Reviewer #2:

Reviewer summary:

The study of Lapides et al. investigates the inclusion of rock moisture into a global dynamic vegetation model. To test the approach the team tests modifications to the storage compartments of the model and compare the model results to two study sites and available data for the continental US.

Overall the authors present a very strong research paper, which unfortunately currently lacks clarity and a discussion of uncertainty. Due to imprecise language and a lot of figures it is sometimes hard to follow the authors in their conclusions. Moving 1-3 figures to a supplement or appendix would help to tell a clearer story.

Thank you for this feedback. We will respond (in blue) below to these comments (in black) as the reviewer expands upon them.

A lot of figures spaced out over short and precise explanations. Moving ~3 Figures to the supplement would help clarity a lot. Currently very hard to read. Figure 12 seems not to be referenced and explained anywhere.

We received similar feedback from Reviewer #1 in regard to the number of figures. Following Reviewer #1's suggestions, we have moved two figures from the main text to the supplement and simplified the remaining figures (see response to Reviewer #1). As for Figure 12, we included the wrong reference in Section 3.4. We have updated the previously incorrect citation to Figure 7 to Figure 12.

I would also appreciate if the authors could comment on how the data uncertainty influences the results. As the authors note data in the US is relatively good compared to global soil and bedrock datasets. It would greatly improve the study if the authors could provide an insight of how their results are impacted by data uncertainty in the presented study regions but also what this possibly entails for the global scale

Thank you for this suggestion. We added a paragraph in the discussion section addressing uncertainty and extending datasets globally:

"The data sources used in this study for soil capacity and to calculate weathered bedrock storage capacity are specific to the United States. To extend this model beyond the United States, it would be most important to extend estimates of total plant-accessible storage since soils datasets are generally inadequate to represent plant-accessible water stores even where they are available (McCormick et al., 2021), due to widespread plant water uptake from layers deeper than those in traditionally mapped soils databases. While the specific distributed water flux datasets used in this study are not globally available, Wang-Erlandsson et al. (2016) used a similar deficit-based strategy to the one used in this study to estimate plant-accessible water storage globally with alternative water flux datasets. The accuracy of estimates of this type is limited by both (i) the accuracy of the input water flux data and (ii) the time period of data availability. In the case of the present study, the PML-V2 evapotranspiration and PRISM precipitation datasets used to calculate the root-zone storage deficit close mass balance well with USGS streamflow gages (Rempe et al., 202x) in undisturbed watersheds in the western United States. However, data concordance should be confirmed with any data sources to be combined for use in a root-zone storage deficit calculation since mass balance errors can compound over time. Second, the time period of data availability is important since the maximum root-zone storage deficit provides only a minimum bound on plant-accessible water storage. With a longer timeseries, the minimum bound is more likely to approach the actual plant-accessible water storage, particularly if dry periods or disturbances like fire or logging are included in the timeseries. Shorter timeseries or timeseries that fall during a particularly wet period of history may be more likely to underestimate plant-available water storage. Based on the findings of this study, underestimating storage capacity would result in lower evapotranspiration and less tree growth in LPJ-GUESS, and overestimating storage capacity would result in the opposite. "

Linked to the comment on uncertainty the authors should provide a clearer link to how the two sites are representative for the US and globally. What other regions should future research investigate to figure out on how to represent rock water on a global scale in these models?

We have added a paragraph to the discussion in Section 4.2 on this topic:

"Elder Creek (480 mm storage capacity) and Dry Creek (180 mm storage capacity) have storage capacities at the 79th and 4th percentiles of storage capacities in the Mediterranean region included in this study (Cs label in Figure 8e). As such, they capture two broad sets of behavior found in Mediterranean sites that are also common beyond Mediterranean regions, but they do not fall at the mode of the distribution of storage capacities, which is 330 mm. Thus, it would be valuable to continue with more site-specific studies to identify whether additional complexity or alteration to the model structure would be valuable. In particular, it would be valuable to explore rock moisture dynamics in DGVMs in snow-dominated sites, which was not explored in detail in this study."

The authors now have modified mainly the storage component of the model. Would it be beneficial to also provide a lateral groundwater connection inside the model?

Absolutely. Fan et al. (2019) advocate for landscape position as an important driver of rooting depths and therefore vegetation dynamics. However, incorporating lateral groundwater flows into LPJ-GUESS would require fundamental changes to the model structure that are far beyond the scope of this study. Given the extensive changes this would require, it may make sense to explore this question by coupling a DGVM with a more complex hydrological model (like the coupling of CLM and LPJ for CLM-DGVM). However, this is still a good point, and we added a note to the discussion to address this:

“Other aspects of hydrology may also be essential to account for in certain regions or landscape positions, such as lateral groundwater flows (Fan et al., 2019). However, most DGVMs are not structured to account for topography, making the inclusion of both subsurface and surface water flow subsidies highly challenging. Future efforts could explore more complex hydrology by restructuring a DGVM like LPJ-GUESS to take into account topography or coupling the plant dynamics in a DGVM such as LPJ to an existing hydrological model.”

Fan, Ying, et al. "Hillslope hydrology in global change research and earth system modeling." *Water Resources Research* 55.2 (2019): 1737-1772.

Additional detailed comments:

P2 11: You motivate your paper with the Mediterranean and then evaluate it for the US, why?

We think that the reviewer’s question may relate to the two possible definitions of “Mediterranean regions,” which we did not adequately define in the original manuscript. First off, we define Mediterranean areas as those with a Mediterranean climate. We adjust the language throughout to “regions with Mediterranean climate” and include a brief description of a the Mediterranean climate: “which experience hot dry summers and cooler, wetter winters.” We also added a sentence clarifying the importance of carryover moisture in mediterranean regions on page 3, line 2:

“Generally, it is the case that water stored from the wet winter is essential for supporting plant function during the dry summer in regions with Mediterranean climates.”

To further address this comment: we motivate the study with areas with Mediterranean climate, which provided a clue that something may be missing from the models. We then want to evaluate it at a large scale, not just in areas with Mediterranean climate, to demonstrate that the model change is generally applicable, not something that applies specifically to areas with Mediterranean climate. Areas with Mediterranean climate are the most affected, but models can be generally improved everywhere if the subsurface hydrology and structure is properly represented. We added notes to that effect at revised page 2 lines 11 and page 4 line 14:

“This performance gap provides a clue that there may be an essential component missing from these models. “

“We test this hypothesis in detail at two intensively monitored sites in Northern California with similar climate but distinct vegetation communities (Hahm et al., 2019) and more broadly at 4 km resolution across CONUS to demonstrate that these changes result in realistic predictions not just in Mediterranean areas but across all biomes represented in CONUS.”

P3 1: Again unclear.

It wasn’t clear to us what the confusion was. Could you be more specific so we can improve the text?

P4 10: You lost me. How do these places relate to the MED issues you highlighted? How much will they be transferable?

We understand the confusion. We think that we may have not explained clearly enough what the change is. We are not just adding bedrock water underneath the existing LPJ soil storage but altering the total available storage in the subsurface layers, as described in the methods. While this is clear in the methods, it is not clear in the introduction. To help with this issue, we added a statement at lines XX:

“We expect to see the largest impacts in Mediterranean areas, but these improvements should show up more modestly in other areas as well since this change will result in a more realistic depiction of subsurface water availability everywhere.”

Table 2: Nice but very difficult to read. Could you move the justification just to text and pivot the table?

Great suggestion. We pivoted the table, and left the justification to the text.

P8 1-4: This is unnecessary. Cite one key paper and be done with it. This seems more like self-advertisement than actual scientific proof

We shortened the list of references.

P8 6: More recent versions than what? The one used in this paper? If so this doesn't matter then. Or if it matters explain why.

Yes, we are referring to versions more recent than included in this paper. We have removed the comment.

P11 23: In the model or in the real-world site?

Both! We clarified with the following comment:

"For the case study locations both in reality and in the model…"

P11 27: For what time frame?

This is clearer now that we have added more details about the model run in response to Reviewer #1, but we also added a comment at Line XX for further clarity:

"for the full study period (1981-2021)"

P11 28: First reference to PML-V2 is that a dataset or a model? Why do you use it as benchmark?

We added a clarification on what PML-V2 is ("distributed ET data product PML-V2"). PML-V2 is first mentioned in the Section 2.3 (prior to this), in which we note that "PRISM precipitation and PML-V2 have been found to perform well for mass balance closure compared with USGS streamflow gages (Rempe et al., 20xx)."

29: Why did you not use a hydrological model which might compute runoff instead?

Using a hydrological model would introduce additional uncertainty and complexity into the study. For simplicity and for consistency with the data used for evaluation in this study, the mass balance approach provides a simple estimate of runoff that incorporates no lateral water flows (similar to LPJ-GUESS), making it a good comparison.

P12 8: Unclear and confusing sentence. The mass balance of what? What spatial distribution metric and what is it used for?

We updated that sentence for clarity. It now reads: “We used annual runoff from mass balance between PRISM P and PMl-V2 ET for comparison with LPJ-GUESS runoff and annual and summer ET from PML-V2 for comparison with LPJ-GUESS ET to evaluate model performance based on Kling-Gupta Efficiency”

10: This is a result and belongs into section 3.

We moved this information to Section 3.3 in the results: “In terms of overall model performance, summer ET improvements with the fully modified model drive strong model improvements, although annual runoff and overall ET performance is slightly decreased (Supplemental Information S1).”

Fig 3: Missing section reference. This does not need to be an extra figure, e.g., add as small legend to Fig. 5.

We tried to turn Figure 3 into an inset. However, the descriptive text, which we think is very useful, becomes hard to read, so we left Figure 3 as a separate figure to improve readability of the paper.

Fig 5: Please explain first what a-d show. It is unclear whether a and c both show results for the whole CONUS. It is currently easy to get lost in the information.

We removed extra information from the caption and clarified that panels a, c, and d show results across CONUS, while b shows only pixels across CONUS where storage increases by at least 200 mm.

P15 18: I assume T stands for transpiration? Make it explicit also in Figure 6

We clarified that T stands for transpiration at both of these locations.

Fig. 6: The effect seems strongest if storage capacity is large but what is the explanation for transpiration underestimation in small capacity sites?

We agree that there is still a lingering question here. This is addressed in the discussion in Section 4.2.

PML should be a different color since you already use black to indicate change. Also, the colored arrows make the plot hard to read and may be confused with datapoints. For all figs: the annotation is helpful but it should be very clear that it is an annotation. Should also be added to legend in all plots.

We changed the color for PML to grey and added legends instead of the labeled lines so that all annotations are clear.

Fig 7 and 8: Maybe these two could be combined? Because they do not show that much new content and you already have a lot of complex figures and in the text and you jump between these two a lot. Maybe some of this could be moved to the supplement?

At the suggestion of Reviewer #1, we moved Figure 8 to the supplement. We agree that there is a lot of similar material.

P16 6: Indeed, but where are the reductions the lowest and why?

This is a good question. Comparing Figure 9c to 9d, we see that error is reduced essentially everywhere, and the error becomes centered around 0 rather than mostly negative. Where error reductions are small, error was small to begin with. As demonstrated in Figure 9a, sites with very little difference between the modified and default storage capacity or with a smaller storage capacity in the modified model tend to see little change in the resulting summer ET signal. We added a comment about this to Section 3.2: "The change in storage capacity between the modified and default models does an excellent job determining how large of a change there will be to summer ET (Figure 9a), so that places with little change in summer ET are those where the storage change was negative or very small."

P22 4: Is this supposed to be Fig 12?

Yes, thank you. We have corrected that.