In the following responses, reviewers' comments are reproduced in their entirety in black, and the authors' responses are noted in blue.

Reviewer 3

Title: Substantial root-zone water storage capacity observed by GRACE and GRACE/FO

Author(s): Meng Zhao et al.

MS No.: egusphere-2024-1939

Reviewer: The manuscript derives "root water storage capacity" (Sr) from GRACE and GRACE-FO observations of terrestrial water storage (TWS), along with uncertainty estimates. The GRACE-based Sr estimates are compared to Sr estimates derived (i) from soil parameters (soil depth and soil water holding capacity) and (ii) water balance estimates (using precipitation and evapotranspiration [ET] observations). The authors find that the GRACE-based Sr estimates are 50% larger than those derived from water balance estimates and 380% than those derived from soil parameters. The different Sr estimates are further used to parameterize a USGS "bucket model", with TWS and ET output from the model validated against GRACE TWS observations and ET estimates from a water balance approach. Finally, the authors find that their GRACE-based Sr estimates correlate "realistically" with vegetation productivity data.

The authors address a clear need for accurate estimates of root zone water storage capacity, a topic of interest to HESS readers. However, the findings of the manuscript are not supported with independent observations and are largely circular. It is no surprise that the GRACE-based Sr estimates have a relatively lower error against GRACE-based TWS observations. Specifically, the GRACE-based Sr estimates essentially reflect the range of the GRACE TWS observations, and the NSE metrics primarily measures skill in terms of the mean-square error (MSE). Additionally, it remains unclear to me how the authors remove the groundwater signal from the TWS observations. I recommend that the manuscript be rejected.
Response: We appreciate the reviewer's feedback. In the revised manuscript, we will perform a new validation effort that employs independent datasets that do not have GRACE/FO inputs. This validation approach will be as robust as those used in prior studies. Below, we summarize key validation methods used in similar studies, clarify the rationale for the revised validation approach that we will adopt, and outline how we will present the strengths and limitations of this new validation effort.

1. *Challenges in validating S_r and methods used in previous studies:* Validating large-scale S_r remains inherently difficult because direct measurement of S_r is challenging. Previous studies have primarily employed two indirect validation methods:
    a. Rooting-depth comparison: Stocker et al. (2023) converted their deficit-based S_r estimates (~5 km resolution) into rooting-depths using soil texture and water-holding capacity parameters, and then compared them to field rooting-

depth measurements aggregated at biome levels to mitigate the scale mismatch. However, this approach is not suitable for our study. Resolved at a much coarser resolution (~300 km), GRACE/FO-derived S_r samples multiple biome types within a single observational footprint, making biome-level aggregation less meaningful. Additionally, the rooting-depth validation method overlooks groundwater contributions to S_r, which Stocker et al. (2023) found to be significant in over half of their measurement sites. This omission will likely become more critical at the spatial scale of GRACE/FO, which averages larger areas and includes more diverse biome types. These factors make the rooting-depth comparison unsuitable for evaluating GRACE/FO-derived S_r.

b. Implementation in a hydrological model: Wang-Erlandsson et al. (2016) used deficit-based S_r estimates in a simple hydrological model and assessed improvements in simulating hydrologic time series. While this approach better aligns with the spatial scale of GRACE/FO, it faces challenges, too. One is the limited availability of high-quality global hydrologic data, which can lead to a circular use of the same data for both S_r estimation and model evaluation, as Wang-Erlandsson et al. (2016) did with satellite-based ET data. This reduces the independence of the validation process. Additionally, the mechanistic linkage between S_r and commonly used hydrological indicators (e.g., ET and streamflow) is complex – pinpointing decisive indicators that are strongly sensitive to S_r is an important research topic yet to be addressed in the literature. Resolving such a complex relationship can be further complicated by the structural errors or uncertainties in other parameters adopted in the model. Together, these challenges can obscure the true impact of accurate S_r parameterization on ecohydrology. For example, in our study, streamflow simulated by the USGS model is mainly driven by precipitation and shows little sensitivity to S_r, similar to what was described in the open peer review file of Wang-Erlandsson et al. (2016), which also did not use streamflow measurements for model evaluation.

2. *Revised validation approach and its rationale*: We will use the latest version (v4.1) of the Global Land Evaporation Amsterdam Model (GLEAM) ET dataset (https://www.gleam.eu/) to validate our model results. The GLEAM ET addresses key shortcomings present in other gridded ET products. For example, it combines hybrid learning from eddy-covariance and sap flow to capture vegetation response to drought more accurately (Koppa et al., 2022), and it explicitly accounts for plant access to groundwater (Hulsman et al., 2023). Importantly, the GLEAM ET is independent of GRACE/FO and, therefore, allows robust validation that is free from circularity. To mitigate the impact of possible biases embedded in GLEAM ET, forcing data, and those caused by model uncertainty (as the USGS model is uncalibrated), we will use standardized ET anomalies (i.e., Z-scores) as the target of validation and focus on assessing whether S_r improves the temporal dynamics of ET simulations (i.e., seasonal and interannual variations) rather than the absolute values of ET.

3. *Strengths and limitations of the proposed new validation efforts*: The key strength of our revised validation approach lies in its use of an independent dataset (GLEAM

ET), which addresses the potential circularity of our current validation efforts. We will also examine and discuss the following limitations in the revised manuscript.

a. Challenges in detecting S_r influence: Given the uncertainties of modeling S_r's role in ET dynamics, the improvements in S_r may be challenging to detect, particularly when using large-scale models that rely heavily on precipitation-driven processes. We will examine if and to what extent this can be mitigated by using standardized ET anomalies (Z-scores) as the validation target.

b. Focus on temporal dynamics over absolute values: The proposed use of standardized ET anomalies (Z-scores) shifts the focus from absolute ET values to temporal dynamics (seasonal and interannual variations). While this helps mitigate the impact of data biases, it may also limit the scope of the validation to detecting only temporal variations and not necessarily capturing the full range of hydrological dynamics influenced by S_r.

Despite these limitations, the revised validation effort will represent a substantial improvement over Wang-Erlandsson et al. (2016) by using independent, high-quality ET data and focusing on the temporal dynamics of ET.

We will also clarify the definition of root zone storage capacity (S_r), acknowledging the inclusion of natural groundwater fluctuations to meet plant water demands, supported by recent studies and field evidence, including our comparison dataset S_r^{accum} from Stocker et al. (2023). Although this broadens the traditional definition of the root zone, it helps delineate the true amount of water available to plants and is consistent with evolving research on groundwater-vegetation interactions.

Major comments:

- Reviewer: The validation approach is circular (contrary to the statement in Lines 137-140). The GRACE-based Sr estimates reflect, by construction, approximately the dynamic range of the validating GRACE TWS observations (as shown in Figure 1). The surface meteorological forcing inputs to the USGS model are the same for all three simulations, and the only difference between the USGS model configurations is in the Sr parameters. The simulated TWS and ET will therefore have very similar *standardized* anomalies (Z-scores), and the key determinant of the NSE metric will be whether the dynamic range of the simulated TWS anomalies matches that of the verifying observations. The latter were used to determine the GRACE-based Sr, thereby essentially guaranteeing a lower MSE and higher NSE for the simulation with the GRACE-based Sr relative to the other simulations. (As an aside, Line 226 refers to "performance in simulating TWS temporal dynamics". This is a bit of an overstatement given the fact that the experiment design primarily measures how well the estimated Sr reflects the dynamic range of the TWS observations. "Temporal dynamics" suggests skill differences in seasonal and interannual variations, which are not explicitly examined and which are likely to be small, given the experiment setup.)

  Response: You raised a good point here. To minimize the influence of dynamic range on the NSE metric, we reanalyzed our results using standardized anomalies (i.e., Z-scores) for both simulated and GRACE/FO-observed TWS time series. By using Z-scores, we standardized the dynamic range while preserving temporal dynamics, including seasonal and interannual variations. Contrary to the reviewer's assumption, our analysis shows that, even after standardizing the anomalies, the TWS simulations with different S_r parameterizations exhibit distinct patterns.

For example, Fig. RC3_1a compares the Z-scores of TWS from GRACE/FO and three simulations (*HydroModel*($S_r^{GRACE/FO}$), *HydroModel*($S_r^{RD\times WHC}$), and *HydroModel*(S_r^{accum})) for the mascon location in Figure 1 of the original submission. The NSE values for the Z-scores time series indicate that $S_r^{GRACE/FO}$ outperforms $S_r^{RD\times WHC}$ and S_r^{accum} in capturing TWS temporal dynamics (Fig. RC3_1b-d). This improvement is widespread (Fig. RC3_2) and overlaps with those based on the original time series (Figure 5 of the original text). Notably, this enhancement extends into many subtropical and Southern Hemisphere regions, where the USGS model struggles to simulate the dynamic range of GRACE/FO TWS.

To address the circularity concern, we will no longer use GRACE/FO TWS as a validation dataset. Instead, we will validate the model using GLEAM ET and evaluate it using standardized ET anomalies (i.e., Z-scores). GLEAM ET is independent of GRACE/FO and, therefore, allows robust validation and avoid the circularity concern. The Z-score approach also allows us to assess the model's ability to capture seasonal and interannual variations without undue influence from potential biases embedded in GLEAM ET, forcing data, and those caused by the uncalibrated nature of the USGS model.

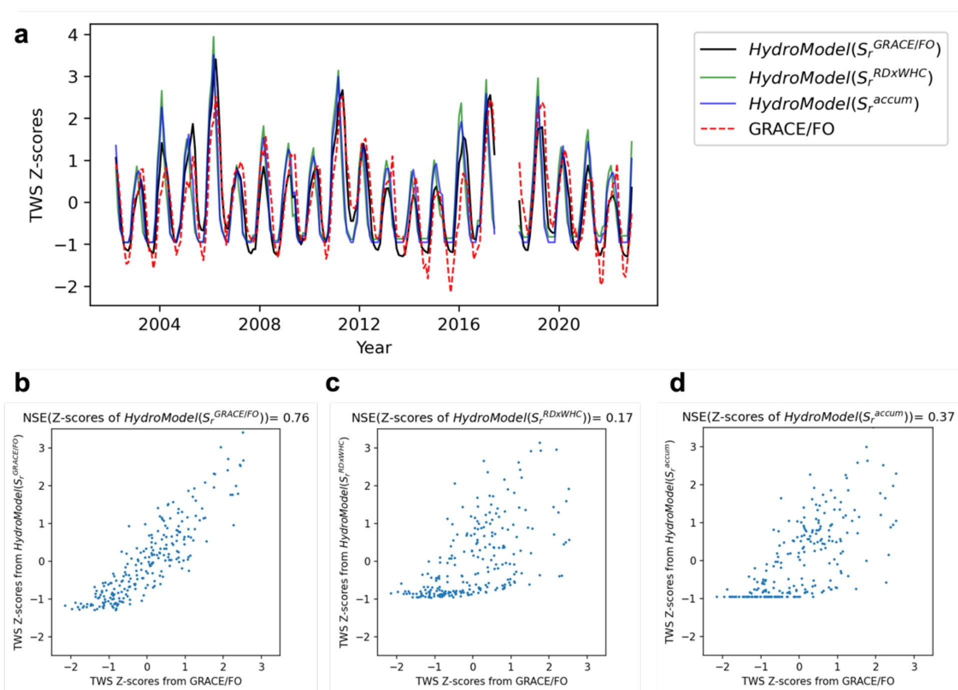

Figure RC3_1. A comparison of model predictive skills for TWS z-scores. (a) Z-score time series comparison between GRACE/FO TWS and model simulations. (b)-(d) Scatterplots of

GRACE/FO TWS z-scores and simulated TWS z-scores from *HydroModel*($S_r^{GRACE/FO}$), *HydroModel*($S_r^{RD \times WHC}$), *HydroModel*(S_r^{accum}), respectively.

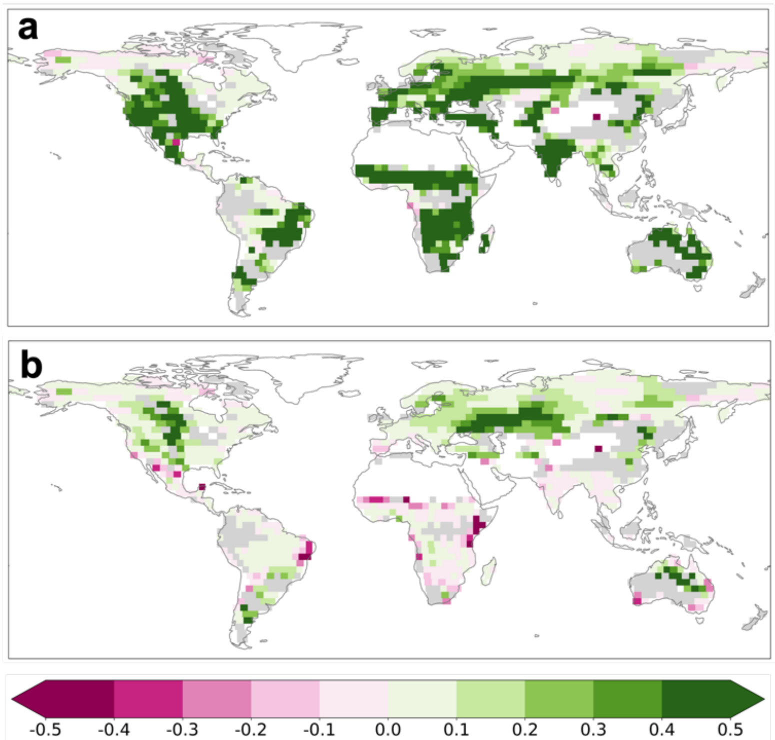

Figure RC3_2. Predictive skill differences for TWS z-scores. (a) The NSE difference between *HydroModel*($S_r^{GRACE/FO}$) and *HydroModel*($S_r^{RD \times WHC}$). (b) The NSE difference between *HydroModel*($S_r^{GRACE/FO}$) and *HydroModel*(S_r^{accum}). The gray colors indicate areas where all models fail to achieve a positive NSE value.

- Reviewer: The ET estimates used to validate the USGS model simulations are based on water balance estimates derived from precipitation and water storage change datasets, which is similarly circular when it comes to validating the model output from the simulations that use Sr estimates based on GRACE observations or water balance estimates.
  Response: In our revised manuscript, we will use GLEAM ET, which is independent of GRACE/FO and addresses key shortcomings in other gridded ET products.

- Reviewer: The definition of Sr as "root zone storage capacity" seems inconsistent the derivation from GRACE TWS observations. The authors explain how they remove the snow signal and anthropogenic groundwater signals from the TWS observations when they derive the GRACE-based Sr estimates. However, it remains unclear how natural groundwater fluctuations are handled. TWS observations include natural variations in groundwater levels that are not related to water storage in what would usually be considered the “root zone” (e.g., in grasslands). Perhaps it is intentional that such fluctuations are included, but then the derived parameter is then no longer a “root zone storage capacity” in the sense that the control volume is no longer what is commonly understood to be the “root zone”.

Response: Thank you for this comment. Natural groundwater variability is indeed embedded in our calculation of root-zone water storage capacity (S_r), and we provide clarification below. As Reviewer#1 correctly pointed out, root-accessible water does not require roots to physically occupy the entire storage domain. Processes such as the capillary rise can move deep water upward to the root zone for vegetation transpiration, especially during dry seasons and droughts. Many studies have shown that natural groundwater variability (such as its seasonal variation) strongly correlates with the net effect of precipitation and ET (e.g., Li et al., 2015).

Including groundwater in the calculation of S_r extends the traditional definition of the “root zone,” beyond the soil layer by recognizing the fact that the root zone is dynamic and can access deep groundwater and bedrock moisture during prolonged droughts and high transpiration demand (Gao et al., 2024). Several recent studies (McCormick et al., 2021; Singh et al., 2020; Stocker et al., 2023) have also included groundwater in their definitions of S_r. This inclusion is well-supported by recent studies based on in situ groundwater (Baldocchi et al., 2021; Fan et al., 2017; Li et al., 2015; Thompson et al., 2011), remote sensing observations (Koirala et al., 2017; Rohde et al., 2024), and modeling efforts (Hain et al., 2015; Miguez-Macho & Fan, 2021), all of which showed that groundwater significantly contributes to ET and is accessible to plants, especially during extreme droughts.

In many ecosystems, water stress can stimulate root growth into deep subsurface through the capillary rise effect, with roots extending to the capillary fringe and the water table, as observed in both field and laboratory studies (Fan et al., 2017; Kuzyakov & Razavi, 2019; Naumburg et al., 2005; Orellana et al., 2012). Although individual shallow-rooted plants (e.g., grassland sites) may not directly tap into groundwater, the large spatial scale of GRACE/FO data likely captures water uptake across diverse vegetation types. This blending makes it likely that vegetation types not typically associated with groundwater use may still access it indirectly, such as through hydraulic redistribution by neighboring deeper-rooted plants (e.g., Espeleta et al., 2004; Orellana et al., 2012). Indeed, satellite observations have revealed widespread plant-groundwater interactions at large spatial scales (Koirala et al., 2017), even in dryland regions dominated by grasslands (Rohde et al., 2024; Wang et al., 2023).

Neglecting groundwater in root zone storage capacity can lead to underestimation of land and air interactions (Dong et al., 2022; Maxwell & Condon, 2016; Schlemmer et al., 2018), affect accurate simulation of runoff (Hahm et al., 2019), and misrepresent vegetation resilience to droughts and heat waves (Esteban et al., 2021; Jiménez-Rodríguez et al., 2022).

Overall, our $S_r^{GRACE/FO}$ definition aligns with our comparison dataset S_r^{accum} from Stocker et al. (2023) and helps explain why the traditional rooting depth approach ($S_r^{RD \times WHC}$), which does not include groundwater, yields lower values than $S_r^{GRACE/FO}$ and S_r^{accum}. This expanded definition is consistent with emerging research on groundwater-vegetation interactions. We will add these discussions to the revised manuscript.

- Reviewer: It is highly concerning that no model attains positive NSE values for 40% of the global *vegetated* domain (Lines 216-217). This area includes most of the subtropics and Southern Hemisphere! If the model is so poor that for nearly half of the domain of interest a time-invariant constant would be a better estimator, what does it say about the skill of the model in the other half of the domain? And what does it mean for the Sr estimates in nearly half of the domain of interest where NSE is negative for all three model simulations?
  Response: The USGS model, which was run without any local calibration, failed to have positive NSE values in 40% of the vegetated regions, primarily due to its inability to capture the dynamic range of GRACE/FO-observed TWS (Figs. A4 and A5, and the discussion from lines 328 to 347 of the original submission). This underperformance is likely due to uncalibrated parameters and the model's simplified representation of key hydrological processes. However, by applying the Z-score approach — which minimizes the impact of dynamic range mismatch on the NSE metric — Fig. RC3_2 shows that the USGS model effectively captures TWS temporal variations in many subtropical and Southern Hemisphere regions. The area with no positive NSE values was reduced from 40% to 24%, indicating that the USGS model still provides valuable insights into the relative accuracy of the three S_r estimates in most global vegetated regions.

  Although 24% of the domain continues to show negative NSE values, this does not invalidate the $S_r^{GRACE/FO}$ estimates. Rather, it highlights regions where further investigation and refinement are needed. Future work could involve local calibration of model parameters or using more sophisticated hydrological models to improve accuracy in these challenging areas. Despite the negative NSE values, the $S_r^{GRACE/FO}$ estimate remains informative, offering valuable insights into water storage dynamics when interpreted within the context of known model limitations.

  Given the discussions above, while the Z-score-based GRACE/FO TWS results are informative, we will not include them in the revision. Instead, we will use the GLEAM ET dataset for model validation to ensure our validation is independent of GRACE/FO and free from circularity.

Minor comments:

1. Reviewer: The heading of section 3 should probably be "Results"
   Response: We will use "Results" in our revised manuscript.

2. Reviewer: The caption of Figure 3 does not clearly state the base for the "percentage changes". This can only be understood from the text.
   Response: We will change the caption to "(a) and (b) are the consumption percentages **of** $S_r^{GRACE/FO}$ during the second and third-largest TWS drawdowns."

3. Reviewer: Line 208: Be more specific about the "drier climates and lower-biomass regions"
   Response: We will specify these regions in our revised manuscript.

References

Baldocchi, D., Ma, S., & Verfaillie, J. (2021). On the inter- and intra-annual variability of ecosystem evapotranspiration and water use efficiency of an oak savanna and annual grassland subjected to booms and busts in rainfall. *Global Change Biology, 27*(2), 359-375. https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15414

Dong, J., Lei, F., & Crow, W. T. (2022). Land transpiration-evaporation partitioning errors responsible for modeled summertime warm bias in the central United States. *Nature Communications, 13*(1), 336. https://doi.org/10.1038/s41467-021-27938-6

Espeleta, J. F., West, J. B., & Donovan, L. A. (2004). Species-specific patterns of hydraulic lift in co-occurring adult trees and grasses in a sandhill community. *Oecologia, 138*(3), 341-349. https://doi.org/10.1007/s00442-003-1460-8

Esteban, E. J. L., Castilho, C. V., Melgaço, K. L., & Costa, F. R. C. (2021). The other side of droughts: wet extremes and topography as buffers of negative drought effects in an Amazonian forest. *New Phytologist, 229*(4), 1995-2006. https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.17005

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences, 114*(40), 10572-10577. https://www.pnas.org/content/pnas/114/40/10572.full.pdf

Gao, H., Hrachowitz, M., Wang-Erlandsson, L., Fenicia, F., Xi, Q., Xia, J., et al. (2024). Root zone in the Earth system. *EGUsphere, 2024*, 1-30. https://egusphere.copernicus.org/preprints/2024/egusphere-2024-332/

Hahm, W. J., Dralle, D. N., Rempe, D. M., Bryk, A. B., Thompson, S. E., Dawson, T. E., & Dietrich, W. E. (2019). Low Subsurface Water Storage Capacity Relative to Annual Rainfall Decouples Mediterranean Plant Productivity and Water Use From Rainfall Variability. *Geophysical Research Letters, 46*(12), 6544-6553. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083294

Hain, C. R., Crow, W. T., Anderson, M. C., & Yilmaz, M. T. (2015). Diagnosing Neglected Soil Moisture Source–Sink Processes via a Thermal Infrared–Based Two-Source Energy Balance Model. *Journal of Hydrometeorology, 16*(3), 1070-1086. https://journals.ametsoc.org/view/journals/hydr/16/3/jhm-d-14-0017_1.xml

Hulsman, P., Keune, J., Koppa, A., Schellekens, J., & Miralles, D. G. (2023). Incorporating Plant Access to Groundwater in Existing Global, Satellite-Based Evaporation Estimates. *Water Resources Research, 59*(8), e2022WR033731. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR033731

Jiménez-Rodríguez, C. D., Sulis, M., & Schymanski, S. (2022). Exploring the role of bedrock representation on plant transpiration response during dry periods at four forested sites in Europe. *Biogeosciences, 19*(14), 3395-3423.

Koirala, S., Jung, M., Reichstein, M., de Graaf, I. E. M., Camps-Valls, G., Ichii, K., et al. (2017). Global distribution of groundwater-vegetation spatial covariation. *Geophysical Research Letters, 44*(9), 4134-4142. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL072885

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nature Communications, 13*(1), 1912. https://doi.org/10.1038/s41467-022-29543-7

Kuzyakov, Y., & Razavi, B. S. (2019). Rhizosphere size and shape: Temporal dynamics and spatial stationarity. *Soil Biology and Biochemistry, 135*, 343-360. https://www.sciencedirect.com/science/article/pii/S0038071719301452

Li, B., Rodell, M., & Famiglietti, J. S. (2015). Groundwater variability across temporal and spatial scales in the central and northeastern U.S. *Journal of Hydrology, 525*, 769-780. https://www.sciencedirect.com/science/article/pii/S0022169415002929

Maxwell, R. M., & Condon, L. E. (2016). Connections between groundwater flow and transpiration partitioning. *Science, 353*(6297), 377-380. https://www.science.org/doi/abs/10.1126/science.aaf7891

McCormick, E. L., Dralle, D. N., Hahm, W. J., Tune, A. K., Schmidt, L. M., Chadwick, K. D., & Rempe, D. M. (2021). Widespread woody plant use of water stored in bedrock. *Nature, 597*(7875), 225-229. https://doi.org/10.1038/s41586-021-03761-3

Miguez-Macho, G., & Fan, Y. (2021). Spatiotemporal origin of soil water taken up by vegetation. *Nature, 598*(7882), 624-628. https://doi.org/10.1038/s41586-021-03958-6

Naumburg, E., Mata-gonzalez, R., Hunter, R. G., McLendon, T., & Martin, D. W. (2005). Phreatophytic Vegetation and Groundwater Fluctuations: A Review of Current Research and Application of Ecosystem Response Modeling with an Emphasis on Great Basin Vegetation. *Environmental Management, 35*(6), 726-740. https://doi.org/10.1007/s00267-004-0194-7

Orellana, F., Verma, P., Loheide II, S. P., & Daly, E. (2012). Monitoring and modeling water-vegetation interactions in groundwater-dependent ecosystems. *Reviews of Geophysics, 50*(3). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011RG000383

Rohde, M. M., Albano, C. M., Huggins, X., Klausmeyer, K. R., Morton, C., Sharman, A., et al. (2024). Groundwater-dependent ecosystem map exposes global dryland protection needs. *Nature, 632*(8023), 101-107. https://doi.org/10.1038/s41586-024-07702-8

Schlemmer, L., Schär, C., Lüthi, D., & Strebel, L. (2018). A Groundwater and Runoff Formulation for Weather and Climate Models. *Journal of Advances in Modeling*

Earth Systems, 10(8), 1809-1832. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017MS001260

Singh, C., Wang-Erlandsson, L., Fetzer, I., Rockström, J., & van der Ent, R. (2020). Rootzone storage capacity reveals drought coping strategies along rainforest-savanna transitions. *Environmental Research Letters, 15*(12), 124021. https://dx.doi.org/10.1088/1748-9326/abc377

Stocker, B. D., Tumber-Dávila, S. J., Konings, A. G., Anderson, M. C., Hain, C., & Jackson, R. B. (2023). Global patterns of water storage in the rooting zones of vegetation. *Nature Geoscience*. https://doi.org/10.1038/s41561-023-01125-2

Thompson, S. E., Harman, C. J., Konings, A. G., Sivapalan, M., Neal, A., & Troch, P. A. (2011). Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater. *Water Resources Research, 47*(10). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009797

Wang, T., Wu, Z., Wang, P., Wu, T., Zhang, Y., Yin, J., et al. (2023). Plant-groundwater interactions in drylands: A review of current research and future perspectives. *Agricultural and Forest Meteorology, 341*, 109636. https://www.sciencedirect.com/science/article/pii/S0168192323003271

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., et al. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrol. Earth Syst. Sci., 20*(4), 1459-1481. https://hess.copernicus.org/articles/20/1459/2016/