworks is essential. Such an evaluation requires a complete integrated modeling setup with external GW and SW solvers, which is beyond the scope of this initial model description paper. We have revised the Discussion (Sect. 4.5) and Limitations sections to explicitly identify direct numerical benchmarking against coupled hybrid models and real-world field evaluations as the immediate next phase of model assessment.
3. *Linking model innovations to numerical performance:* We have added a new opening synthesis to the Discussion section (Sect. 4) that explicitly connects each of GWSWEX's core design innovations to the observed quantitative outcomes: the compact in-memory handling of state and output yields the sub-linear scaling exponents; the dual-solver architecture permits the regime-matched accuracy visible in the phase-resolved metrics; and the strict per-step mass-conservative formulation prevents spurious fluxes, enabling the outer GW–UZ–SW coupling iteration to converge without under-relaxation. This synthesis is then developed through the subsections that follow.

We hope these revisions address the reviewer's concern substantively and hope they make the novelty and performance contributions of the model transparent to the reader.

## 0.8 Comment 2 (Specific Comment on Abstract)

*The abstract is overly lengthy with excessive elaboration on research motivation, while lacking descriptions of the innovations of GWSWEX and its specific accuracy metrics. The current version reads more like a project summary rather than a standard academic abstract.*

**Response:** The abstract has been significantly revised and condensed. We shortened the research motivation to a single introductory sentence, clearly stated GWSWEX's primary methodological innovations up front, and added specific quantitative metrics of model accuracy and computational performance from our benchmark results.

## 0.9 Comment 3 (Specific Comment on Section 2)

*Section 2: As a research article, the model overview section should explicitly specify the advanced techniques or design philosophies adopted in model development, as well as the underlying reasons why the proposed model outperforms existing 3D models and bucket-type models. However, this section mainly describes each component of the model, which fails to meet the academic communication requirements of a research paper.*

**Response:** We agree that presenting the underlying design philosophy explicitly improves the paper's academic context. The model description section now opens with an explicit design-philosophy statement: GWSWEX deliberately trades three-dimensional subsurface flow details for per-element computational efficiency, physically interpretable constitutive relations, and a prognostic moving groundwater head that can interface directly with external solvers. These features distinguish GWSWEX from conventional conceptual bucket schemes, while its 1-D vertical column structure and split-operator solver formulation make it computationally viable for regional-scale ensembles where fully integrated 3-D Richards models are prohibitive.

We have also ensured that the manuscript does not frame GWSWEX as a replacement for fully integrated 3-D models, but rather as an intermediate-complexity alternative designed for specific regional-scale applications. This design philosophy is directly linked to the performance results in the new opening synthesis of the Discussion.

## 0.10 Comment 4 (Specific Comment on Line 125)

*Line 125: "Two forcing setups probe complementary regimes." Although the authors state that two complementary experimental schemes are designed to evaluate model performance, the general applicability of these experimental configurations needs further clarification, along with the real-world environmental conditions they represent. If these are idealized numerical experiments, the authors should also specify the general evaluation objectives for adopting these two experimental designs.*

**Response:** The two configurations are indeed idealized numerical stress tests designed to isolate and evaluate specific numerical behaviors under controlled conditions, and we have clarified this in the Methods section. We have also explicitly tied each configuration to the real-world hydrological regimes and environmental conditions they represent: - The *basic* scenario represents a slow-recharge, deep-water-table, active-capillary-fringe regime typical of dominant field-scale seasonal water budgets. Its objective is to evaluate gravity-driven drainage, evapotranspiration-induced recession, and capillary exchange under smooth daily atmospheric forcing. - The *intensive* scenario represents a sustained, high-intensity wetting regime that raises a shallow groundwater table to the soil surface. Its objective is to test the transition to full-column saturation, specifically evaluating dynamic top-boundary condition switching in the implicit solver and the activation of the saturated-column formulation in the explicit solver.

By distinguishing these regimes, we clarify why different soil profiles (e.g., clay vs. sand) exhibit contrasting surface ponding and recession behaviors, linking the numerical experiments directly to physical mechanisms.

### 0.11 Comment 5 (Specific Comment on Lines 221–223)

> *Lines 221–223: "In the intensive scenario, both GWSWEX solvers report higher peak ponding and longer ponded duration than HYDRUS-1D in the four simulations" What physical implications do these results carry? Does this observation indicate that GWSWEX yields reasonable or biased outputs? What potential mechanisms could account for such discrepancies?*

**Response:** We have conducted a detailed, step-by-step water-balance analysis of the saturated interval in the intensive scenario to investigate this discrepancy. The results, which have been added to the Results and Discussion sections of the revised manuscript, reveal that the discrepancy is structural rather than a numerical error or bias in GWSWEX. We also acknowledge that our previous explination of terming it "structural" without further details might have been inadequate and therefore, we have followed it up with a more detailed explination.

*Mechanism:* The difference stems from how the two models partition water once the soil profile becomes fully saturated ($\theta = \theta_s$), the lower boundary flux is negligible (verified as zero), and additional atmospheric inputs can no longer be stored in the soil pore space. - In GWSWEX, evaporation is drawn from the surface pond and transpiration from the soil layers up to their prescribed potential rates. The remaining net rainfall ($P - E - T$) is routed to surface-water storage. Since the soil column is saturated and bottom flux is zero, this is the only mass-conservative outcome. - In HYDRUS-1D, during the same saturated interval, the diagnostic outputs show an actual cumulative root-water-uptake flux that is approximately 8.4 times larger than the prescribed potential transpiration rate (e.g., 0.84 mm $h^{-1}$ actual vs. 0.10 mm $h^{-1}$ potential in the intensive loam case). This excess internal uptake is sustained by continued infiltration from the surface pond, which artificially lowers the remaining ponded surface storage in HYDRUS-1D (keeping its peak at ~16 mm, compared to GWSWEX's mass-conservative ~75 mm). We reviewed the HYDRUS-1D documentation and the literature on the Feddes root-water-uptake formulation but found no standard physical justification for root-water-uptake exceeding potential rates under saturated stress. We therefore report this as an empirical discrepancy in HYDRUS-1D's water-balance partitioning.

*Reasonable or biased?* The GWSWEX output is physically reasonable and strictly mass-conservative; it represents the exact mathematical residual of the specified boundary conditions when subsurface storage is exhausted. The discrepancy does not indicate a bias or mass leak in GWSWEX, nor does it propagate into the groundwater-head signal.

*Physical and Practical Implications:* The main groundwater head and subsurface storage dynamics are unaffected, as both models correctly represent a saturated column. However, there is a practical implication for coupling: if GWSWEX's surface-water storage is coupled to a downstream routing or runoff model, the ponding-depth series will read higher and persist longer than a HYDRUS-1D reference under saturated conditions. We have added this caveat to the Discussion, noting that field-scale validation or comparison with integrated models will be necessary to benchmark the surface partition under real-world conditions.

## 0.12 Comment 6 (Specific Comment on Figures 3–7)

> *Figures 3–7: Descriptions of these figures only identify the variables illustrated, without interpreting the physical characteristics shown in the figures or linking these features to the physical mechanisms embedded within the GWSWEX model. The authors are advised to either incorporate observational ground truth data for comparison or thoroughly discuss the physical meaning of the discrepancies between different models reflected in these figures.*

**Response:** We have updated the captions of Figures 3–7 and expanded the corresponding discussion in the Results section to interpret the physical characteristics and mechanisms shown: - We now explain the transmission-zone delay that governs the dry-period recession lag in the basic scenario. - We discuss the saturation-excess mechanism that causes the clay profile to pond rapidly despite having the lowest hydraulic conductivity. - We explain the localization of the GWSWEX-HYDRUS-1D discrepancy in the clay profile to the capillary coupling zone above the water table.

Regarding observational data, the present paper is designed strictly as a code-to-code numerical verification. We have expanded the Limitations section to explicitly acknowledge this scope boundary and to frame validation against field-scale observations as the immediate next step in model development.

## 0.13 Comment 7 (Specific Comment on Line 482)

> *Line 482: "Several GW–UZ–SW couplers in the same architectural family warrant specific mention." It is recommended to relocate this part to the introduction section. Additionally, the authors need to provide a clear justification for selecting HYDRUS-1D as the benchmark instead of conducting direct comparisons with these existing hybrid coupling models mentioned herein.*

**Response:** We have addressed both points: 1. We moved a concise overview of the most closely related hybrid GW-UZ-SW couplers to the Introduction, referencing the new Table 1. The detailed model-by-model comparison is retained in the Discussion to keep the Introduction focused, and the sections now cross-reference each other. 2. We have clarified the justification for selecting HYDRUS-1D as our benchmark in the revised manuscript. GWSWEX introduces a new 1-D vertical unsaturated-zone *component*. To evaluate its core physics (e.g., matric-potential gradients, boundary switching, and column mass balance), we must isolate it from the confounding influences of particular 2-D or 3-D groundwater/surface-water routing solvers. HYDRUS-1D is the natural, like-for-like reference: it solves the same 1-D mixed-form Richards equation with Mualem-van Genuchten physics, isolating the GWSWEX implicit solver's accuracy and the explicit solver's cascade abstraction under identical soil properties, geometry, and boundary conditions. Furthermore, HYDRUS-1D is the widely accepted community standard for unsaturated-zone flow. Benchmarking against a fully coupled integrated model would conflate the performance of the unsaturated-zone component with that of the surrounding coupler, which we leave to future coupled-model studies.