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).

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.

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.

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.

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.

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.

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.

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.