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

Reviewer 1

Reviewer: Review of "Substantial root-zone water storage capacity observed by GRACE and GRACE/FO" by Zhao et al. . This paper describes the use of TWS estimates from the GRACE satellite project to estimate multi-year water storage changes. Negative changes are used to estimate a lower-bound on root-zone water storage (Sr). The estimates are compared to two alternative (Sr) methodologies, and all three Sr estimates are used to parameterize a hydrologic model. The main result is that the authors' Sr is significantly larger than the previously described Sr estimates.

Comments

General

I found the authors' use of GRACE data to be novel, and the results interesting. The paper is well-written and generally clear.
Response: Thanks for the positive feedback.

As with other GRACE studies, the spatial resolution of the data is relatively coarse, so I suggest adding some discussion of how these results might be applied in the operational configuration of land models, which would typically have finer spatial resolution.
Response: Thanks for your comment and suggestion. In our revised manuscript, we will discuss two ways to use our S_r estimates for land surface modeling. First, $S_r^{GRACE/FO}$ can be used for evaluating default S_r parameterization used in the model at the coarse-scale of GRACE/FO data in conjunction with other analyses. For instance, if a model simulates low ET during droughts in a region where the S_r value is also low compared to $S_r^{GRACE/FO}$, the default value may be increased based on $S_r^{GRACE/FO}$. Second, we will discuss approaches for developing finer-scale GRACE-based S_r products, such as using downscaled TWS products developed through machine learning and data assimilation techniques (Gou & Soja, 2024; Li et al., 2019) .

Reviewer: Abstract
The maximum water held would be the difference from saturation to wilting point. But saturated conditions are unlikely to occur at these spatial scales in many regions.
Response: Our root zone storage capacity includes water uplifted from groundwater, and thus, it is not limited by the wilting point and saturation. In the revision, we will rephrase it to "the maximum amount of water available for plant uptake."

Reviewer: 1st sentence defines Sr, and the next sentence discusses simulations. Perhaps add a sentence indicating how Sr is used in a modeling context after the 1st sentence to provide context.
Response: Thanks. We will add a couple of sentences to explain why and how S_r is relevant to model simulation.

Reviewer: Line 15: to be clear, GRACE measures gravity and TWS is inferred from that, so the use of the word 'direct' can be problematic. There are other geophysical processes that affect time-varying gravity.
Response: We will remove 'direct.'

Reviewer: Line 20: what does 'correlates realistically' mean? Can you use a more specific or quantitative description?
Response: We will specify the relationship as a saturating relationship in the revision.

Reviewer: Introduction
Line 26: 'plants can store during wet periods' should be 'plants can access'? i.e. plants aren't storing the water, the soil is storing water.
Response: We will revise the sentence to "the more water plant roots can access for use in droughts."

Reviewer: Line 37: why would it overlook rock moisture and groundwater? This sentence implies a different reason besides uncertainties in rooting depths or hydraulic properties, which are mentioned previously.
Response: Thank you for this comment. The reason water stored in weathered rocks and groundwater is often overlooked is that most approaches typically set rooting depth shorter than simulated soil thickness and assume that roots do not extend into deeper unsaturated zones. However, recent studies have shown that this assumption is not always accurate (Rempe and Dietrich, 2018; Fan et al., 2017). In fact, in many ecosystems, plant roots can penetrate beyond the shallow soil layer into weathered bedrocks to access deep water storage, including groundwater, especially during dry seasons and droughts (Mccormick et al., 2021; Maxwell and Condon, 2016).

We will clarify this and highlight the importance of bedrock moisture and groundwater in our definition of root-zone storage capacity in the revised manuscript.

Reviewer: Line 49: again, the word 'direct' I find problematic. If you wish to use this word, perhaps add a sentence explaining its use.
Response: We will remove 'direct.'

Reviewer: Methods

Line 61: clearly, 'root-zone' implies vegetated areas, but what might one learn from this method in more arid regions?
Response: Thank you for your question. In more arid regions such as deserts, our approach may capture moisture storage capacity for bare soil evaporation. We will clarify this point in the revision.

Reviewer: Line 68: typically P, ET, and R refer to fluxes. To be more consistent with other literature, consider using rate or flux units consistently and include a summation symbol in equation 1.
Response: We will use flux units and add a summation symbol in equation 1 in the revision.

Reviewer: Line 75: 'consumed' could be changed to 'transpired' or 'returned to the atmosphere'
Response: We will change it to 'transpired.'

Reviewer: Line 82: in areas in which widespread groundwater use is absent, how will this trend removal affect your results? Is it likely to increase or decrease your Sr estimates for such areas? Could you use maps of irrigated area, such as AQUASTAT, to confine this operation to areas where widespread irrigation occurs?
Response: You raised a good point here. In some cases, long-term trends in TWS can be associated with precipitation trends in responses to climate change. In those cases, removing long-term linear trends likely leads to underestimation of S_r. However, we found that regions showing significant TWS decreasing trends largely coincide with known irrigation areas identified in AQUASTAT data, except in some high Arctic locations (Figs. RC1_1a, b). Thus, our S_r estimates may be underestimated in these high Arctic regions. We will discuss this limitation in the revision.

The AQUASTAT dataset has its own uncertainties and limitations. For instance, it is based on statistics from 2000-2008 and is particularly uncertain in high-latitude regions (Fig. RC1_1c). Consequently, it may not provide reliable information on groundwater use in some areas of the globe.

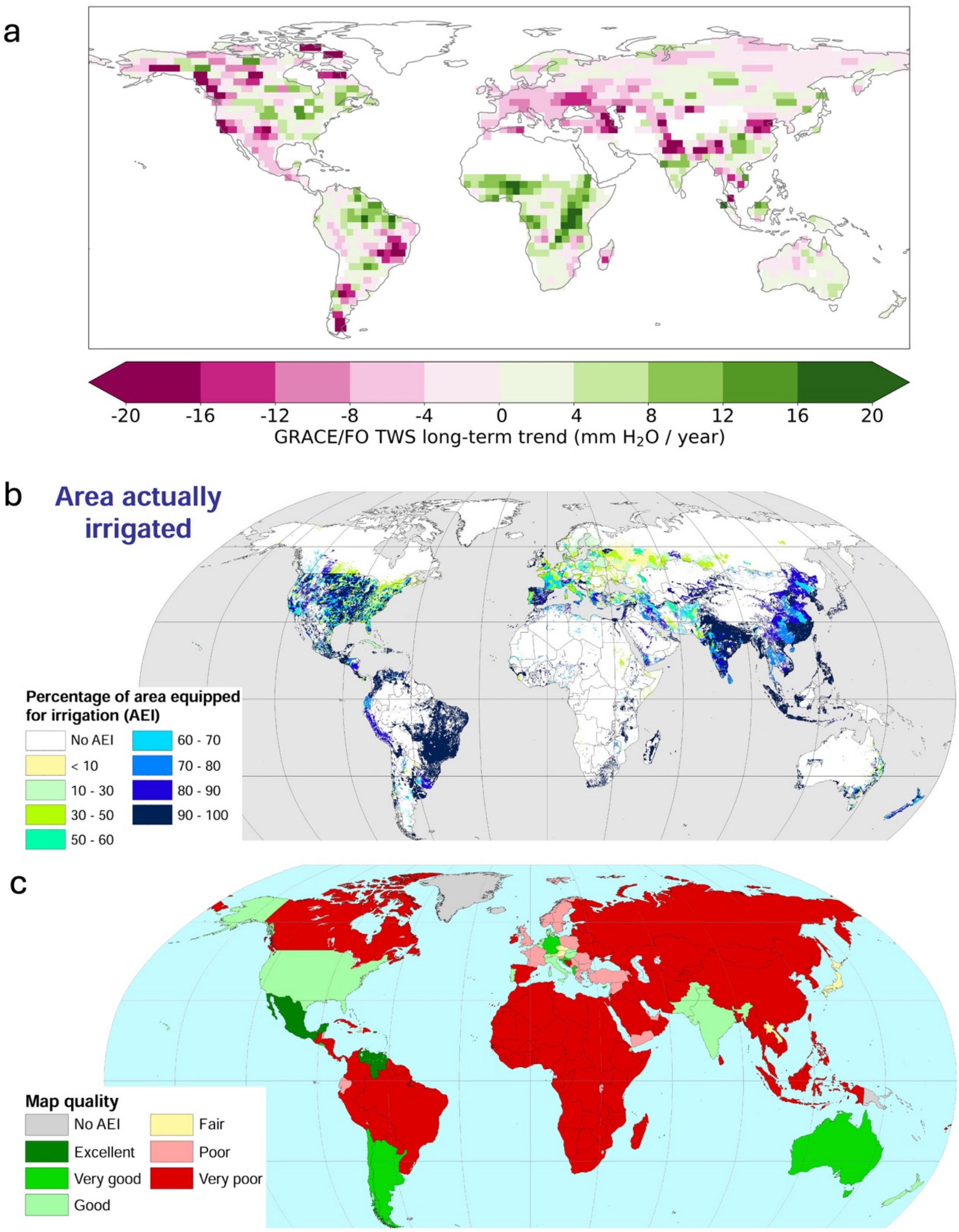


Figure RC1_1. (a) Trends in TWS obtained from GRACE/FO observations from 2002 to 2022. (b) Percentage of area equipped for irrigation that is actually irrigated. (c) Map quality marks assigned to each country for area equipped for irrigation in (b). (b-c) are from the Global Map of Irrigation Areas – version 5.0 by AQUASTAT available at https://firebasestorage.googleapis.com/v0/b/fao-aquastat.appspot.com/o/PDF%2FMAPS%2Fgmia_v5_lowres.pdf?alt=media&token=d098a48f-ab49-4eae-a16e-82a5779f924e

Reviewer: Line 91: how runoff is used here is not clear to me. Is there a budget equation that could be shown? What does 'surface water' encompass; rivers, lakes, reservoirs, ...?
Response: Yes, surface water here encompasses water stored in rivers, lakes, and reservoirs. In GRACE/FO TWS decomposition studies (Bhanja et al., 2016; Getirana et al., 2017; Shamsudduha & Taylor, 2020; Thomas et al., 2017; Wang et al., 2023), surface runoff (Q) is commonly used as a proxy for surface water storage change (ΔSW), expressed as $\Delta SW = Q$. This approach assumes Q directly contributes to an increase in surface water levels within the drainage network. This approach also assumes that it takes approximately one month for Q to exit the drainage system, aligning with the monthly time step of GRACE/FO data.

In our study, we used total runoff (R), which includes both surface runoff (Q) and subsurface runoff, as a proxy for ΔSW (i.e., $\Delta SW = R$), as subsurface runoff which is groundwater discharge to rivers also contributes to surface water storage changes.

We will clarify the methodology and justification further in the revised manuscript and include the water budget equation ($\Delta SW = R$) for clarity.

Reviewer: Line 109: to what extent is Yang 2016 a model-based dataset versus an observational dataset?
Response: Yang et al. (2016) is a fully model-based dataset. It relies on Guswa (2008)'s analytical model that estimates rooting depth, which balances the carbon gain and cost of any additional roots. While such model-based datasets are valuable for providing comprehensive coverage and insights into complex processes, they do not incorporate direct observational data for validation or correction. We will discuss this caveat in the revision.

Reviewer: Line 111: how is water holding capacity defined? Field capacity minus wilting point?
Response: The reviewer is correct. Field capacity is defined as the difference between field capacity and permanent wilting point. We will add this definition in our revised manuscript.

Reviewer: Line 132: why is this an approximation? Are there other modeled water storage components in HydroModel that were ignored?
Response: There are no other modeled water storage components in the USGS model. We will clarify this in the revision.

Reviewer: Line 141: 'ET anomalies'
Response: We will correct it in the revised manuscript.

Reviewer: Line 146: Does Xiong 2023 use GRACE water storage for their ET estimates? If so, does that reduce its independence from your results?
Response: Yes, Xiong et al. (2023) used GRACE for their ET estimates. In the revised manuscript, we will replace Xiong et al. (2023) ET estimates with 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 is an improved dataset, addressing key shortcomings present in other gridded ET products. For example, it combines hybrid learning from eddy-covariance and sap flow to capture vegetation responses 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.

Reviewer: Line 169: you say that you compare the two datasets, but you don't explicitly say what your hypothesized relationship between them is, so the justification here seems weak. In areas that are not water limited, one could imagine that GPP would be high, but a deep root zone is not necessary. Perhaps expand further on your reasoning in this paragraph.
Response: Thank you for this suggestion. We will revise this paragraph as follows to justify the GPP analysis in the revised manuscript (new text in bold).

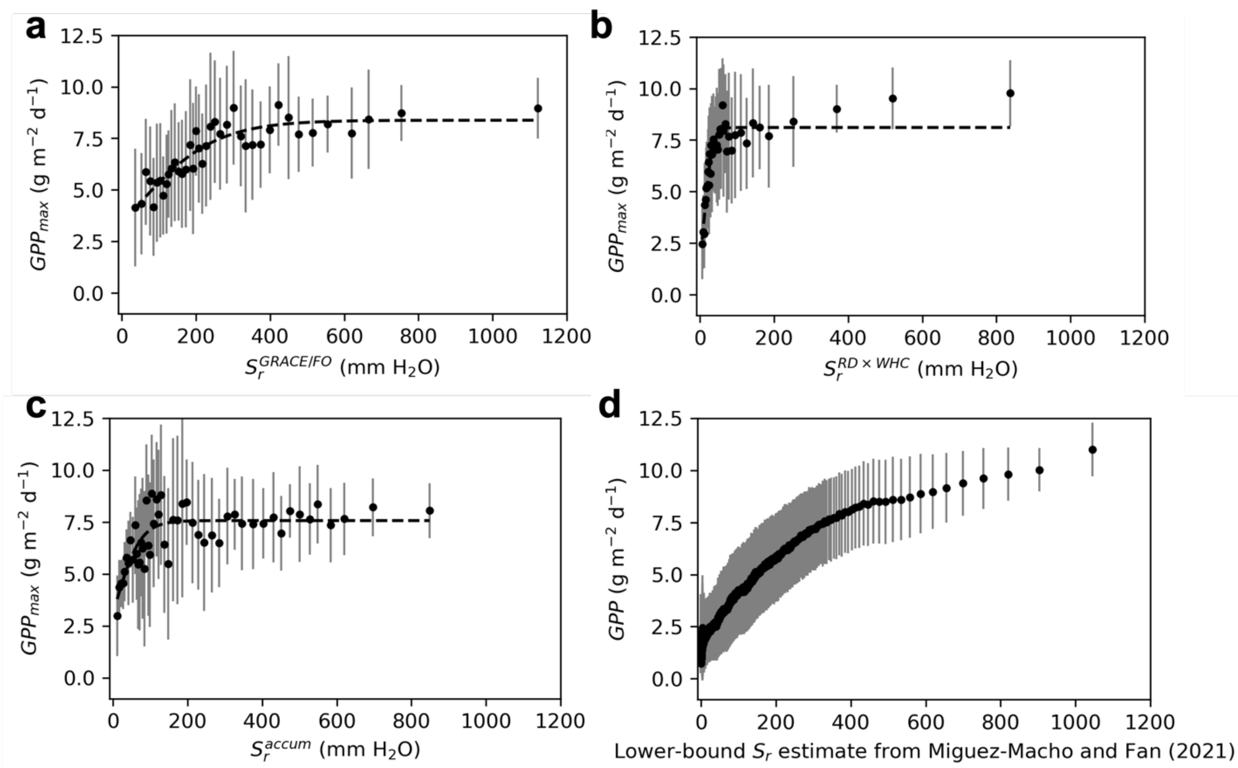
"The $S_r^{GRACE/FO}$ is derived from the water balance, but its ecological relevance remains undetermined. ***We hypothesize that ecosystems with higher biomass, particularly in wetter regions, develop larger S_r as a long-term adaptation to manage periodic water surpluses, promoting water retention and maintaining productivity during dry periods. Even in regions that are not water-limited, a larger S_r may still serve as a buffer against interannual precipitation variability, thus contributing to sustained productivity despite the less direct influence of immediate water limitations. To investigate this hypothesis, we compared $S_r^{GRACE/FO}$ with an independent measure of ecosystem productivity****..."*

Reviewer: Line 194: is this saying that the durations shown in 3c) and 3d) are often larger than that shown in 2b)?
Response: No. The average duration of the first, second, and third-largest TWS drawdowns are 2.8, 1.6, and 1.2 years, respectively. This indicates that the durations in Fig.3 c) and d) are often shorter than those shown in Fig. 2b.

Reviewer: Line 225: Do these patterns correlate with a particular land cover or vegetation type?
Response: We did not find a clear correlation with a particular land cover or vegetation type. This may be due to the large spatial resolution of GRACE/FO data, which represents combined signals from various land cover types within its 3° × 3° footprint. As

a result, it is challenging to isolate patterns specific to individual land cover or vegetation types.

Reviewer: Line 271: plot d) is unclear to me. You create an Sr estimate from Miguez-Macho 2021, but then plot it against transpiration instead of GPP; why is this done differently from a) - c)?
Response: Thank you for pointing this out. The reason for this difference is that Miguez-Macho and Fan (2021) only provided transpiration data, not GPP. To maintain consistency across all plots a) – d) in our revised manuscript, we will convert their transpiration data to GPP. This conversion will be done using a linear regression (R^2 = 0.86) between their transpiration data and the GPP derived from MODIS and FLUXNET. This will ensure that plot d) aligns with plots a) – c) in its use of GPP, addressing the inconsistency you noted. The new figure will be as follows:



Reviewer: Figure 8: why are the x- and y-axis ranges different for plots a) - c)? It is harder to compare the scatterplots because of this.
Response: We will adjust plots a) – d) to make x-axis and y-axis ranges identical for all subplots (see the new figure above).

Reviewer: Discussion

Line 321: does root-accessible water require that the roots physically occupy the entire storage domain? For example, as soils dry, upward moisture fluxes can occur which might replenish soil moisture deficits near roots. Might this help explain the mismatch between observed rooting depths and the Sr estimates here?
Response: Thank you for this insightful comment. You are correct that roots do not necessarily need to physically occupy the entire storage domain. Processes such as the capillary force can indeed move deep water storage upward to replenish moisture near the roots, especially during dry conditions. Such mechanism could be the reason for the observed differences between the rooting depth-based estimation and our GRACE/FO-based estimation. We will include this discussion in the revised manuscript.

Reviewer: Line 325: one could also interpret your Sr/WHC as simply the effective soil depth. For land models that do not use an explicit Sr variable, this could indicate that models with a soil depth < 2m (i.e. some of the GLDAS models) are likely incapable of simulating these kinds of drawdowns, which would have implications for studies of groundwater that have used GLDAS to remove the soil moisture component of TWS.
Response: Agreed. We will discuss these implications in the revised manuscript.

Reviewer: Line 326: 'tap'
Response: Corrected. Thank you.

Reviewer: Figure A1: how does this result relate to the relationship between magnitude and duration? Does it imply that during the largest drawdowns, there is also the largest 'net precipitation'? That seems counterintuitive.
Response: Thank you for your observation. Figure A1 currently shows the cumulative sum of P – R during the drawdown periods. The largest drawdown does indeed correspond to the longest duration, which results in a higher cumulative sum of P – R. We recognize that this might seem counterintuitive, as it suggests that the largest drawdowns also have the largest 'net precipitation.' To clarify this, we will revise the figure to present the average P – R instead of the *cumulative* sum. This adjustment will remove the influence of duration and reflect the mean P – R during the drawdown periods.

References

Bhanja, S. N., Mukherjee, A., Saha, D., Velicogna, I., & Famiglietti, J. S. (2016). Validation of GRACE based groundwater storage anomaly using in-situ groundwater level measurements in India. *J. Hydrol., 543*, 729-738.

Getirana, A., Kumar, S., Girotto, M., & Rodell, M. (2017). Rivers and Floodplains as Key Components of Global Terrestrial Water Storage Variability. *Geophysical Research Letters, 44*(20), 10,359-310,368. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074684

Gou, J., & Soja, B. (2024). Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms. *Nature Water, 2*(2), 139-150. https://doi.org/10.1038/s44221-024-00194-w

Guswa, A. J. (2008). The influence of climate on root depth: A carbon cost-benefit analysis. *Water Resources Research, 44*(2). https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007WR006384

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

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

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research, 55*(9), 7564-7586.

Shamsudduha, M., & Taylor, R. G. (2020). Groundwater storage dynamics in the world's large aquifer systems from GRACE: Uncertainty and role of extreme precipitation. *Earth Syst. Dyn., 11*, 755-774.

Thomas, B. F., Caineta, J., & Nanteza, J. (2017). Global assessment of groundwater sustainability based on storage anomalies. *Geophys. Res. Lett., 44*, 11 445-411 455.

Wang, S., Li, J., & Russell, H. A. (2023). Methods for estimating surface water storage changes and their evaluations. *Journal of Hydrometeorology, 24*(3), 445-461.

Yang, Y., Donohue, R. J., & McVicar, T. R. (2016). Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resources Research, 52*(10), 8260-8276. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019392