Reviewer #1

This is a well-written and timely paper that provides a comprehensive overview of two decades of progress in ParFlow–land surface model coupling. It offers clear insights and a thoughtful vision for future sustainable model integration. I've included a few specific suggestions and questions to help improve clarity and flow in certain sections, but overall, I find the paper to be of high quality and fully supportive of its publication after minor revisions.

We sincerely thank the reviewer for carefully reading our manuscript and providing valuable suggestions. Please find our point-by-point responses below. Thank you again for your constructive comments.

Abstract

- Lines 17–24: The first two sentences are quite long. Consider splitting after "Earth system models (ESMs)" to improve readability.

  Suggestion: "Groundwater plays a vital role in terrestrial water and energy cycles. Yet, it remains oversimplified in most Earth system models (ESMs)…"

  Revised. Please see line 18 in the revised clean manuscript. All line numbers in the following response refer to those in the changes tracked manuscript.

Introduction

- Lines 44–52: The sentence from "It introduced the coupling…" to "…incorporating physically based groundwater dynamics…" is long and dense.

  As suggested by the reviewer, we revised the first paragraph in introduction for clarity. The previous long and dense sentence has been streamlined and rewritten into several clearer statements. Please see lines 45–52.

- Line 66: The phrase "groundwater models are often run offline with limited interaction…" could be clearer if you add "thereby missing dynamic feedbacks with the land–atmosphere system."

  Thanks! Added and please see lines 68–69.

- Lines 77–87: The transition from describing model evolution to the need for a "sustainable framework" is abrupt.

We have rewritten this paragraph to provide a smoother logical flow. Please see lines 92–104.

Section 2 – Review of Coupled Modeling

- Line 109: "Three key areas" — consider making these bold or in numbered format (1), (2), (3) for visual clarity.

  We have changed the number format to (1), (2), and (3). Please see lines 139–144.

- Lines 116–124: The paragraph is rich in technical context but could use a short topic sentence like "Groundwater strongly modulates the surface energy balance through its effect on soil moisture distribution."

  We rephrased the first sentence to a topic sentence. Please see lines 150–151.

- Lines 147–153: Consider simplifying the long sentence beginning with "Simulation results showed…" into 2–3 shorter sentences.

  We split the long sentence to shorter sentences. Please see lines 185–191.

- Lines 213–224: The description of rooting depth could be condensed or summarized for focus; much of the detail could be cited rather than explained.

  Thanks for the suggestion. This paragraph provides a complete explanation of how rooting depth influences the range of WTD critical zone. Therefore, we kept the overall level of detail. However, following the reviewer's suggestion, we added additional references to support the discussion. Please see lines 288, 292–295.

Section 3 – Foundational Recoupling

- Lines 319–324: The transition from "domain description" to "vertical discretization" is abrupt.

  We added a bridging sentence to improve the logical flow. Please see lines 399–401.

- Lines 378–386: The sentence "The overall advancement in model performance…" is strong; consider adding a quantitative statement (e.g., RMSE reduction or bias improvement) if available.

Great suggestion. We added RMSE in Figure 3. Please see the new figure.

Section 4 – Sustainable Framework

- Lines 470–483: Clarify how "logging system" differs from a typical software version control or provenance tracker — perhaps add a brief example.

Thanks for the question. The logging system described here is not a general software version-control or provenance-tracking tool. Instead, it specifically records the evolution of each component model that affects the coupling interface—for example, changes in how a module defines its coupling-relevant fields. For example, the introduction of the plant hydraulics module requires new soil hydraulic conductivity fields that are not part of the current interface; such changes would be explicitly logged. We have further clarified this distinction in the revised manuscript (lines 562–569).

Section 5 – PLCMIP

- Lines 502–517: The description of existing models could benefit from a small schematic figure showing coupling hierarchies.

We thank the reviewer for the valuable suggestion. We agree that a schematic figure can be helpful for illustrating coupling hierarchies. However, since our manuscript focuses on the recoupling framework rather than a detailed comparison of existing model architectures, adding a schematic with structural details may distract readers from the main message. To maintain clarity, we therefore keep this section concise while improving the textual description. Please see lines 630–633.

- Lines 529–537: Replace "To quantify... To compare... To identify..." with bullet points for easier reading.

Thanks! We added bullet points. Please refer to lines 649–657.

Section 6 – Summary

- Lines 547–558: Strong conclusion, but it repeats parts of the abstract.

Suggestion: Focus this section on future directions — e.g., "Key next steps include expanding the recoupling to other LSMs, refining coupler standards, and launching PLCMIP."

We thank the reviewer for the insightful suggestion. We agree that highlighting future directions will strengthen the conclusion, and we have revised this section accordingly by adding a clearer topic sentence and sharpening the discussion of next steps. At the same time, we retained a brief summary of the main contributions to ensure that the conclusion reflects both the key findings in the past and the forward-looking perspective. Please see lines 682–688.

- Table 1. I would also include this paper published recently on ParFlow-NASA/LIS-Noah-MP. Abbaszadeh, P., Zaouna Maina, F., Yang, C., Rosen, D., Kumar, S., Rodell, M., and Maxwell, R.: Coupling the ParFlow Integrated Hydrology Model within the NASA Land Information System: a case study over the Upper Colorado River Basin, Hydrol. Earth Syst. Sci., 29, 5429–5452, https://doi.org/10.5194/hess-29-5429-2025, 2025.

  Added. Please see line 628 and the updated table.

Lines 299–306: When you mention "basic water and energy modules of CoLM," which specific subroutines were included or excluded (e.g., snow, plant hydraulics)?

In this study, "basic water and energy modules" refer to the core components required to close the surface water and energy budgets, including canopy interception, snow processes, and the surface energy balance (radiation, sensible and latent heat fluxes, and ground heat flux). Modules such as plant hydraulics, urban, and other recently added extensions in CoLM were not activated in the current recoupling tests, as they are beyond the scope of this initial technical evaluation. We have clarified this in the revised manuscript (lines 374–376).

Lines 343–356: Why was 2018 chosen as the baseline year? Does this year represent typical hydrometeorological conditions for the domain?

The study area was clipped from the CONCN model domain where the model evaluation was performed for the year of 2018. We therefore selected 2018 as the baseline year in this study. We clarified this in the revision. Please see lines 421–422.

Lines 379–381: You note timing discrepancies in SWE accumulation. Have you explored adjusting the snow albedo or canopy interception parameters to resolve this?

We thank the reviewer for this insightful suggestion. Adjusting snow albedo or canopy interception parameters is indeed a promising way to further diagnose the timing discrepancies in SWE accumulation. However, such parameter sensitivity experiments are beyond the scope of this initial recoupling study. We have added a brief discussion of this point in the model comparison section of the revised manuscript (lines 652–653).

Lines 395–398: Could the deeper water table in CoLM/PF be partly influenced by changes in infiltration or soil parameterizations, rather than transpiration alone?

Yes, this is possible. Figure 4b represents the combined influence of both infiltration and evapotranspiration (ET): positive anomalies indicate higher infiltration, whereas negative anomalies with larger magnitudes reflect enhanced transpiration, which is particularly evident in Figure 4b. The soil configuration was kept as consistent as possible between the two models. Although some discrepancies are unavoidable due to differences in soil parameterizations, these differences are minor and have been noted in the manuscript (lines 495–497).

Lines 531–537: Will PLCMIP include standardized benchmark datasets or synthetic cases (similar to IH-MIP) to ensure comparability?

Yes. Thanks for this great point. We added it into the revised manuscript. Please see lines 658–661.

Reviewer #2

Thank you for the opportunity to review the manuscript titled “Twenty Years of Trials and Insights: Bridging Legacy and Next Generation in ParFlow and Land Surface Model Coupling” by Yang et al. In this manuscript, the authors first reviewed milestones in coupling a groundwater model, ParFlow, to land surface models and atmosphere models. Then the authors presented their preliminary results in updating ParFlow-CLM to the most recent version of ParFlow-CoLM. They ended by reviewing current models that have incorporated ParFlow, and providing high-level recommendations for future model coupling efforts. My major and minor comments are detailed below.

We appreciate the reviewer’s constructive comments. Please see our point-by-point responses below. All line numbers in the following response refer to those in the changes tracked manuscript.

1. I struggle to decipher the type of this article. It was not clear whether this is intended to be a review, insight, or research paper. The current presentation has a feel of mixture. Section 2 is a review of previous ParFlow-CLM and ParFlow-WRF efforts. Section 3 seems to reporting the preliminary ParFlow-CoLM efforts. Section 4 outlines insights moving forward, although a brief table reviewing other ParFlow efforts were also included. Readers would prefer the authors having a clear target where they want this manuscript to be and do a good job at that goal, instead of squeezing 3 very different goals in one paper but only scratching the surface. Personally, I would be very excited about the insights component. I wish the authors

could detail their implementation of the sustainable coupling, which would be of great benefit to the community as ParFlow or the LSMs will upgrade inevitably.

We thank Reviewer #2 for the suggestion. We understand R#2's confusion. While the manuscript does describe the connections among the three components, we acknowledge that the explanation may have been relatively weak. To improve overall readability and help readers follow the logical flow, we added an overarching summary in the Introduction (lines 94–121) and provided additional clarifications at the beginning of each section (lines 145–148, 371–394, 520–521, 645–648).

Here, we would also like to further clarify the connections among the three parts and why we organized the manuscript in this way. We intend to include three aspects in this paper at this important twenty-year juncture: reviewing the past, presenting our preliminary current exploration, and outlining the future plan. We regret the reviewer's misunderstanding, as each part is based on substantial work rather than being superficial.

- We conducted an in-depth review and synthesis of nearly twenty years of scientific findings based on ParFlow–land surface and ParFlow–atmospheric coupled models. We summarized and explained the groundwater–land–atmosphere interactions and clarified the role of groundwater in terrestrial water and energy cycles, again presenting the importance of groundwater in these processes to a broader range of readers. This also shows why such coupled modeling is necessary and forms the basis and motivation for our future coupling efforts.
- We also implemented the coupling between the latest CoLM and ParFlow through the existing interface, which is not trivial. This required a large amount of work because CoLM's code and data structures have undergone significant changes. This is also why we hope to use this process to guide a large-scale, structured development effort based on coupler in the next 3–5 years (or even longer):
    - quantitatively evaluate whether the parameterization improvements in land surface models over the past twenty years have improved model performance, and by how much, which provides theoretical support for further coupling. To our knowledge, such a comparison has not been done before, and we hope this could serve as a starting point for future model comparison efforts;

    - understand the current CoLM code structure and data structure, so that large-scale coupler-based coupling work can be carried out more efficiently;
    - provide the community with a new CoLM option immediately, without waiting 3–5 years or longer for a full system-level update.
- We also summarized the future perspective, which can better guide the next steps and stimulate community thinking and participation.

Therefore, the positioning of this manuscript is to bridge the past and the future, providing a comprehensive summary of the past, the present, and the future plan, and to clarify where we currently stand.

Regarding the sustainable coupling, please refer to comment 4 in this response letter.

2. The scope of this work seems very narrow. The authors focused on parflow-CLM and parflow-WRF, both of which are efforts led by the author team themselves. Some other modeling efforts were also mentioned, but to an extremely brief extent (one sentence, or one line in a table). To make it appealing to the broader community, I suggest the authors looked into work beyond their own. For instance, in addition to CLM, several other LSMs have also be coupled to ParFlow. In addition to ParFlow serving as the subsurface boundary, other groundwater models are also being coupled to LSMs to study similar processes – influence of groundwater on surface fluxes.

   We thank R#2 for the constructive suggestion. In the revision, we have added a synthesis of other coupled groundwater–land surface modeling studies, including relevant work such as Fang and Tai's findings on plant mortality (lines 241–256, 274–276). The most relevant contributions from Ying Fan's studies have already been incorporated in the original manuscript (lines 284–297). However, after an extensive literature review, we found that most existing studies primarily focus on model development (Zeng et al., 2018), calibration (Fang et al., 2019), or sensitivity analysis (Engdahl, 2025), rather than producing process-level scientific findings comparable to those emerging from PF–CLM. Some studies do include process analyses, but these processes are not related to groundwater–land surface interactions.

   For example, two studies using the CATHY/NoahMP coupled system clearly illustrate this point (Niu et al., 2014a; Niu et al., 2014b). One paper concentrates on model development and resolution effects, while the other demonstrates how lateral overland and near-surface flows keep lowland areas wetter by re-infiltrating runoff generated upslope. Although both studies are valuable, neither explicitly investigates groundwater–land surface feedbacks, particularly those involving water-table dynamics.

To further acknowledge these efforts, we have included these models in the revised manuscript by encouraging their participation in our proposed model intercomparison plan (PLCMIP) (lines 661–666), where they may contribute to broader community-based evaluation and potentially generate comparable scientific insights in the future.

3. The section on the preliminary coupling of CoLM with ParFlow is weak in several aspects. 1) It was not clear in this coupling if the authors had implemented a coupler to make future coupling efforts easier. If they only replaced CLM with the updated CoLM, the model development itself is incremental. 2) The test case was also poorly designed. It was not clear why the authors decided to use the model setup in China but the met forcing in Colorado. It was not clear which region did the authors use for model validation. Does that indicate the subsurface setup or the met forcing does not matter that much? 3) A significant amount of results from this section is essentially about the difference between CLM and CoLM, which is not really new results from this study. Instead, papers from the CoLM development should be appropriately cited. 4) since ParFlow-CLM in 2005, ParFlow has also went through several updates as well. It would be helpful to summarize parflow updates and report the influence of model simulations as well.

We thank R#2 for these thoughtful questions and suggestions.

(1) We coupled the new CoLM with ParFlow through the existing interface. Even so, this required substantial work because both the code structure and the data structure in CoLM have undergone major changes. This effort also provides a foundation for guiding a large-scale, structured development plan based on a coupler in the future. More details are provided in our response to the first comment.

(2) Regarding the test cases, the China case and the Colorado case are two completely different experiments. The China case is used for general evaluation, whereas the Colorado case is a single-column setup used specifically for snow processes. They are based on different subsurface configurations and different meteorological forcings. We used the Colorado case for two reasons: first, there is no snow in the China study area; second, in our previous work we found noticeable deviations between simulations and observations of SWE in the Colorado single-column experiment, so we wanted to re-examine it with the newly coupled model. Therefore, the two test cases serve different purposes. We revised it a bit for clarification. Please refer to lines 422–426, 431.

(3) The differences between CLM and CoLM are indeed not new results. Our intention is not to restate these model differences themselves, but to identify and explain the different parameterizations (mechanisms) within the models that lead to different simulation outcomes. These mechanisms are not apparent without the analysis we provide here. We added references in the revised manuscript. Please see lines 483, 487–488, 489–491.

(4) Since 2005, ParFlow has undergone several major developments that significantly affect model performance, such as the integration of overland flow and the implementation of the terrain-following grid. Both test cases in this study already use these modern capabilities. We do not run simulations with the 20-year-old version of ParFlow that had no overland-flow integration and used an orthogonal grid, nor is there a scientific need to revert to such an outdated configuration. Features such as GPU acceleration and reservoir capabilities in ParFlow exist, but they are not relevant to the test cases here. Nevertheless, following the reviewer's suggestion, we have added these contents for clarification. Please refer to lines 500–505.

4. The 4th section on sustainable recoupling framework was too high level and not very specific. I think the community would be more interested in how a sustainable framework has been or can be implemented to ease the process of model upgrades.

   We thank R#2 for the suggestion regarding sustainable recoupling.

   Following the comment, we added several implementation-level details that improve maintainability—such as modularized logging, a modular coupling interface, and modular variable registration (lines 562–569, 576–580, 597–602). These additions help streamline the coupling process. However, we would like to clarify that such technical conveniences alone do not determine long-term sustainability.

   From our perspective, sustainability mainly depends on two structural challenges.

First, the current module-embedded paradigm fundamentally restricts the independent evolution of the coupled components. For systems such as ParFlow and CoLM—each with mature pre-/post-processing ecosystems—embedding one model within another inevitably limits the sustainable development of at least one component. A coupler-based architecture, rather than continued embedded interfaces, is therefore essential.

Second, a sustainable coupling framework requires a clearly defined and well-established grid-mapping and exchange mechanism that reconciles ParFlow's grid with the grid–subgrid structures of land surface models. Establishing this mapping at the outset of the coupling process is essential for ensuring long-term maintainability and interoperability of the coupled system.

Beyond these structural foundations (points 1 and 2 in Section 4), the sustainability of the coupling ecosystem also relies on shared development protocols and consistent community participation (points 3 and 4). Together, these four elements strengthen the maintainability and longevity of the system. However, they do not make an inherently complex multi-model coupling problem "simple." Effective participation still requires developers to possess adequate domain knowledge and technical expertise.

We have clarified these points in the revised manuscript (lines 523–532).

5. regarding the section on the model inter-comparison project, I have some thoughts (not as criticism). I agree the model comparison would be useful exercise. But I think before making model inter-comparison, it would be important to assess model parameter sensitivity. Model development is much more expensive, compared to parameter change. Before we can make any attribution that model development/structure is causing the difference, we need to ensure that the difference cannot be resolved by model parameterization.

   This is a great suggestion, and we have incorporated this point into the revised manuscript (see lines 653–654). Parameter sensitivity analysis is indeed an essential precursor to any structural attribution, and we fully agree that structural differences should only be interpreted after ruling out parameter-related effects. We also clarify that the proposed model intercomparison focuses solely on evaluating existing coupled models rather than developing new ones.

Specific comments.

Figure 1. I did not find this figure from the cited reference paper. Please check.

We double checked the reference paper, and the citation is correct. Our Figure 1 was modified from Figure 4 in

Yang, C., Maxwell, R., McDonnell, J., Yang, X., and Tijerina-Kreuzer, D.: The Role of Topography in Controlling Evapotranspiration Age, Journal of Geophysical Research: Atmospheres, 128, e2023JD039228, https://doi.org/10.1029/2023JD039228, 2023.

Line 377 why evaluating against model reanalyzed data. why not using a EC tower? where is the validation data located?

ERA5-Land was selected because it provides spatially continuous reanalysis fields that fully cover the entire modeling domain, allowing us to evaluate domain-wide variations of sensible heat, latent heat, skin temperature, and transpiration. In contrast, EC tower measurements are point-scale observations with sparse spatial distribution, and therefore cannot serve as a suitable validation dataset for domain-wide averages as shown in Figure 3.

As shown in Figure 3 and clarified in the captions, the validation is performed against spatial averages over the whole modeling area.

Line 399 these are new changes in CLM, and should be properly cited,

Thank you for pointing this out. We have now added the relevant CoLM development references to clarify the source of these updates. Please refer to lines 483, 487–491.

Line 418 main changes in what? CLM or ParFlow.

Revised by adding *in CoLM*. Please see line 506.

References

Engdahl, N. B.: Impacts of Uncertain Permeability Fields on the Transient Hydrologic Response in Coupled Surface-Subsurface Simulations of a Headwaters Catchment, Water Resour Res, 61, e2025WR040668, https://doi.org/10.1029/2025WR040668, 2025.

Fang, K., Ji, X., Shen, C., Ludwig, N., Godfrey, P., Mahjabin, T., and Doughty, C.: Combining a land surface model with groundwater model calibration to assess the impacts of groundwater pumping in a mountainous desert basin, Adv Water Resour, 130, 12-28, https://doi.org/10.1016/j.advwatres.2019.05.008, 2019.
Niu, G.-Y., Paniconi, C., Troch, P. A., Scott, R. L., Durcik, M., Zeng, X., Huxman, T., and Goodrich, D. C.: An integrated modelling framework of catchment-scale ecohydrological processes: 1. Model description and tests over an energy-limited watershed, Ecohydrology, 7, 427-439, https://doi.org/10.1002/eco.1362, 2014a.
Niu, G.-Y., Troch, P. A., Paniconi, C., Scott, R. L., Durcik, M., Zeng, X., Huxman, T., Goodrich, D., and Pelletier, J.: An integrated modelling framework of catchment-scale ecohydrological processes: 2. The role of water subsidy by overland flow on vegetation dynamics in a semi-arid catchment, Ecohydrology, 7, 815-827, https://doi.org/10.1002/eco.1405, 2014b.
Zeng, Y. J., Xie, Z. H., Liu, S., Xie, J. B., Jia, B. H., Qin, P. H., and Gao, J. Q.: Global Land Surface Modeling Including Lateral Groundwater Flow, J Adv Model Earth Sy, 10, 1882-1900, 10.1029/2018ms001304, 2018.