Reviewer #1

This manuscript explores a case study of a new model coupling, the intregrated hydrologic code ParFlow and the land surface code LIS/Noah-MP, to simulate hydrologic processes across the Upper Colorado River Basin (UCRB). The authors compare results between the standalone land surface model and the coupled model to in situ and remote sensing observations of soil moisture, streamflow and groundwater levels. When published, this paper will make a valuable contribution to the literature. At the moment, several aspects of the manuscript require clarification and revision before publication. Notably, the details of the model coupling are not fully described in this paper and are instead referenced in a manuscript currently under review, which limits my ability to evaluate the robustness of the methods and results. In addition, it the introduction and conclusions could be revised to clarify the novelty of the manuscript. Therefore, I am recommending major revisions.

We would like to thank the reviewer for their thoughtful comments and constructive suggestions. These insights have significantly enhanced both the quality and clarity of our work, helping us refine key aspects of the study and improve its overall readability. We appreciate the time and effort the reviewer dedicated to providing detailed feedback, which has been instrumental in strengthening the presentation and impact of our paper.

Major comments:

1) It is difficult to review this paper, as it presents a case study of a new model coupling, but the details of that coupling are described in a paper currently in review. Thus, I cannot evaluate the results presented here as the underling methods are not fully described. I recommend that the authors either (1) wait to publish this manuscript until the paper by Maida et al. (2024) is fully published, (2) publish a preprint of Maida et al. (2024), or (3) describe the ParFlow-LIS/Noah-MP coupling in depth.

This paper by Maina et al. (2025) was recently published in the Journal of Advances in Modeling Earth Systems. Here is the link to the paper: https://doi.org/10.1029/2024MS004415. We have updated the references in the revised manuscript.

References

Maina, F. Z., Rosen, D., Abbaszadeh, P., Yang, C., Kumar, S. V., Rodell, M., & Maxwell, R. (2025). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment. Journal of Advances in Modeling Earth Systems (JAMES), 17(2). https://doi.org/10.1029/2024MS004415

2) I recommend the authors edit the introduction to emphasize the motivation for and novelty of this manuscript. ParFlow has long been coupled to one land surface model within the LIS framework (the Community Land Model), so what additional functionality is provided by coupling ParFlow to LIS? There are certainly advantages provided by the data assimilation and uncertainty estimation tools within LIS, but those tools are not used in this manuscript. Perhaps, then, the novelty of this paper is the difference in process representation between CLM and NoahMP, but the comparison to ParFlow-CLM is not presented in this paper.

ParFlow has not been previously coupled with any land surface model within LIS. This study, following our previously published work, is the first to explore the robustness of coupling ParFlow with Noah-MP within LIS to simulate land surface and subsurface hydrologic processes. We would also like to note that ParFlow is coupled to a different version of CLM than what is in LIS and that this version is incorporated into ParFlow. That is it's not an external, community modeling platform. So in addition to the differences between CLM and NoahMP, there are software differences too. To address this comment, we have revised the introduction section to emphasize the main novelty of this paper and its advantages.

"*The main novelty of this work is to demonstrate the capability of the newly coupled ParFlow and LIS/Noah-MP model in simulating land surface and subsurface hydrologic processes. Although LIS/Noah-MP has been widely used in many studies, its ability to model groundwater processes has been limited. In this study, we assess the performance of the ParFlow groundwater hydrology model when coupled with LIS/Noah-MP, focusing on its ability to simulate subsurface hydrologic processes, such as groundwater and soil water content, and their interactions with land surface processes. It is important to note that the primary goal of this paper is not to compare the performance of the ParFlow-LIS/Noah-MP system to LIS/Noah-MP or any other coupled system. Instead, the focus is on how ParFlow is integrated with LIS/Noah-MP and the resulting improvements, not only in simulating soil moisture (as accurately as LIS/Noah-MP) but also in enabling the simulation of groundwater and other subsurface* *hydrologic processes, such as pressure head—processes that could not be modeled using LIS alone. Unlike LIS/Noah-MP, the ParFlow-LIS/Noah-MP coupling tracks subsurface water movement by solving the three-dimensional Richards equation, providing a more realistic representation of groundwater storage and water table dynamics*."

We have also updated figure 1 to better represent the main novelty of the proposed coupling system.

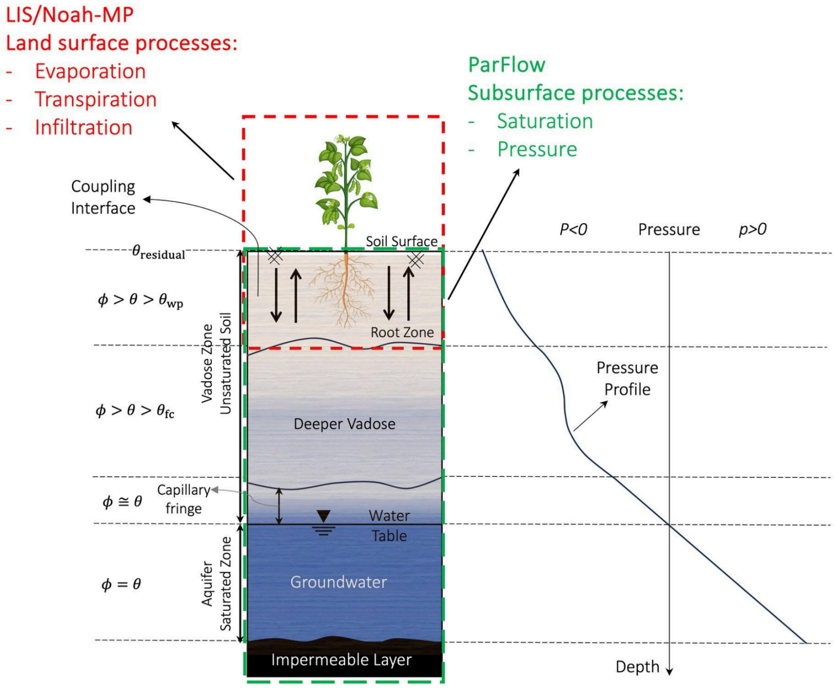

Figure 1. Schematic of the coupled PF-LIS/Noah-MP model. Single soil column representing the coupling zone between the LIS/Noah-MP and ParFlow. θ_{wp} *and* θ_{fc} *are wilting point and field capacity, respectively.*

3) While reviewing the results, I'm not sure I come to the same conclusions as the authors. The abstract (lines 33-34) and conclusions (lines 556-557) find that coupling ParFlow to LIS/Noah-MP improves accuract in regions with complex topography. However, the metrics presented in figures 5, 6, 7, S2 and S3 show that root mean squared error and correlation coefficients are nearly identical between LIS/Noah-MP and ParFlow-LIS/Noah-MP. However, those metrics are averaged over the entire domain, but perhaps there's a difference when those metrics are averaged over areas with complex topography? Please clarify.

Yes, over the regions with complex topography (regions with high elevation gradient – shown in Figure 2), the ParFlow-LIS/Noah-MP resulted in relatively better model performance in terms of soil moisture simulation compared to standalone LIS/Noah-MP model. To further clarify this point and address the reviewer's comment, we have added two additional figures to the revised manuscript, revised Figure 1, and added the following text to the revised manuscript.

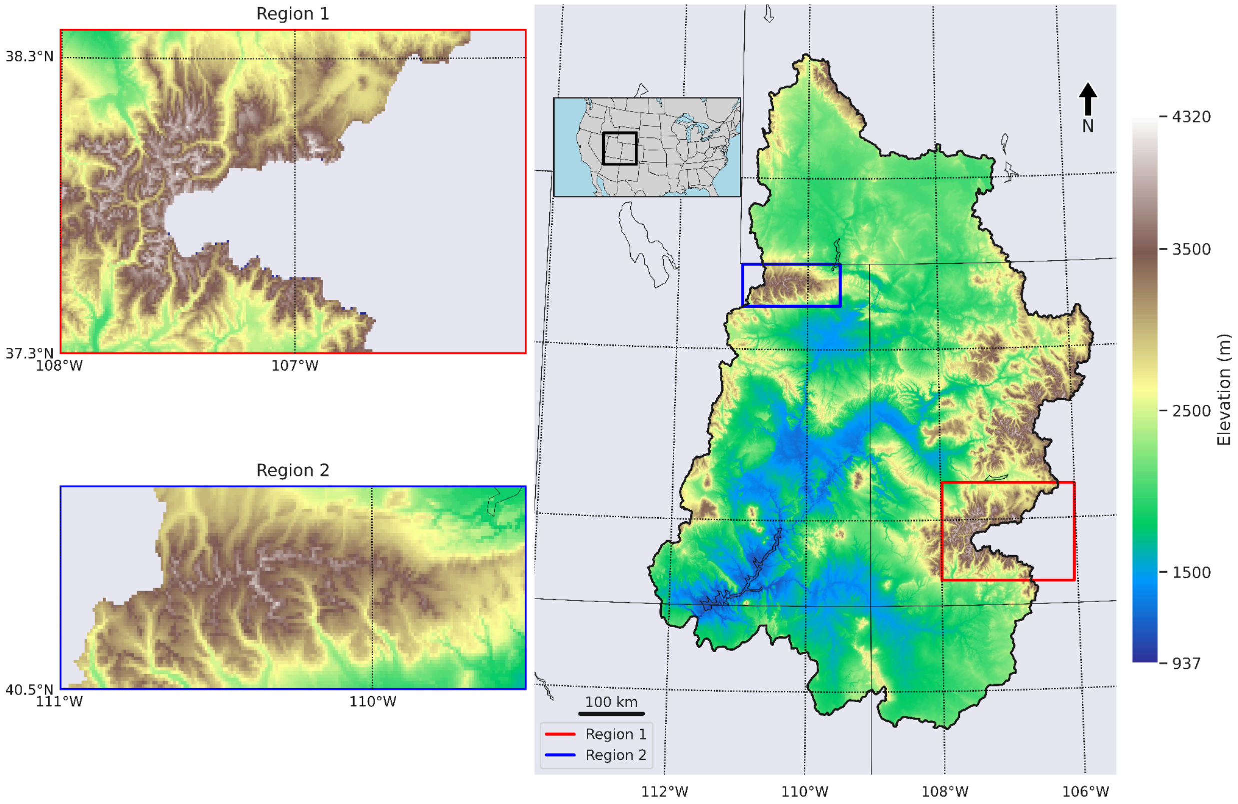

Figure 2. Topography of the Upper Colorado River Basin (UCRB) and its location in the US. Regions 1 and 2 represent the areas with complex topography - regions with high elevation gradient.

"In general, the results indicate that the coupled ParFlow-LIS/Noah-MP model produces soil moisture simulations comparable to those of the LIS/Noah-MP model across the entire UCRB. Figures 5, 8, and 9 (S2 and S3) show that the root mean squared error and correlation coefficient are nearly identical between the two models. For instance, in Figure 5, these metrics are reported as 0.036 m^3/m^3 and 0.608, respectively. However, further analysis—when these metrics are averaged over areas with complex topography—revealed that, in regions with a high elevation gradient (for instance, regions 1 and 2 shown in Figure 2), the ParFlow-LIS/Noah-MP model outperforms the standalone LIS/Noah-MP model in terms of soil moisture simulation. Figures 6 and 7

demonstrate the performance of the LIS/Noah-MP and PF-LIS/Noah-MP models compared to SMAP observations, specifically zooming in on two regions with latitude and longitude ranges: Region 1 (37°N to 38.2°N, -108°W to -106°W) and Region 2 (40.5°N to 41°N, -111°W to -109.5°W). In Region 1, the LIS/Noah-MP model yielded a ubRMSE of 0.0323 m^3/m^3 and a correlation coefficient (R) of 0.308, whereas the ParFlow-LIS/Noah-MP model showed slightly higher values of 0.0358 m^3/m^3 and 0.343, respectively. In Region 2, the LIS/Noah-MP model reported a ubRMSE of 0.0388 m^3/m^3 and R of 0.482, while the ParFlow-LIS/Noah-MP model performed better with a lower ubRMSE of 0.0330 m^3/m^3 and a higher R of 0.539. These regions were selected due to their complex topography characterized by high elevation gradients (see Figure 2).”

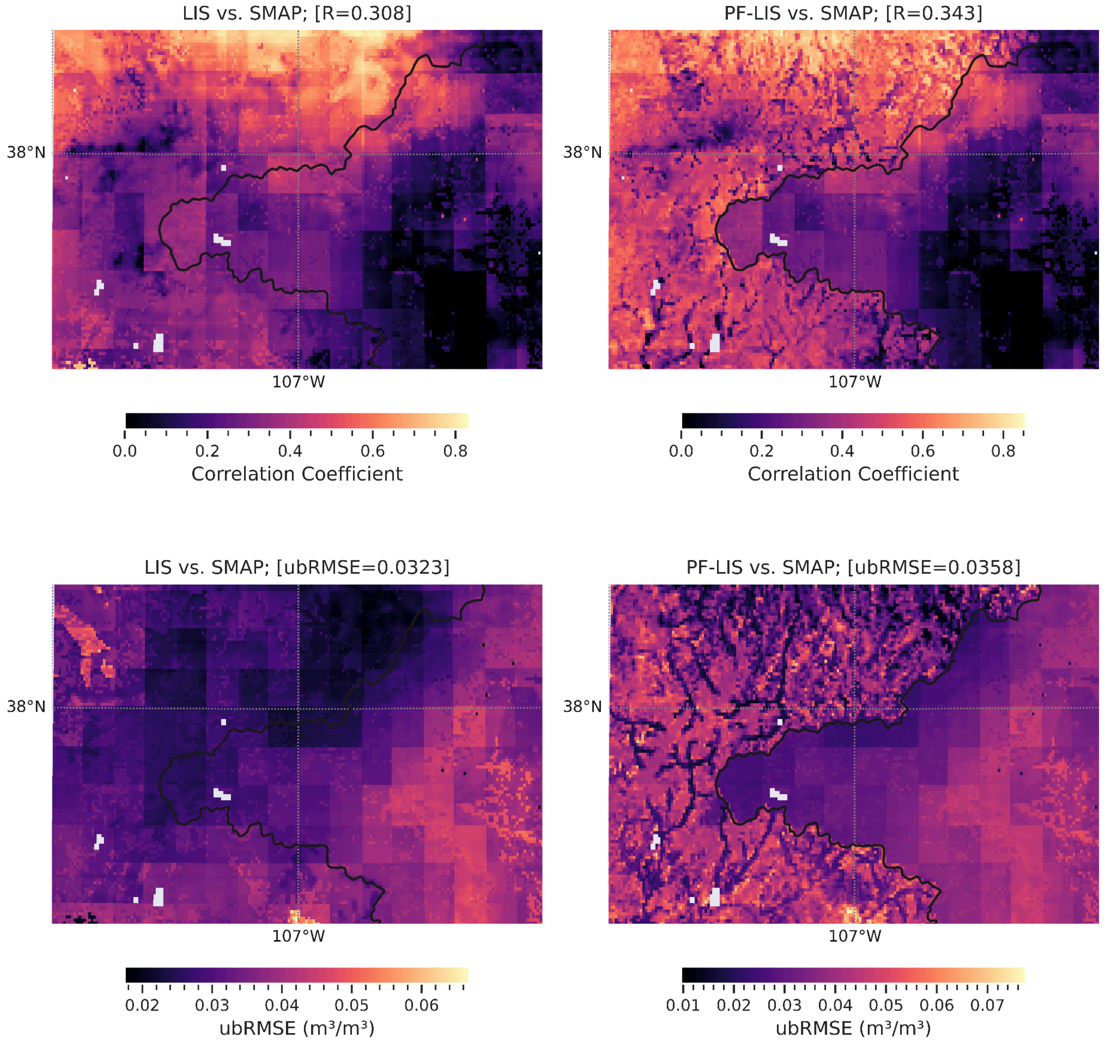

Figure 6. Spatial distribution of soil moisture performance metrics (ubRMSE and R) for Region 1 (shown in Figure 2), comparing LIS/Noah-MP and ParFlow-LIS/Noah-MP models against SMAP observations.

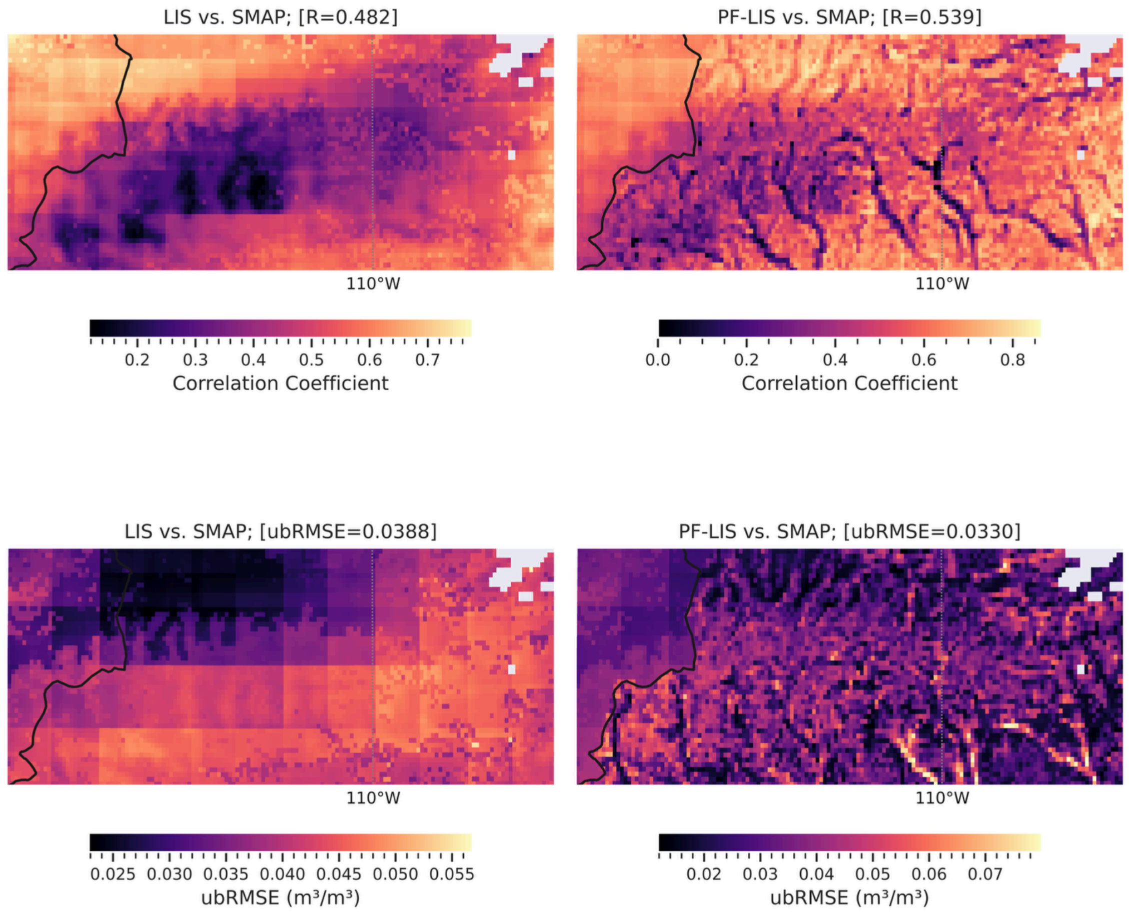

Figure 7. Spatial distribution of soil moisture performance metrics (ubRMSE and R) for Region 2 (shown in Figure 2), comparing LIS/Noah-MP and ParFlow-LIS/Noah-MP models against SMAP observations.

Minor comments:

Lines 61-78: This paragraph focuses on the importance of simulating irrigation, groundwater pumping and other water management infrastructure, but those processes are not included in the simulations in this paper. Thus, it would be helpful to clarify why this paragraph is included here

Toward the end of the manuscript, we emphasize the significance of human impacts, particularly groundwater regulation, in the UCRB. This is connected to the discrepancies observed between the ParFlow-LIS/Noah-MP groundwater simulations and GRACE groundwater observations. Including this information in the introduction helps highlight the role of groundwater pumping and irrigation in the UCRB, providing context for their influence on model validation. To further address this comment, we have added the following text to the revised manuscript.

"While the current study does not directly simulate irrigation or groundwater pumping, these processes are critical in understanding the water dynamics of the region and can influence model outputs. The discrepancy between the ParFlow-LIS/Noah-MP groundwater simulations and GRACE groundwater observations, discussed later in the manuscript, highlights the importance of including such human impacts in hydrologic models. Although addressing these processes is not the primary objective of this study, this work serves as a foundational step toward that goal. As an alternative to direct simulation, data assimilation techniques could be employed in future research to incorporate observed groundwater data or other relevant measurements. This would enable better representation of irrigation and groundwater pumping processes in the model, improving simulation accuracy and addressing the observed mismatch in groundwater observations."

Lines 105-108: The reference to WRF feels a little out-of-place since it isn't used in this study. It would be helpful to briefly introduce Noah-MP here instead, since it is mentioned in the abstract and in the next paragraph

We have removed the reference to WRF in this paragraph and have instead provided a brief introduction to Noah-MP. The following text has been added to the revised manuscript:

"One of the land surface models available within LIS is Noah-MP (Niu et al., 2011), an advanced version of the Noah land surface model. Noah-MP is specifically designed to simulate a range of land surface processes, including soil moisture, temperature, snowpack dynamics, vegetation dynamics, and energy fluxes between the land surface and the atmosphere. It incorporates multiple soil layers, a detailed representation of vegetation types and their properties, and advanced treatments of surface energy exchanges, all of which are important for capturing the complexity of land-atmosphere interactions. In this study, we utilize the LIS framework with the Noah-MP model to simulate these land surface processes, which are critical for accurately representing hydrologic fluxes in the UCRB. Noah-MP's flexibility in representing diverse land surface characteristics allows for a more realistic simulation of hydrological processes such as evapotranspiration, infiltration, and runoff. Its detailed soil-vegetation-atmosphere interactions make it especially useful for understanding water fluxes in regions like the UCRB, where land surface conditions have significant impacts on groundwater recharge and surface water availability."

Line 116: Should Fadji et al. (2024) be Maida et al. (2024)?

Corrected.

“Maina et al., (2024).”

Line 111: It would be helpful to mention some examples of these couplings (ParFlow-CLM, ParFlow-WRF, etc)

To address this comment, we have added the following text to the revised manuscript:

“ParFlow is a robust and versatile groundwater model that integrates advanced numerical techniques to simulate both saturated and unsaturated flow conditions. This model has been coupled with different land surface and atmospheric models, such as the CLM (Community Land Model) and WRF (Weather Research and Forecasting model), to better understand the interactions between the subsurface, surface, and atmospheric processes

Some examples of ParFlow-CLM applications include studies by O'Neill et al. (2021), Tijerina et al. (2021), and Tijerina-Kreuzer et al. (2023), which highlight its use in high-resolution, coupled hydrology–land surface modeling at continental scales. O'Neill et al. (2021) introduced the ParFlow–CLM model (PFCONUSv1) configured over the U.S. to evaluate water balance components, identifying areas for model improvement, such as streamflow biases and shallow water table depth. Tijerina et al. (2021) compared two continental-scale, high-resolution models—ParFlow-CONUS v1.0 and WRF-Hydro—in the first phase of the Continental Hydrologic Intercomparison Project (CHIP), highlighting the importance of model performance evaluation in large-scale hydrologic predictions. Tijerina-Kreuzer et al. (2023) focused on the evaluation of subsurface property configurations for integrated hydrological modeling, emphasizing the significance of accurate datasets for effective model performance and recommending a 1 km resolution subsurface dataset for large-scale hydrologic modeling. All these studies are based on the ParFlow-CLM framework, underscoring its capability in simulating complex hydrological processes at continental scales.

Some examples of ParFlow-WRF applications include studies by Maxwell et al. (2011) and Xu et al. (2022), which highlight its use in coupled atmospheric and hydrologic modeling. Maxwell et al. (2011) introduced the PF-WRF model, coupling the WRF atmospheric model with ParFlow to simulate subsurface flow and overland flow. Their study, applied to the Little Washita watershed, demonstrated improvements *in water resources and wind-energy forecasting, particularly in simulating rainfall, runoff, and the effects of soil moisture on wind power output. Xu et al. (2022) used an integrated process model (IPM) combining WRF with ParFlow-CLM to simulate hydrometeorological conditions in the East River Watershed. Their findings highlighted the significant impact of subgrid-scale physics configurations on simulated hydrological metrics like discharge, snowpack, and evapotranspiration, providing guidance for future modeling in mountainous watersheds. Both studies showcase the versatility of ParFlow-WRF in simulating complex hydrologic processes."*

References:

O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M.: Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States, Geosci. Model Dev., 14, 7223-7254, 10.5194/gmd-14-7223-2021, 2021.

Tijerina, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M. M., Sampson, K., Gochis, D., and Maxwell, R.: Continental Hydrologic Intercomparison Project, Phase 1: A Large-Scale Hydrologic Model Comparison Over the Continental United States, Water Resour Res, 57, e2020WR028931, https://doi.org/10.1029/2020WR028931, 2021.

Tijerina-Kreuzer, D., Swilley, J. S., Tran, H. V., Zhang, J., West, B., Yang, C., Condon, L. E., and Maxwell, R. M.: Continental scale hydrostratigraphy: basin-scale testing of alternative data-driven approaches, Groundwater, n/a, https://doi.org/10.1111/gwat.13357, 2023.

Maxwell, R.M.; Lundquist, J.K.; Mirocha, J.D.; Smith, S.G.; Woodward, C.S.; Tompson, A.F.B. Development of a Coupled Groundwater–Atmosphere Model. Mon. Weather Rev. 2011, 1, 96–116.

Xu, Z., Siirila-Woodburn, E. R., Rhoades, A. M., and Feldman, D.: Sensitivities of subgrid-scale physics schemes, meteorological forcing, and topographic radiation in atmosphere-through-bedrock integrated process models: A case study in the Upper Colorado River Basin, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-437, 2022.

Lines 131-150: Description of ParFlow could be more clear

To address this comment, we have added an appendix to the revised manuscript.

“Appendix

The ParFlow model operates in three distinct modes: (1) variably saturated; (2) steady-state saturated; and (3) integrated watershed flows. This adaptability enhances its utility across a range of hydrological scenarios. Here we summarize each mode following the work of Kollet and Maxwell (2006).

Variably Saturated Flow

ParFlow can operate in variably saturated mode through the well-known mixed form of the Richards’ equation:

$$S_s S_w(p)\frac{\partial p}{\partial t} + \phi\frac{\partial\left(S_w(p)\right)}{\partial t} = \nabla q + q_s \qquad (1)$$

$$q = -\, k_s k_r(p)\nabla(p - z) \qquad (2)$$

where S_s is the specific storage coefficient [L-1], S_w is the relative saturation [−] as a function of pressure head p, t is time, ϕ is the porosity of the medium [−], q is the specific volumetric (Darcy) flux [LT−1], k_s is the saturated hydraulic conductivity tensor [LT−1], k_r is the relative permeability [−], which is a function of the pressure head p, q_s is the general source or sink term [T−1] (includes wells and surface fluxes, e.g., evaporation and transpiration). z represents depth below the surface [L]. ParFlow has been utilized for numerical simulations, including the modeling of river–aquifer exchange involving both free-surface flow and subsurface flow. It has also demonstrated efficacy in addressing highly heterogeneous problems under variably saturated flow conditions. For the situations where the saturated conditions are predominant, the steady-state saturated mode in ParFlow becomes a valuable tool.

Steady-State Saturated Flow

The fully saturated groundwater flow equation is expressed as follows:

$$\nabla q - q = 0 \qquad (3)$$

$$q = -\, k_s \nabla P \qquad (4)$$

where P represents the 3-D hydraulic head-potential [L]. ParFlow does include a direct solution option for the steady-state saturated flow that is distinct from the transient solver. When studying more sophisticated or complex processes, such as when simulating a fully coupled system is of interest (i.e., surface and subsurface flow), an overland flow boundary condition is employed.

Overland Flow

Surface water systems are interlinked with the subsurface system; this interaction plays a critical role for rivers. However, explicitly representing the connections between the two systems in numerical simulations is a difficult task. In ParFlow, overland flow is implemented as a two-dimensional kinematic wave equation approximation of the shallow water equations. The continuity equation for two-dimensional shallow overland flow is expressed as follows:

$$\frac{\partial \psi_s}{\partial t} = \nabla\left(\upsilon \psi_s\right) + q_s \qquad (5)$$

where υ *is the depth-averaged velocity vector [LT−1] and* ψ_s *is the surface ponding depth [L]. Ignoring the dynamic and diffusion terms results in the momentum equation, which is known as the kinematic wave approximation:*

$$S_{f,i} = S_{o,i} \qquad (6)$$

The $S_{f,i}$ *and* $S_{o,i}$ *represent the friction [−] and bed slopes (gravity forcing term) [−], respectively.* i *indicates the* x *and* y *directions in the following equations. Therefore, Manning's equation can be used to build a flow depth-discharge relationship as follows:*

$$\upsilon_x = \frac{\sqrt{S_{f,x}}}{n} \psi_s^{2/3} \qquad (7)$$

$$\upsilon_y = \frac{\sqrt{S_{f,y}}}{n} \psi_s^{2/3} \qquad (8)$$

where n *is the Manning roughness coefficient [TL−1/3]. The shallow overland flow formulation (Eq. 9) assumes the vertical averaging of flow depth and disregards any vertical change in momentum within the surface water column. To incorporate vertical flow (from the surface to the subsurface or vice versa), a formulation that couples the system of equations through a boundary condition at the land surface becomes essential. We can modify Equation (5) to include an exchange rate with the subsurface,* q_e*:*

$$\frac{\partial \psi_s}{\partial t} = \nabla\left(\upsilon \psi_s\right) + q_s + q_e \qquad (9)$$

In ParFlow, the overland flow equations are directly coupled to the Richards' equation at the top boundary cell under saturated conditions. Conditions of pressure continuity (i.e., equal pressures at the ground surface for the subsurface and surface domains) and flux at the top cell of the boundary between the subsurface and surface systems are assigned. Setting pressure head in Equation (1) equal to the vertically averaged surface pressure, ψ_s*:*

$$p = \psi_s = \psi \qquad (10)$$

and the flux, q_e, equal to the specified boundary conditions (for example, Neumann-type boundary conditions):

$$q_{BC} =- k_s k_r \nabla(\psi - z) \qquad (11)$$

and one solves for the flux term in Equation (10), the result becomes:

$$q_e = \frac{\partial\|\psi, 0\|}{\partial t} - \nabla\upsilon\|\psi, 0\| - q_s \qquad (12)$$

where the $\|\psi, 0\|$ operator is defined as the greater of the quantities, ψ, 0. Putting the equations (10) and (11) together results in the following relationship:

$$- k_s k_r \nabla(\psi - z) = \frac{\partial\|\psi, 0\|}{\partial t} - \nabla\upsilon\|\psi, 0\| - q_s \qquad (13)$$

As we see here the surface water equations are represented as a boundary condition to the Richards' equation. For more information about the coupled surface and subsurface flow systems in ParFlow, we refer the interested readers to Kollet and Maxwell (2006)."

References:

Kollet, S. J. and Maxwell, R. M.: Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, Adv Water Resour, 29, https://doi.org/10.1016/j.advwatres.2005.08.006, 2006.

Lines 148-149: "groundwater may take a longer time (for example compared to soil moisture) to reach a steady-state due to such a complicated subsurface configuration" Does "long time" here refer to simulation time or computational time? I'm unclear if this sentence is meant to describe the long time it takes to spin-up water content in the deep vadose zone due to slow rates of groundwater recharge, or if it refers to long computational time due to the difficulty of solving the Richards equation across a thicker vadose zone.

We thank the reviewer for this comment. What we mean is really the former of the two; that groundwater and the deeper vadose zone takes a longer time to reach steady state than the shallower subsurface stores. This longer simulation time can result in longer simulation times, but not due to difficulty in solving Richards' equation, just because the equilibrium times for the deep vadose zone are so long. We have revised this sentence to read:

"It is important to note that groundwater and the deeper vadose zone may take long simulation times (for example compared to shallow soil moisture) to reach a

steady-state due to slow rates of groundwater recharge and subsurface heterogeneity, which can make it a computationally expensive problem to solve (Maxwell et al., 2014).”

References:

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J. O., Ferguson, I. M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S. J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y. J., Phanikumar, M. S., Shen, C., Sudicky, E. A., and Sulis, M.: Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour Res, 50, https://doi.org/10.1002/2013WR013725, 2014

Line 155: Which variables are included in the initial conditions? Soil moisture, surface temperature, what else?

It includes the total volumetric soil moisture and liquid water volume, soil temperature, canopy intercepted water (ice and liquid), canopy temperature, ground surface temperature, snow water equivalent and snow depth. For more information, please see section 3.1.1.3.2 in https://land-da-workflow.readthedocs.io/en/release-public-v1.2.0/CustomizingTheWorkflow/Model.html.

We have also revised the text to further address this comment.

“Land surface modeling within LIS relies on three key inputs: (1) initial conditions, describing the land surface's starting state (i.e., total volumetric soil moisture and liquid water volume, soil temperature, canopy intercepted water (ice and liquid), canopy temperature, ground surface temperature, snow water equivalent and snow depth)… ”

Lines 168-171: I recommend mentioning that the LIS data assimilation framework is not used in this study

We have included this in the revised manuscript. Thank you.

“Please note that the LIS data assimilation framework is not used in this study.”

Lines 181-182: Is the coupling specific to NoahMP? Or could someone use this same code to use VIC or HySSIB instead of NoahMP?

The coupling is specific to Noah-MP.

Section 4: The description of this coupling could be more detailed. How is transpiration from the root zone handled? Is transpiration only from the top soil layer or does NoahMP draw water from deeper layers as well? Also, how is overland flow handled? In this description (and the image in Fig 1), it appears as though ParFlow only simulates subsurface flow.

To address this comment, we added the following text to the revised manuscript.

"The LIS/Noah-MP model is designed to simulate the energy and water fluxes at the land surface, along with key state variables like ET and its components, snow-related variables (such as SWE and snow cover), and infiltration. It computes the surface energy balance by representing vegetation with a detailed canopy model, incorporating its dimensions, orientation, density, and radiometric properties. A two-stream radiation transfer scheme is employed to account for the complex interactions of solar radiation within the canopy. For snow processes, the model features a multi-layer snowpack, capable of storing liquid water and simulating melt and refreeze processes. It also includes a snow interception component, which models the loading and unloading of snow, sublimation, and other snow-related processes. The ET and infiltration values (which combine snowmelt and rainfall) produced by LIS/Noah-MP are passed on to ParFlow. ParFlow then calculates the surface, soil, and subsurface hydrodynamics, generating important hydrological outputs such as water table depth, groundwater storage (derived from pressure-head and saturation), soil moisture, and streamflow (Maina et al., 2025). In particular, transpiration is computed by LIS/Noah-MP using the soil moisture computed by ParFlow. Within LIS/Noah-MP, transpiration is computed using a Penman-Monteith based approach, where stomatal resistance (influenced by solar radiation, vapor pressure deficit, temperature, and soil moisture) controls canopy conductance. Actual transpiration is obtained by scaling potential transpiration with a soil moisture stress function, considering vegetation type, root distribution, and dynamic LAI."

References

Maina, F. Z., Rosen, D., Abbaszadeh, P., Yang, C., Kumar, S. V., Rodell, M., & Maxwell, R. (2025). Integrating the interconnections between groundwater and land surface processes through the coupled NASA Land Information System and ParFlow environment. Journal of Advances in Modeling Earth Systems (JAMES), 17(2). https://doi.org/10.1029/2024MS004415

Line 254: Should be "USGS stream stations"

Corrected.

Line 255: I'm surprised at how few monitoring wells there are. Have the authors considered adding water level measurments from either the Colorado Water Conservation Board or the Utah ? From a cursory glance (https://dwr.state.co.us/Tools/GroundWater/WaterLevels), it seems like there are many water level measurements not included in the USGS database. Also, what are the screened intervals for each well? If a well screen extends across multiple model cells, how are modeled and observed values compared?

Thank you for your suggestion. We have utilized the 18 USGS stations available within the study region. We also studied the observational datasets from the Colorado Water Conservation Board, as you recommended. However, most of these datasets are not recorded at a daily time scale, and some fall outside the period of our study, which limits their use for model simulation validation. Additionally, we studied the distribution of the wells and found that they are spread across the model grid cells in such a way that each grid cell has only one USGS station available for use. We also reviewed the USGS documentation for each station but did not find any information about the screened intervals. The only related information available was the well depth, which we have added to Table 1 in the revised manuscript. To determine the screened intervals, we need well logs or well-completion reports. Since these are not available, we can estimate the screened intervals (using rule-of-thumb method), such as screening 20% of the well depth.

USGS Station ID	*Well Depth (m)*
362936109564101	*259.9*
363850110100801	*407.4*
364255108053202	*18.6*
364338110154601	*264.4*
382427107491401	*4.9*
382656107500701	*7.5*
382917107483101	*4.5*
382947107465801	*5.9*
383051107525501	*5.5*
383315107525201	*10.4*
383626107581501	*6.1*

USGS Station ID	*Well Depth (m)*
384110107591801	*4.4*
384240108000701	*6.9*
395136108210000	*195*
395136108210001	*265.4*
395136108210004	*75.9*
395755108211400	*384*
395755108211401	*534.8*

Fig. 3: I recommend clarifying that WTD corresponds to "water table depth". Also, do all of these monitoring wells truly represent the depth to the water table? Or do they represent groundwater head? It's unclear whether these wells are screened across the water table.

Yes, the monitoring wells represent the depth to the water level according to the USGS webpage. We included the following text in the revised manuscript to further clarify this.

"In this study, the monitoring wells are used to measure the depth to the water table (WTD), not groundwater head."

Line 283: The manuscript could be improved by expanding this section and including additional details on model set up, such as boundary conditions and the extent and discretization of the domain. What are the lateral boundary conditions for the PF-LIS model? I assume that cells outside the UCRB would be inactive, but results for those cells are shown in Fig 4, 5, etc. Similarly, what is the extent of the LIS/Noah-MP domain? What are the lateral boundary conditions?

ParFlow is run over the UCRB, with areas outside the defined region masked out and are inactive in the simulation. We showed the model result only using LIS-Noah-MP on both maps to highlight the difference between using the coupled ParFlow/LIS-Noah-MP and standalone LIS-Noah-MP model. For example, to what extent the coupled system is able to provide more detailed predictions of land surface process across different regions with different land surface characteristics. The boundary conditions for ParFlow are set as no-flow (Neumann conditions) along the lateral edges of the region, reflecting the natural limits where lateral flow into the model domain is negligible. Similarly, the bottom layer is also assigned a no-flow condition, as the model extends deep enough to reach a zone where vertical flow is minimal. At the top of the domain, overland flow conditions are applied, corresponding to the land surface.

We have added the following text to the revised manuscript to further address this comment.

"The total extent of the UCRB model is 608 km in the east–west (x) direction and 896 km in the south–north (y) direction, with a horizontal resolution of 1 km. The model depth is 392 m and consists of 10 layers with variable thicknesses of 200, 100, 50, 25, 10, 5, 1, 0.6, 0.3, and 0.1 m from bottom to top. ParFlow is run over the UCRB, with areas outside the defined region masked out. We present model results using only LIS-Noah-MP on both maps to highlight the difference between the coupled ParFlow/LIS-Noah-MP system and the standalone LIS-Noah-MP model. This comparison demonstrates the extent to which the coupled system provides more detailed predictions of land surface processes across regions with varying land surface characteristics. The boundary conditions for ParFlow are set to no-flow (Neumann conditions) along the lateral edges of the region, reflecting the natural limits where lateral flow into the model domain is negligible. Similarly, the bottom layer is assigned a no-flow condition, as the model extends deep enough to reach a zone where vertical flow is minimal. At the top of the domain, overland flow conditions are applied, corresponding to the land surface Maina et al., (2025)"

Line 293: 1 km lateral resolution?

Yes. Revised.

Line 296: Are these depths or thicknesses?

It refers to the thickness. We have included this information in our response to the above comment.

Line 317: Were these three 20-year periods run sequentially? Also, how was 60 years determined to be an adequate spin-up period? Are there metrics to determine whether the system is at dynamic steady state?

Yes, the model was run sequentially, and 60 years of simulation were sufficient to ensure that the system reached quasi-equilibrium.

Lines 320-321: How different were the initial conditions across the shared portions of the PF-LIS/NoahMP domain? How were differences in the two soil moisture fields reconciled before starting the first coupled simulation?

Prior to running the coupled ParFlow-LIS/Noah-MP spin-up simulation, both models—ParFlow and LIS/Noah-MP—were spun up individually. When we compared the soil moisture simulations from both models, the results were very similar.

Line 323: What metrics were used to determine that the system was at quasi-equilibrium?

We considered the system to have reached quasi-equilibrium when the total storage change was less than 1% of the potential recharge.

Line 329: What size time step was used for input forcing and the output analysis for these simulations? Hourly meteorological forcing? Daily pressure/saturation output?

We used hourly meteorological forcing to run the model and employed daily output for analysis.

Line 349: What is the difference in input forcing that provides this finer spatial resolution in PF-LIS/Noah-MP than in LIS/Noah-MP alone? Weren't both codes run using the same lateral resolution?

PF-LIS/Noah-MP and LIS/Noah-MP use a form of Richards' equation with some different assumptions. LIS/Noah-MP uses a different function for retention (not the van Genuchten function used within ParFlow) and it is 1D (one-dimensional). The main difference between PF-LIS/Noah-MP and LIS/Noah-MP is the deeper subsurface in PF-LIS/Noah-MP and the fact that it accounts for lateral flow, resulting in a more physically realistic representation of water movement through the soil. This enables the PF-LIS/Noah-MP model to capture the complex influence of topography and specific land surface features on soil moisture.

Fig. 4: What are the values outside of the UCRB watershed boundary and why does the resolution appear to be lower beyond that boundary in PF-LIS? Are those cells identical between the two simulations?

Yes, the cells are identical between the two simulations. The outer boundary of the UCRB was simulated using LIS/Noah-MP only. The right panel shows the simulated values for the interior of the UCRB when ParFlow is activated, and the coupled system is used for soil moisture simulation. This figure highlights the contribution of the coupled system in providing more detailed information about soil moisture simulation and its relationship to the region's topographic characteristics.

Line 360: How does the vertical resolution of the simulations compare to the SMAP penetration depth?

The soil moisture simulation at the topsoil layer (10 cm depth) from both the ParFlow-LIS/Noah-MP and LIS/Noah-MP models was compared with SMAP soil moisture data.

Fig. 5: In the caption, it could be useful to clarify the depth interval for the simulated soil moisture values.

We have added this to the revised manuscript.

"The comparison of soil moisture was made using data from the 10 cm soil depth."

Line 384: Could the difference in overland flow between PF-LIS/Noah-MP and LIS/Noah-MP also contribute to the increased spatial heterogeneity observed in PF-LIS/Noah-MP simulations?

Yes, this is correct. Figure S4 in the supplementary information also shows the surface runoff (along with its spatial heterogeneity) simulated by the coupled system, PF-LIS/Noah-MP.

Lines 399-400: It's not immediately clear from the figures that PF-LIS/Noah-MP improves the accuracy of soil moisture in high altitude regions. Is there an alternate figure that more clearly shows this result?

Thank you for the suggestion. We have addressed this point earlier in our response to comment #3 under "Major Comments".

Line 455: Why do you think PF-LIS/Noah-MP is unable to capture the timing of runoff? Is this due to errors in hydraulic conductivity, which cause inaccurate estimates of the timing of the rainfall-runoff response?

Over some USGS stations, PF-LIS/Noah-MP has shown marginal efficiency in capturing the timing of runoff, and this is likely not solely due to errors in hydraulic conductivity. As discussed in Maxwell and Condon (2016), the algorithm used for topographic processing resulted in spatial inconsistencies between the modeled and actual stream networks. To address this, USGS gauges were mapped to the PF-LIS/Noah-MP grid using nearest-neighbor mapping and manual adjustments, ensuring the gauges were correctly placed on the appropriate ParFlow stream cells. These inconsistencies in stream network representation may contribute to inaccuracies in runoff timing, in addition to any potential errors in hydraulic conductivity.

To further address this comment, we have added the following text to the revised manuscript, where we discuss the Condon diagram.

"*Over some USGS stations, PF-LIS/Noah-MP has shown marginal efficiency in capturing the timing of runoff, and this is likely not solely due to errors in hydraulic conductivity. As discussed in Maxwell and Condon (2016), the algorithm used for topographic processing resulted in spatial inconsistencies between the modeled and actual stream networks. To address this, USGS gauges were mapped to the PF-LIS/Noah-MP grid using nearest-neighbor mapping and manual adjustments, ensuring the gauges were correctly placed on the appropriate ParFlow stream cells. These inconsistencies in stream network representation may contribute to inaccuracies in runoff timing, in addition to any potential errors in hydraulic conductivity*.

Reference

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377–380, https://doi.org/10.1126/science.aaf7891, 2016."

Line 471: A minor point, but it could be useful to add to this diagram the number of points that are in each quadrant/category.

Thanks for the suggestion. We added the following text to the revised figure caption.

"This diagram includes 177 purple points, 197 green points, 4 red points, and no blue points."

Line 492: How were groundwater heads compared between simulated and observed values? Given that some cells are up to 200 m thick and PF-LIS/Noah-MP reports a single pressure value per cell, do these calculations assume hydrostatic equilibrium within a given cell to calculate the exact water table depth within that cell? Similarly, for wells that have a long screen length and are screened entirely below the water table, the reported water level measurements integrate pressure across the length of the screen.

The reviewer is correct that we assume hydrostatic equilibrium within a cell to interpolate the exact water table depth. We generally do try to compare predicted and observed heads in a way that honors the well construction, integrating over the screen if confined. However, it’s often difficult to determine this from the observation database and errors can occur. We have added some clarifying text to the sentence a few above this section:

“It is important to note that all wells were assigned to the nearest grid cell center without any additional adjustments, that water table depths are interpolated within grid cells assuming a hydrostatic equilibrium and that information regarding screen depth and well construction are used in the comparison when available.”

Lines 495-497: "Stations located in topographically complex surroundings tend to yield lower model performance compared to those in areas with smoother and flatter environments." Would it be possible to include a figure in the supplement to support this statement? It might be more clear to show this relationship in a map rather than in a table of latitude and longitude values.

We have added the following figure to the supplementary file as you suggested. Thank you.

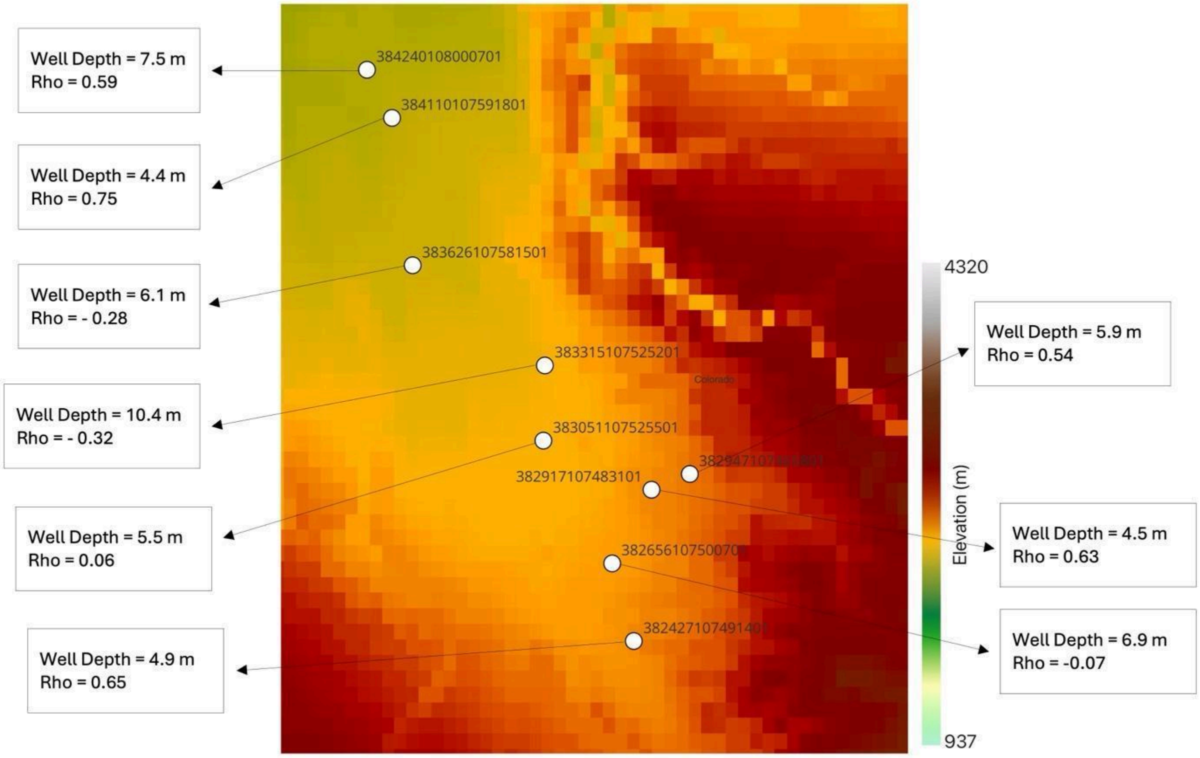

Figure S6. Spatial distribution of the estimated WTD and its comparison with well observations across various locations featuring different land characteristics.

Lines 528-530: This is an interesting result! Does this discrepancy also suggest that PF-LIS/Noah-MP underestimates evapotranspiration because croplands in the simulations do not receive any groundwater-fed irrigation? Another option for future work would be to compare remote-sensing-based estimates of ET with estimates from both PF-LIS/Noah-MP and LIS/Noah-MP.

We did not focus on evapotranspiration in this study, so we cannot be certain that PF-LIS/Noah-MP underestimates evapotranspiration because simulated croplands don’t include groundwater-fed irrigation. In our first paper (Maina et al., 2025), we compared evapotranspiration simulations from PF-LIS/Noah-MP and LIS/Noah-MP in irrigated areas. That paper gives more details about how the coupled system handles evapotranspiration, and we recommend the reviewer refer to it for more information.

Thank you for suggesting ideas for future work. There are many observational datasets and methods available to validate the PF-LIS/Noah-MP model outputs. In this paper, we focused on a subset of them, but future studies will expand the validation using additional observational data and covering other regions.