Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL115164

Special Collection:
Advances in Modeling Soil System Science, Earth System Science, and Beyond

Underestimation of Historical Terrestrial Water Storage Droughts in Global Water Models

Amar Deep Tiwari[1], Yadu Pokhrel[1], Farshid Felfelani[2], Ahmed Elkouk[1], Julien Boulange[3], Simon N. Gosling[4], Naota Hanasaki[5], Aristeidis Koutroulis[6], Vimal Mishra[7], Hannes Müller Schmied[8,9], Yusuke Satoh[10], Sebastian Ostberg[11], Tobias Stacke[12], and Jiabo Yin[13]

[1]Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI, USA, [2]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [3]United Graduate School of Agricultural Science, Tokyo University of Agriculture and Technology, Tokyo, Japan, [4]School of Geography, University of Nottingham, Nottingham, UK, [5]National Institute for Environmental Studies (NIES), Tsukuba, Japan, [6]School of Chemical and Environmental Engineering, Technical University of Crete, Chania, Greece, [7]Department of Civil Engineering, Indian Institute of Technology, Gandhinagar, India, [8]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany, [9]Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany, [10]Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, [11]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [12]Max Planck Institute for Meteorology, Hamburg, Germany, [13]State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China

Abstract Enhanced drought modeling is crucial for realistic prediction and effective management of water resources, especially with climate change anticipated to exacerbate drought frequency and severity. Global water models (GWMs) simulate historical and future terrestrial water storage (TWS) with continuous spatial and temporal coverage. However, a global evaluation of TWS simulations by GWMs focused on drought is lacking. Here we evaluate, for the first time, GWMs' capability to represent TWS droughts by comparing simulations with Gravity Recovery and Climate Experiment satellite data. We find notable underestimation of drought severity and coverage by GWMs, across diverse regions, including North America, South America, Africa, and Northern Asia. When examined without trend removal, the underestimation of TWS droughts is more pronounced in recent years (2016–2019) compared to 2002–2015, especially in northern latitudes. This underrepresentation highlights the necessity to improve GWMs to simulate TWS droughts. Our results imply that previously reported future TWS projections could have underestimated droughts.

Plain Language Summary Droughts are worsening, likely due to climate change, making accurate water resource projections crucial. This study evaluates global water models (GWMs) against Gravity Recovery and Climate Experiment (GRACE) satellite data for terrestrial water storage (TWS)—an integration of water stored in snowpacks, rivers, lakes, floodplains, canopies, soils, and aquifers—to assess their accuracy in simulating droughts. Results indicate that GWMs systematically underestimate the severity and spatial extent of TWS droughts, particularly in tropical and high latitude regions such as North and South America, and Africa. This underestimation suggests that future drought hazards might be higher than projected, posing challenges for water management and agriculture. These discrepancies highlight the need for model enhancements and more rigorous validation to improve drought projections, which are critical for effective water resource management and climate adaptation strategies.

Key Points:

- Global water models (GWMs) underestimate terrestrial water storage (TWS) drought severity
- The underestimations span diverse climate regions
- Better representation of human interventions in the GWMs could improve TWS drought prediction

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Pokhrel,
ypokhrel@msu.edu

Citation:

Tiwari, A. D., Pokhrel, Y., Felfelani, F., Elkouk, A., Boulange, J., Gosling, S. N., et al. (2025). Underestimation of historical terrestrial water storage droughts in global water models. *Geophysical Research Letters*, *52*, e2025GL115164. https://doi.org/10.1029/2025GL115164

Received 31 JAN 2025
Accepted 16 SEP 2025

Author Contributions:

Conceptualization: Amar Deep Tiwari, Yadu Pokhrel
Data curation: Amar Deep Tiwari
Formal analysis: Amar Deep Tiwari
Funding acquisition: Yadu Pokhrel
Investigation: Yadu Pokhrel
Methodology: Amar Deep Tiwari
Project administration: Yadu Pokhrel
Resources: Amar Deep Tiwari, Yadu Pokhrel
Software: Amar Deep Tiwari
Supervision: Yadu Pokhrel
Validation: Amar Deep Tiwari

1. Introduction

Terrestrial water storage (TWS), a fundamental component of the terrestrial hydrological budget (Pokhrel et al., 2021; Tang et al., 2010), represents the total water stored above and below the land surface (Syed et al., 2008), including in snowpacks, rivers, lakes, floodplains, canopies, soils, and aquifers. Changes in TWS are profoundly influenced by both natural and anthropogenic factors, including climate extremes (i.e., droughts, floods, and heat waves) as well as human activities such as groundwater abstraction (Liu et al., 2020) and land use changes (Shi et al., 2024). Recent trends in TWS (B. Li & Rodell, 2024; Liu et al., 2020; Rodell et al., 2018; Wang et al., 2018) have raised global concerns due to their implications for freshwater availability and the sustainability of water resources (Döll et al., 2016; Scanlon et al., 2023). Climate change is already threatening water

Visualization: Amar Deep Tiwari
Writing – original draft: Amar Deep Tiwari
Writing – review & editing: Amar Deep Tiwari, Yadu Pokhrel, Farshid Felfelani, Ahmed Elkouk, Julien Boulange, Simon N. Gosling, Naota Hanasaki, Aristeidis Koutroulis, Vimal Mishra, Hannes Müller Schmied, Yusuke Satoh, Sebastian Ostberg, Tobias Stacke, Jiabo Yin

availability in many populous regions of the globe (Konapala et al., 2020; X. Li et al., 2022; Wu et al., 2020; Q. Zhang et al., 2023). Consequently, accurate quantification and future projections of TWS are essential for effective and sustainable water resource management under changing climate.

The Gravity Recovery and Climate Experiment (GRACE) mission has unveiled Earth's changing freshwater landscape, which has profound implications for water and food security (Tapley et al., 2019). GRACE data sets have helped enhance our ability to monitor natural and anthropogenic variations in water availability (Ahmed et al., 2014; Bibi et al., 2021; H. Deng & Chen, 2017; S. Deng et al., 2022; Tapley et al., 2019; X. Yang et al., 2021). Moreover, GRACE products have proven valuable in capturing historical drought events (Houborg et al., 2012; Huang et al., 2018; Jin & Zhang, 2016; Khorrami et al., 2023; Schumacher et al., 2018; Sinha et al., 2017; Z. Zhang et al., 2015) and in predicting future droughts (Hasan et al., 2021; Kenea et al., 2020). Although the GRACE record is relatively short (monthly anomalies from April 2002 with some missing months and coarse spatial resolution), the data have provided important early confirmation of large-scale changes in the global hydrological cycle projected by global water models (GWMs) for the twenty-first century (IPCC, 2021).

To compensate for the spatial and temporal resolution limitations and shorter periods of GRACE data sets, GWMs have been utilized to simulate TWS changes worldwide (Chaudhari et al., 2019; Kenea et al., 2020; Yan et al., 2021; Yin et al., 2022). In parallel, several studies have also extended the GRACE data period itself using various methods (Humphrey & Gudmundsson, 2019; F. Li et al., 2021; Yin et al., 2023). However, future projections of water availability remain model-dependent and require systematic evaluation (Scanlon et al., 2018; Tiwari, Pokhrel, et al., 2025). Several studies have assessed the performance of simulated TWS against GRACE data. For instance, Vergnes and Decharme (2012) incorporated a groundwater model and evaluated its TWS simulation, while van Dijk et al. (2011) assessed the simulated TWS by the Australian Water Resources Assessment model for Australia. Seyoum and Milewski (2016) evaluated the long-term TWS simulated by an integrated hydrologic model using in situ data in the Northern High Plains. Further, Long et al. (2017) analyzed spatiotemporal variability in merged TWS changes using global hydrological models against GRACE products. Bibi et al. (2024) benchmarked the multimodel TWS seasonal cycle against GRACE observations over major global river basins. More recently, Tiwari, Pokhrel, et al. (2025) provided a global intercomparison of GWMs by quantifying mean TWS trends and systematic biases against GRACE. Despite these advances, most prior studies emphasize long-term trends and seasonality. Critically, no global evaluation has yet examined GWM performance under drought conditions, leaving an important scientific gap in understanding how well models capture extreme hydrological variability.

GWMs have been employed to project TWS droughts (Ahi & Cekim, 2021; Jia et al., 2020; Pokhrel et al., 2021; Y. Yang et al., 2023; Yin et al., 2022), however, there is limited understanding of how well these models can capture TWS droughts to monitor historical droughts, which hinders accurate future drought projections. Numerous studies evaluated the GWMs, but they have not comprehensively examined TWS droughts (Heinicke et al., 2024; Zaherpour et al., 2018). This study aims to fill these gaps by evaluating the capability of models to accurately represent TWS droughts. We answer the following science questions: (a) How accurately do GWMs simulate historical TWS droughts? and (b) What are the specific regions and climate zones where GWMs struggle to accurately simulate TWS droughts? By addressing these gaps, this study seeks to enhance our understanding of TWS dynamics and improve the accuracy of drought predictions, ultimately contributing to more sustainable water resource management strategies.

2. Materials and Methods

Monthly model TWS data are taken from the Inter-Sectoral Impact Model Intercomparison Project ISIMIP3a (Frieler et al., 2024) (https://www.isimip.org/protocol/#isimip3a). The eight GWMs included in this study are: CWatM (Burek et al., 2020), H08 (Hanasaki et al., 2008, 2018), HydroPy (Stacke & Hagemann, 2021), JULES-W2 (Best et al., 2011), LPjmL.5-7-10-fire (Oberhagemann et al., 2024; Wirth et al., 2024), MIROC-INTEG-LAND (Yokohata et al., 2020), ORCHIDEE-MICT (Guimberteau et al., 2018), and WaterGAP2.2e (Müller Schmied et al., 2021, Müller Schmied, Gosling, et al., 2024, Müller Schmied, Trautmann, et al., 2024). Each GWM is forced with daily (except MIROC-INTEG-LAND; three-hourly) inputs from the GSWP3-W5E5 historical climate data set (Lange et al., 2022; Tiwari, Pokhrel, et al., 2025). Different models consider human management to varying degrees. For instance, CWatM, H08, HydroPy, and WaterGAP2-2e models use groundwater withdrawal, CWatM, H08, LPJmL5-7-10-fire, and WaterGAP2-2e models use dams and reservoirs,

and all models account for land-use changes (Table S2 in Supporting Information S1). Details about how different models resolve TWS components are given in Table S3 in Supporting Information S1. More details regarding model parameterizations, human activities considered, and calibration can be found in Telteu et al. (2021) and Müller Schmied, Gosling, et al. (2024).

We compare the ISIMIP3a-based simulated TWS against GRACE TWS during drought conditions. We select GRACE mascon products from the Center for Space Research (CSR; Save et al., 2016), Jet Propulsion Laboratory (JPL; Watkins et al., 2015), and Goddard Space Flight Center (GSFC; Croteau et al., 2020). Mascon solutions enhance signal recovery by reducing leakage along coastlines between land and ocean regions and eliminating vertical striping found in non-regularized Level 2 spherical harmonic solutions (Croteau et al., 2020; Luthcke et al., 2013; Save et al., 2016; Watkins et al., 2015). GRACE TWS data from CSR (0.25°) are re-gridded to 0.5-degree spatial resolution using bi-linear interpolation to make it consistent with the simulations. Further, we use Singular Spectrum Analysis (Yi & Sneeuw, 2021) to fill the missing GRACE data including the gap between GRACE and GRACE Follow-On (GRACE-FO).

We calculate the uncertainty in the three GRACE products using the three-cornered hat (TCH) method (Gray & Allan, 1974; Premoli & Tavella, 1993) which calculates relative covariances without a priori knowledge when at least three data sets exist (Long et al., 2017; T. Xu et al., 2019; Q. Zhao et al., 2019). Widely used for TWS uncertainty estimation (Ferreira et al., 2016; Long et al., 2017; Yan et al., 2021), the TCH method supports our ensemble approach, minimizing processing errors by averaging three GRACE mascon products (Ali et al., 2022; Sakumura et al., 2014; Yan et al., 2021). While GRACE TWS data start from 2002, the ISIMIP simulations provide monthly data from 1901 to 2019. Thus, our analysis is limited to the overlapping period between the models and GRACE observations (2002–2019). To increase robustness and enable model evaluation over a longer period, we incorporate machine learning-based global TWS reconstruction products (GTWS-MLrec; Yin et al., 2023). We quantify the accuracy of GTWS-MLrec data sets for drought simulations by validating against observed standardized runoff index (SRI)-based droughts (Kumar et al., 2022) for eight diverse catchments previously studied by (Gosling et al., 2017) and Hattermann et al. (2017). These data sets provide long-term TWS records from 1940 to 2022 and show good performance in constraining global and regional water balance (e.g., Lu et al., 2025; Saemian et al., 2024).

TWS drought severity index (TWS-DSI hereafter; Zhao et al., 2017), has been employed to examine droughts and used widely in previous studies over different global regions (Kenea et al., 2020; Pokhrel et al., 2021; Satish Kumar et al., 2021; G. Xu et al., 2023; Yin et al., 2022). The time-period is consistently applied in simulations: 2002–2019 for comparisons with GRACE, and 1940–2019 for comparisons with GTWS-MLrec, while reference period to remove the mean from TWS to calculate TWS anomaly is same for GRACE, GTWS-MLrec, and models, which is 2004–2009. The TWS-DSI is calculated as follows.

$$\mathrm{TWS-DSI}_{i,j} = \frac{\mathrm{TWS}_{i,j} - \overline{\mathrm{TWS}_j}}{\sigma_j} \tag{1}$$

where $\mathrm{TWS}_{i,j}$ refers to the TWS anomalies at year i and month j, and $\overline{\mathrm{TWS}_j}$ and σ_j denote the mean value and standard deviation of TWS anomalies at month j, respectively. TWS-DSI values indicate a wet condition for above zero and dry condition for below zero values similar to US drought monitor (Svoboda et al., 2002).

We calculate the latitudinal droughts at every 0.5° in the TWS from −45° to 60-degree latitudes to isolate latitudinal variations. We evaluate TWS droughts by comparing the latitudinal means between the ISIMIP model ensemble median and the observational data from GRACE from 2002 to 2019 as well as Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) climate zones. For IPCC AR6 zones, we have taken the 12-month moving average and then removed the trends from TWS anomalies. We use a 12-month moving average to smooth short-term variability and highlight sustained water deficits, aligning with the seasonal to annual scale of hydrological droughts. We remove trends from both GRACE and ISIMIP TWS using two methods: linear detrending and the Empirical Mode Decomposition (EMD) method (Dong & Zhang, 2012). Furthermore, we compare TWS droughts between GRACE and the models under nonstationary conditions (without trend removal) and under approximately stationary conditions (after removing trends). Continent-wise analysis of TWS droughts is conducted considering both the ensemble median and individual models to

systematically investigate drought characteristics across and their contributions to global signal. Further, we evaluate the peak drought period at the continental scale in GRACE TWS and compare it with the model simulations.

We compute the drought severity and coverage index (DSCI), similar to that in the US drought monitor (Leelaruban et al., 2012), but for TWS droughts, further explained by Tiwari and Mishra (2022) as:

$$\mathrm{DSCI} = 1(\mathrm{D0}) + 2(\mathrm{D1}) + 3(\mathrm{D2}) + 4(\mathrm{D3}) + 5(\mathrm{D4}) \tag{2}$$

Here, D0, D1, D2, D3, and D4 represent the percentage area of abnormal drought, moderate drought, severe drought, extreme drought, and exceptional drought, respectively. DSCI ranges from 0 to 500, with a zero meaning that none of the area is under drought, and 500 meaning that entirety of the area is in D4 (i.e., exceptional TWS drought). We applied the DSCI metric to the Amazon and Congo basins due to their large size and strong TWS drought signals during the GRACE period (2002–2019). These basins offer a robust testbed for evaluating model performance in capturing extreme droughts.

Major river basin shapefiles (a total of 230 basins $\geq$50,000 km^2) are taken from the Global Runoff Data Center, which incorporates data from the HydroSHEDS database (Lehner & Grill, 2013). Further, we have analyzed the contributions of soil moisture and snow water equivalent (SWE) to variations in TWS, providing insights into regional TWS dynamics. GRACE data limitations, including coarse resolution, temporal gaps, and signal recovery uncertainties, have been addressed using an ensemble of three GRACE mascon products (CSR, JPL, GSFC), gap-filling via SSA, and re-gridding to 0.5° × 0.5° resolution. Uncertainty has been quantified using the TCH method, though limitations in GRACE—such as coarse spatial resolution, signal smoothing, and leakage errors—can dampen extreme anomalies and lead to underestimation of drought severity, especially in smaller basins. A detailed uncertainty analysis across mascon products has been conducted to evaluate spatial and temporal variability.

To evaluate the models' performance in simulating TWS droughts and identifying the underperforming models, we calculate the Critical Success Index (CSI; Wilks, 2011), a widely used metric in hydrological and climate studies (Tiwari et al., 2021).

$$\mathrm{CSI} = \frac{\mathrm{hit}}{\mathrm{hit} + \mathrm{miss} + \mathrm{false}} \tag{3}$$

Here “hit” means that TWS drought was observed and simulated on a given period, “miss” indicates that the observed TWS drought was not simulated. “False” refers to cases where the simulated drought did not occur in the observation. The threshold for TWS drought is taken as −1 value of TWS-DSI (Tiwari, Kushwaha, et al., 2025). To account for ensemble variability, we compute the mean and standard deviation of CSI values across all models for each basin. Models with CSI values lower than (mean −1 standard deviation) have been classified as substantially deviating models, allowing multiple models to be flagged when their performance was notably lower than the ensemble average.

Signal-to-Noise Ratio (SNR) is computed to assess the fidelity of simulated terrestrial TWS relative to the observed TWS from GRACE. SNR quantifies the proportion of the variance in the observed signal captured by the models while accounting for the residual errors (Gao et al., 2023). SNR values are computed for 230 basins across eight models, and higher SNR values indicate better model performance in capturing the observed TWS variability. Given a time-series consisting of GRACE-based TWS values (GTWSA) and corresponding simulated values (STWSA), the SNR is defined as:

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{\sigma_{\mathrm{O}}^2}{\sigma_{\varepsilon}^2}\right) \tag{4}$$

where σ_{O}^2 is the variance of the GTWSA values, and σ_{ε}^2 is the variance of the residuals ε = GTWSA-STWSA.

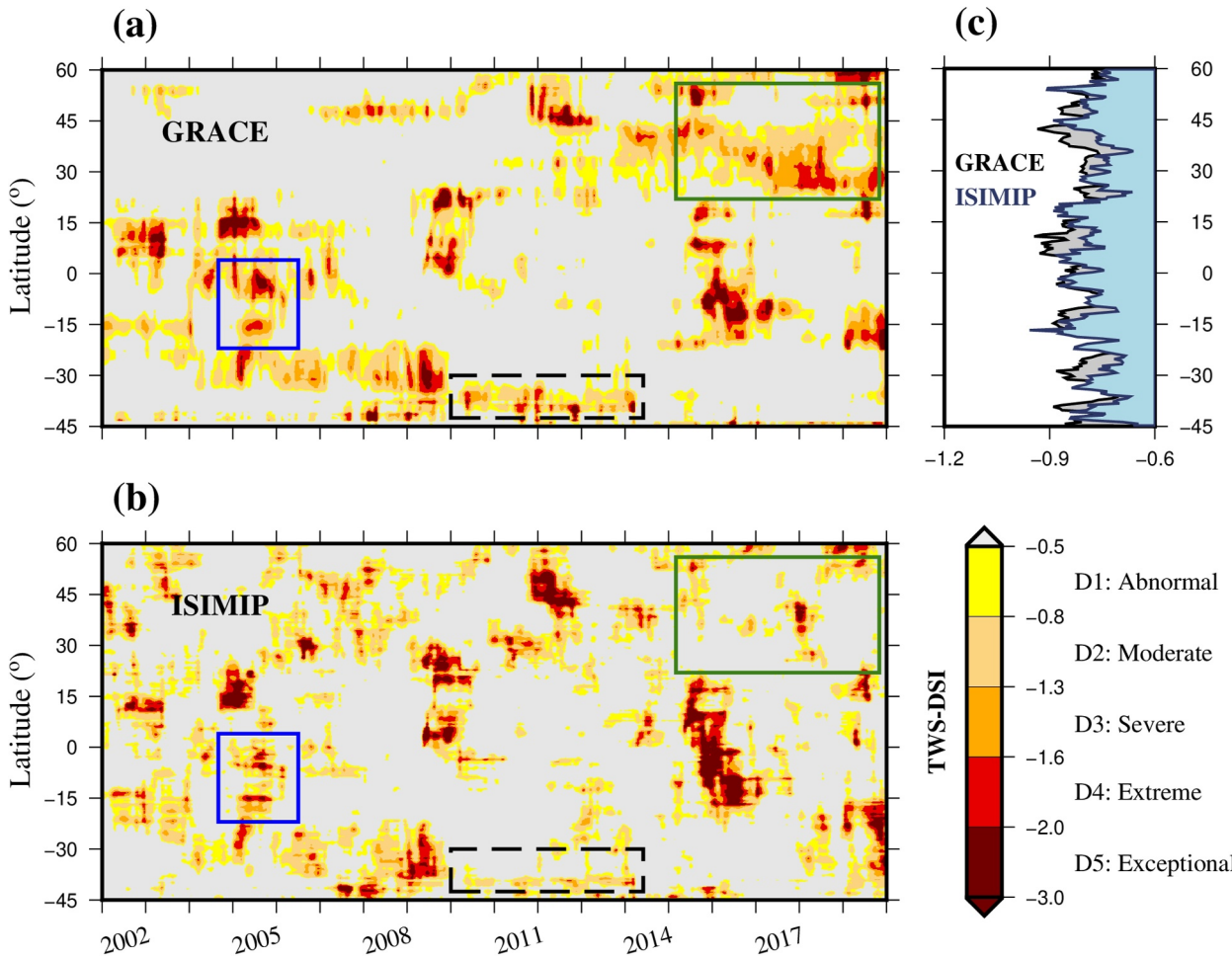

Figure 1. Historical terrestrial water storage latitudinal droughts from (a) Gravity Recovery and Climate Experiment (GRACE), (b) the median ensemble of eight ISIMIP models, and (c) a comparison of long-term (2002–2019) latitudinal simulated droughts against GRACE. Selected rectangles highlight major droughts underestimated in the simulations, with events detected in mid and lower latitudes during 2005–2006 (blue box), 2010–2014 (dashed black box), and 2015–2019 (green box).

3. Results

3.1. Global TWS Droughts and Model Evaluation

Using GRACE-based TWS, we detect the major drought events (Figure 1a) during 2005–2006, 2010–2014, and 2015–2019 on global latitudinal mean. The GRACE data (2002–2019) revealed distinct TWS drought patterns globally (Figure 1a). North America, especially the western U.S., underwent significant water deficits, peaking during the 2011–2013 drought with TWS-DSI dropping to extreme drought conditions (Figure S1 in Supporting Information S1). South America, Europe, Africa, Asia, and Australia also experienced severe negative TWS-DSI, underscoring a global trend of increasing water stress and potential escalation in water scarcity from 2010 onwards. Mid-latitude drought events (blue box in Figure 1a) are observed in South America, Africa, and Australia (Figure S1 in Supporting Information S1). Lower latitude drought events from 2010 to 2014 (black box in Figure 1a) are observed in South America (Figure S1 in Supporting Information S1), while 2015–2019 high-latitude drought events are observed in South America and Africa (Figure S1 in Supporting Information S1).

The comparative assessment of TWS-DSI between median ISIMIP and GRACE reveals that although the ISIMIP models capture the general spatial patterns of TWS-DSI, there are systematic underestimations of the severity and coverage of drought conditions, especially within tropical and high-latitude regions (Figures 1a and 1b). Underestimated drought conditions based on ISIMIP models in several regions are prominent in the GRACE data (Figures 1a and 1b). Though regions indicated by blue and green rectangles are observed in simulations, models

10.1029/2025GL115164

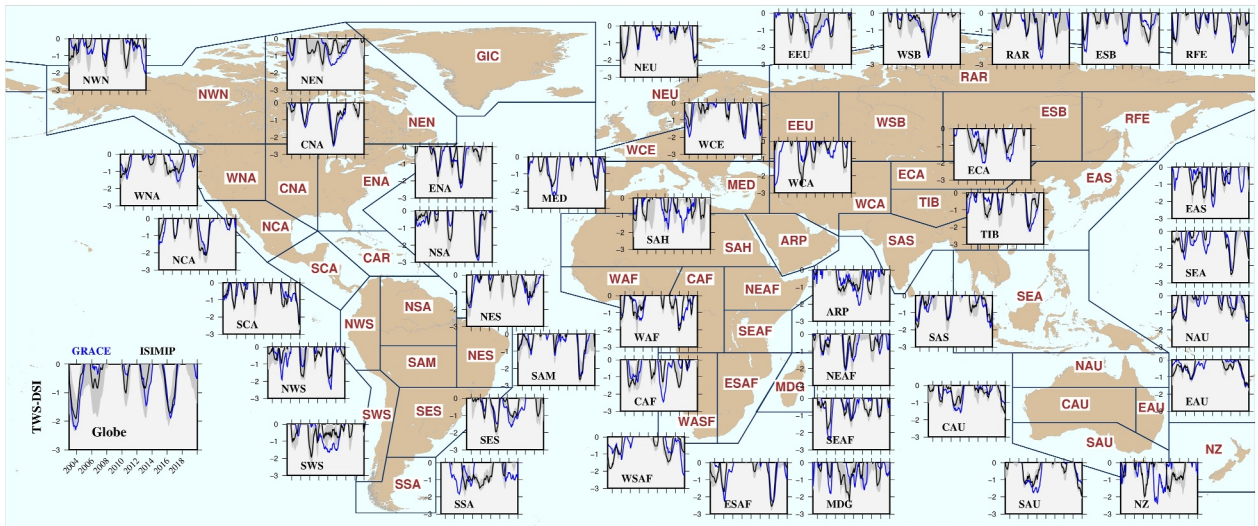

Figure 2. Comparison of simulated terrestrial water storage (TWS) drought against Gravity Recovery and Climate Experiment (GRACE) for Intergovernmental Panel on Climate Change AR6 climate regions. Blue and black lines represent the GRACE and model median TWS-DSI, respectively. The gray shaded area is the range of the TWS-DSI among models. Full names of climate zones are provided in Table S1 in Supporting Information S1.

particularly underestimate GRACE-detected droughts in regions indicated by black rectangles. These discrepancies are pronounced in the regions highlighted by the black and blue lines (Figure 1c), where the underestimation of drought conditions by the model simulations could significantly impact regional water management and climate adaptation strategies. We further compare the individual model simulated TWS droughts against observations (Figure S2 in Supporting Information S1). We find that all the models underestimate droughts except CWatM and WaterGAP2-2e, which better capture the spatial extent of these droughts. When TWS trends are preserved, GRACE TWS-based droughts are found to be severe in recent years, especially in the northern latitudes which are not captured by any model (Figure S2 in Supporting Information S1), indicating that declining TWS trends are not well represented in models. However, when trends are removed models tend to compare better with observations in TWS droughts (Figure S3 in Supporting Information S1).

3.2. TWS Droughts Comparison Based on Climate Regions and Continents

Figure 2 shows the comparative analysis of TWS droughts for different climate regions. In several climate regions such as CNA, ENA, NES, SAM, NEU, WCE, ESAF, WSB, NEAF, SAS, and NAU GWMs well capture TWS drought events from 2002 to 2019. However, in majority of the climate regions (e.g., NEN, WNA, NWS, MED, SSA, SAH, RAR, ESB, ECA, CAF, ARP, SAU, EAS), simulated droughts are underestimated compared to GRACE (Figure 2); the underestimation is pronounced in North America, South America, Africa, and Northern Asia (Russia). Among climate regions, underestimation of simulated drought, including NEN, SWS, SSA, SES, NES, CAF, SEAF, ARP, and NZ are prominent (Figure 2).

We calculate the continent-wise mean time series of TWS droughts and note the peak time of the droughts in each continent, to identify the most severe droughts across continent (Figure S4 in Supporting Information S1). For these peak drought conditions in each continent, we further compared how the TWS droughts from GRACE are captured in the simulated TWS droughts using drought severity and coverage index (DSCI) (Figure S4 in Supporting Information S1). We find that in each continent there is an underestimation in DSCI in simulated TWS droughts compared to the GRACE (Figure S5 in Supporting Information S1). Overall, the simulated TWS consistently underestimates the severity and spatial extent of TWS droughts compared to GRACE observations, highlighting the need for improved model parameterizations and validation.

3.3. TWS Droughts Comparison Based on Basins

The comparison of TWS drought in the Amazon River basin (the largest River basin by area) shows significant variation across the eight models analyzed, as well as in comparison to GRACE satellite observations, particularly for 2016 (Figure 3). Time series analysis from 2002 to 2019 highlights the varied ability of the models to capture

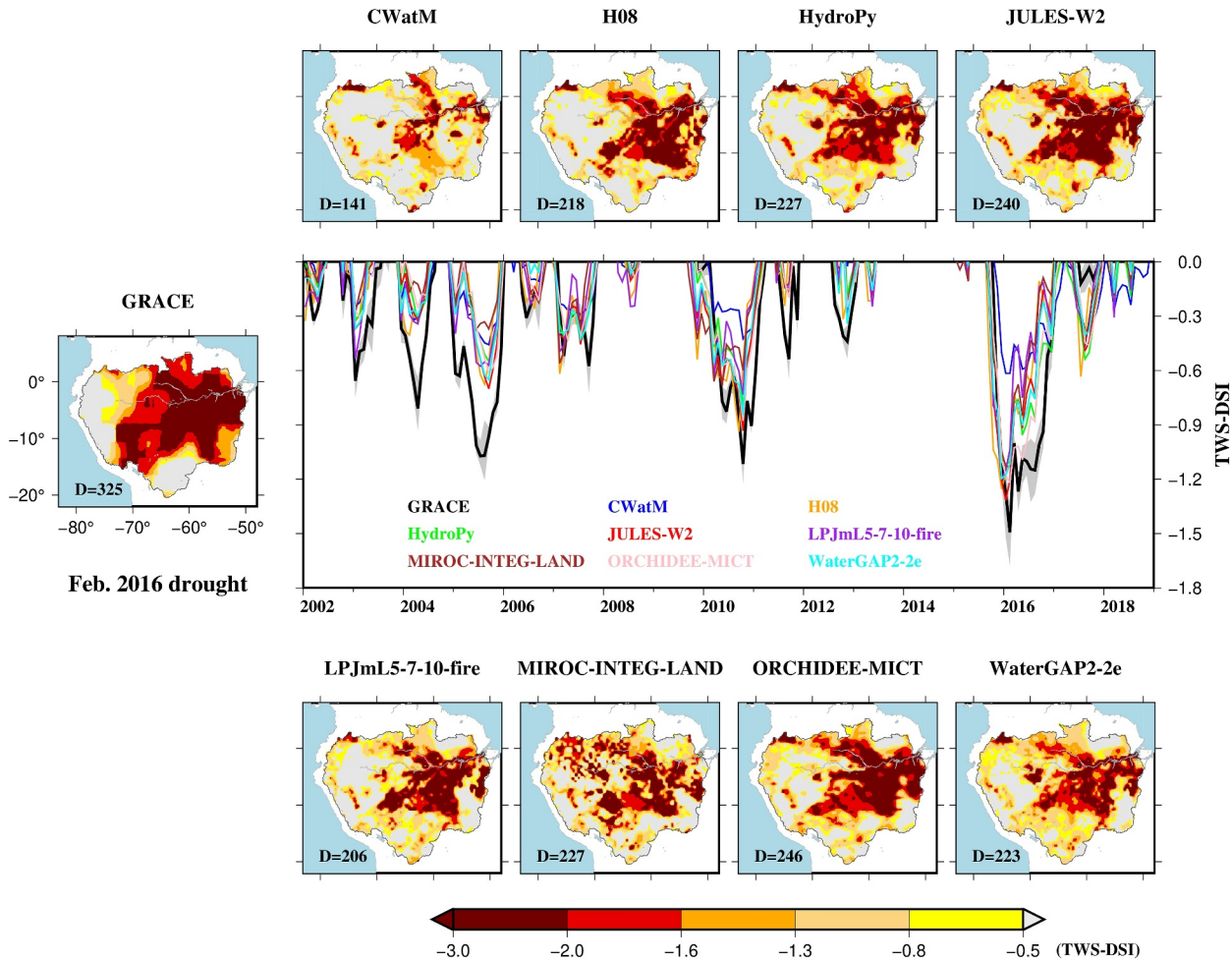

Figure 3. Comparison between GRACE-based and individual model-derived terrestrial water storage drought with the long-term time series (2002–2019) and spatial distribution of TWS-DSI across the Amazon River basin during the peak drought period of February 2016. Gray shaded area shows uncertainty in Gravity Recovery and Climate Experiment data. D values show the drought severity and coverage index.

the temporal dynamics of TWS droughts, with some models showing closer agreement with GRACE during major drought events, such as those in 2005 and 2016. The GRACE data reveals a DSCI value of 325, indicating a severe drought in February 2016. Among the models, ORCHIDEE-MICT and JULES-W2 show relatively closer DSCI of 246 and 240, respectively, suggesting better alignment with the GRACE observations. The improved performance of JULES-W2 and ORCHIDEE-MICT in the Amazon likely comes from their advanced representation of CO_2-sensitive vegetation dynamics, which are critical in high-biomass regions (Reinecke et al., 2025; Stacey et al., 2025; Wei et al., 2025). While other models also include some CO_2 effects, these two structurally capture vegetation–hydrology interactions more explicitly, possibly explaining their better performance. Other models, such as CWatM and LPJmL5-7-10-fire, exhibit lower DSCI of 141 and 206, respectively, indicating a tendency to underestimate the drought's severity (Figure 3). Similarly, we found the varied performance of models for TWS drought simulation of the Congo River basins as well where HydroPy model performs best and JULES-W2 and MIROC-INTEG-LAND underestimate the drought most for the peak drought of December 2005 (Figure S6 in Supporting Information S1). Moreover, unlike the Amazon River basin, in the Congo basin, all the models underestimate most of the peak TWS droughts (Figure S6 in Supporting Information S1). While comparing dry and wet periods both for selected River basins, we find that models generally exhibit lower TWS amplitude, which partially explains the underestimation of drought severity (Figure S7 in Supporting Information S1). This underestimation is more prominent during dry periods, as models tend to underestimate the negative extremes (i.e., droughts) more than the positive extremes (wet conditions).

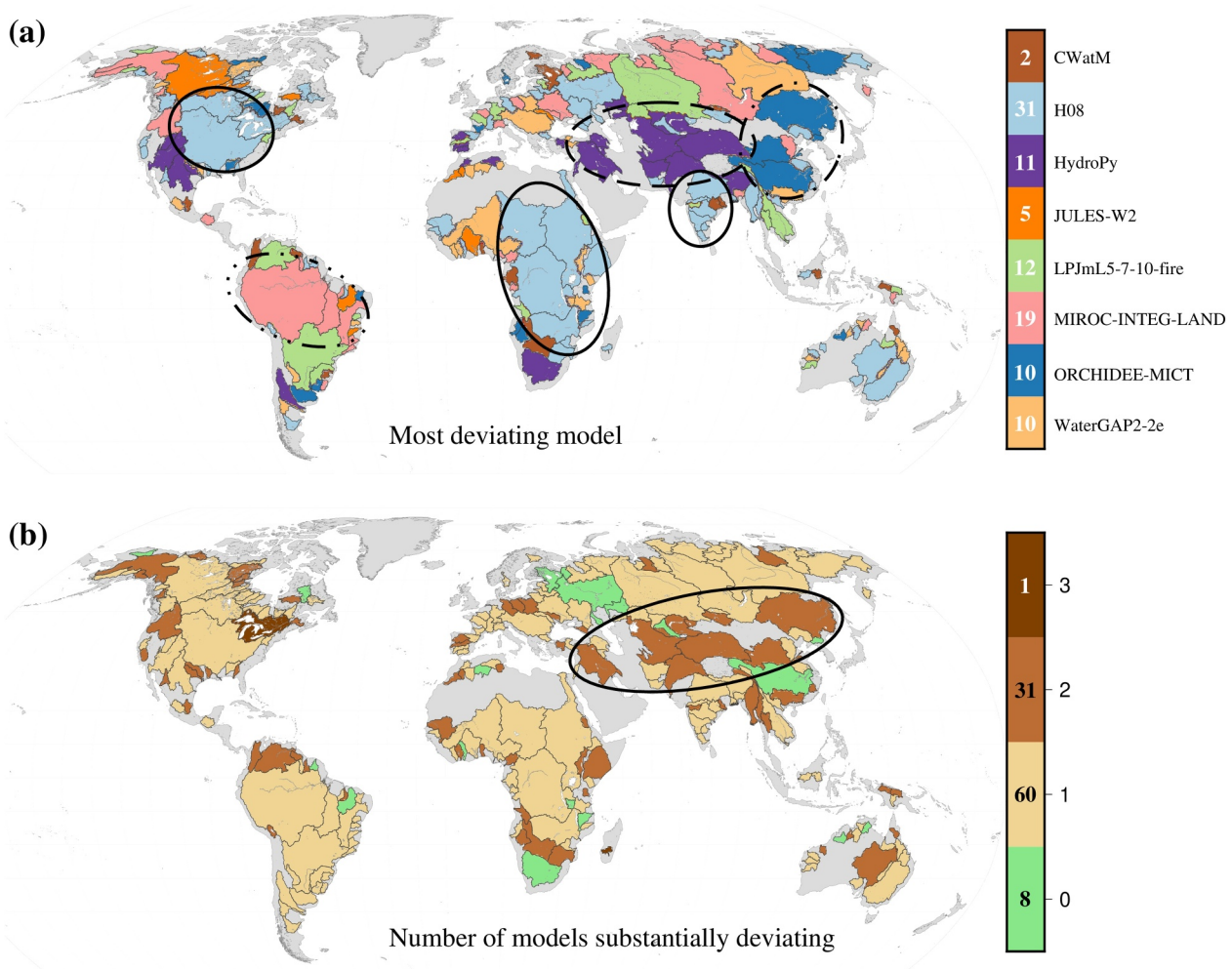

Figure 4. (a) Basin-wise most deviating model for terrestrial water storage (TWS) drought. The regions are specified where a particular model is performing worst using oval shapes. The numbers inside the color bar indicate the percentage of basins corresponding to the model. (b) Number of models substantially deviating from Gravity Recovery and Climate Experiment (GRACE) TWS. The numbers outside the color bar indicate the number of models substantially deviating while numbers inside the color bar indicate corresponding percentage of basins. Oval shape shows the region where 2 models substantially deviate from GRACE.

During the peak drought in February 2016, all models produce results that are closely aligned in terms of percentiles of TWS anomalies (Figure S8 in Supporting Information S1). The anomaly percentiles from the models matched well with the GRACE data. However, spatial comparisons reveal discrepancies between the models and GRACE TWS anomaly percentiles. These differences may be primarily attributed to soil moisture variability (Rodell & Famiglietti, 2001) as we find that soil moisture shows similar variability as TWS for Amazon River basin (Figure S9 in Supporting Information S1). Upon examining soil moisture, we found that it exhibited the highest correlation with TWS in more than 70% of the basins, with correlation coefficients exceeding 0.7 across the models (Figure S10 in Supporting Information S1).

We examine the contributions of individual TWS components. All eight models provide data on soil moisture and SWE, so our analysis focuses on these two variables. We find that soil moisture is the dominant component of TWS across nearly all regions (Figure S11a in Supporting Information S1). In polar regions, however, SWE also makes a substantial contribution to TWS (Figure S11b in Supporting Information S1). Given the dominant influence of soil moisture on TWS, any errors in soil moisture estimates are likely to be the primary cause of the underestimation of TWS droughts. Therefore, improving soil moisture simulations in global water models could be critical in reducing the underestimation of TWS droughts, especially in regions where soil moisture is the predominant component. Furthermore, in polar regions, where SWE also plays a crucial role, both soil moisture and SWE need to be accurately represented to improve the overall reliability of TWS drought assessments.

3.4. Model Benchmarking

Figure 4a shows, for each basin, the model that deviated most from GRACE, based on CSI, when simulating TWS drought, highlighting spatial variability in model performance. We find several clusters where individual models deviate most. The H08 model shows the highest deviation in 31% of basins, particularly in North America, Africa, and India, and basins in Southeast Asia, indicating challenges in capturing drought dynamics in these regions. In Northwest North America, South America, and North Asia, MIROC-INTEG-LAND exhibits significant deviations, reflecting difficulties in modeling complex hydrological processes. The HydroPy model is the most deviating in several central Asian and Middle Eastern basins, while ORCHIDEE-MICT frequently deviates in Eastern Asia, suggesting a need for region-specific model enhancements. These clusters can be used to identify the major issue in a particular model. We also identify most deviating models (See Methods) from GRACE for TWS droughts, finding that in majority of the basins, only one model deviates the most (Figure 4b). While in some basins (e.g., Orange and Yangtze basin) where none of the models exceed the substantial deviation limit, in other basins (e.g., Indus, Tarim, and Amur) multiple models deviate substantially from GRACE. Furthermore, we compare the model-simulated TWS against GRACE TWS using SNR (Figure S12 in Supporting Information S1). Basins with high human impacts, such as the Indus, Ganges, and Yellow, as well as snow-dominated basins, including the Yukon and Mackenzie, exhibited lower SNR values, indicating poorer model performance (Figure S12 in Supporting Information S1). Among the models, JULES-W2 demonstrated the best performance in terms of SNR, whereas H08 showed the lowest performance. Moreover, these findings emphasize the importance of improving model accuracy for better drought assessment and water resource management in diverse climatic and geographic regions.

3.5. Uncertainty in GRACE Observations and Comparison With Alternative GRACE Reconstructions

Analysis using the THC method reveals that JPL-M exhibits the highest average uncertainty among GRACE mascon products, while CSR-M shows the lowest, consistent with prior findings (Gao et al., 2023) (Figure S13 in Supporting Information S1). JPL-M's discontinuous grid block pattern, due to its 3° spherical cap mascon block, contrasts with its 0.5° × 0.5° resolution. Basin-averaged TWS-DSI uncertainties are lower in large basins (e.g., Amazon, Mississippi) but higher in smaller basins (e.g., Colorado, Orinoco) (Figure S14 in Supporting Information S1). Spatial averaging across mascon products reduces uncertainties, particularly during droughts, providing more robust estimates. Validation of GTWS-MLrec data sets in eight climatically diverse basins confirms their reliability, with TWS droughts aligning closely with runoff-based droughts (e.g., Upper Mississippi River basin), despite minor lags (Figure S15 in Supporting Information S1). This underscores GTWS-MLrec's effectiveness in reconstructing historical drought dynamics. TWS droughts (2002–2019) assessed against a 1940–2019 reference period show increased intensity due to recent declining TWS trends, exacerbated by long-term variability (Figure S16 in Supporting Information S1). However, GWMs does not fully capture this intensification, likely due to inadequate representation of human activities (e.g., groundwater extraction) and recent hydrological changes. Under nonstationary conditions, GTWS-MLrec shows more severe and extensive droughts, while ISIMIP models exhibit reduced severity (Figure S17a and S17b in Supporting Information S1). Removing trends linearly does not substantially affect global drought severity patterns (Figure S17a–S17d in Supporting Information S1). ISIMIP models consistently underestimate TWS drought severity relative to GTWS-MLrec in both stationary and nonstationary conditions. In the Amazon basin, where long-term trends are weak, linear trend removal has minimal effect on drought severity. However, using the EMD method brings ISIMIP-simulated droughts closer to those from GTWS-MLrec, although the models still underestimate severity (Figure S17e in Supporting Information S1). In contrast, in the heavily human-impacted Indus basin, retaining trends amplifies drought magnitude in GTWS-MLrec, while ISIMIP remains less sensitive. Although trend removal (both linear and EMD) improves agreement with GTWS-MLrec under stationary conditions, the models continue to underestimate drought severity (Figure S17f in Supporting Information S1). Therefore, negative long-term trends, particularly in recent years, can intensify drought signals and should be considered in model evaluations. While comparing the top drought events between the period of 1981–2000 for selected six River basins from each continent, we found that models mostly underestimate the observed TWS droughts (Figure S18 in Supporting Information S1).

4. Conclusions

This study provides the first comprehensive evaluation of TWS droughts simulated by GWMs using GRACE satellite data (2002–2019). Our analysis reveals a global trend toward increasing negative TWS-DSI, particularly post-2010, with peak severities occurring between 2015 and 2019. GWMs underestimate the severity and spatial extent of TWS droughts, particularly in tropical and high-latitude regions such as North America, South America, Africa, and Northern Asia. The comparison with GRACE satellite observations reveals significant discrepancies, with recent drought trends (2016–2019) in northern latitudes being poorly captured by the models. When using a longer reference period (i.e., GTWS-MLrec) the severity of recent droughts is found to have intensified due to declining TWS. These findings indicate that GWMs struggle to accurately capture historical TWS drought magnitudes, potentially underestimating future projections. Consistent underestimation of severity underscores gaps in parameterization, particularly in soil moisture dynamics, limiting model reliability for forecasting future drought scenarios.

The analysis reveals substantial variability in model performance across major river basins, with some models, like ORCHIDEE-MICT and HydroPy, better capturing the drought dynamics, while others, such as MIROC-INTEG-LAND and LPJmL5-7-10-fire, exhibit notable deficiencies. Recent years have seen profound drought events contributing to declining global TWS as observed by GRACE, which are not captured by any model, suggesting future projections of drought severity may be underestimated. This variability underscores the need for careful interpretation of multi-model ensemble results. To improve accuracy, future studies should use the ensemble median for balanced outputs or selectively combine reliable models tailored to specific regions. These approaches can enhance TWS simulation reliability, making multi-model ensembles more effective for addressing regional hydrological challenges and informing drought prediction and water resource management.

The underestimation of droughts in models stems from specific component-level limitations that are directly relevant for ISIMIP benchmarking and future model development. These include simplified vegetation dynamics, limited representation of human activities (e.g., irrigation, groundwater extraction), and incomplete groundwater-surface water interactions. Errors in soil parameterization and discrepancies in TWS spatial distribution highlight the need to improve lateral water fluxes, soil moisture dynamics, and groundwater processes (Clark et al., 2015; Maxwell et al., 2014). In polar regions, underrepresentation of snow processes and permafrost thaw affects TWS estimates (Elkouk et al., 2024; Lawrence et al., 2011), while groundwater pumping in regions like South Asia and the Central U.S. significantly influences TWS during droughts. Our benchmarking highlights the need to improve soil moisture, snow and permafrost, and human–water interactions, focusing on basins and regions where models underestimate droughts. Future work should focus on integrating detailed human activities (e.g., groundwater extraction, irrigation) to enhance model robustness (Döll, Fritsche et al., 2014; Döll, Müller Schmied et al., 2014; Wada et al., 2016). Furthermore, here we have used a single forcing (GSWP3-W5E5) based simulated TWS that is available for all models and that is bias adjusted against precipitation stations. Future research could explore the use of alternative forcing data sets to assess the sensitivity of model performance to variations in climate forcing, providing deeper insights into their robustness and reliability. Addressing these gaps is crucial for understanding drought processes and regional variability, ultimately supporting improved water management and policymaking to mitigate drought impacts.

Data Availability Statement

The simulated data sets used in this study are openly accessible at this URL/DOI: https://data.isimip.org/search/. Additionally, three openly available GRACE products utilized in this study can be obtained from the following sources: http://www2.csr.utexas.edu/grace/, https://podaac.jpl.nasa.gov/GRACE, and https://earth.gsfc.nasa.gov/geo/data/grace-mascons. We obtain GTWS-MLrec data sets from Yin et al. (2023). Scripts and the data files to reproduce the figures of this study are available in Zenodo from Tiwari (2025).

Acknowledgments
We thank ISIMIP for providing the data. Y.P. acknowledges support from the National Science Foundation (CAREER Award, Grant Number 1752729). A.D.T. and Y.P. acknowledge support from the National Science Foundation (BLUEGEM Award, Grant Number 2127643).

References

Ahi, G. O., & Cekim, H. O. (2021). Long-term temporal prediction of terrestrial water storage changes over global basins using GRACE and limited GRACE-FO data. *Acta Geodaetica et Geophysica*, *56*(2), 321–344. https://doi.org/10.1007/s40328-021-00338-4

Ahmed, M., Sultan, M., Wahr, J., & Yan, E. (2014). The use of GRACE data to monitor natural and anthropogenic induced variations in water availability across Africa. In *Earth-Science Reviews*. Elsevier. https://doi.org/10.1016/j.earscirev.2014.05.009

Ali, S., Wang, Q., Liu, D., Fu, Q., Mafuzur Rahaman, M., Abrar Faiz, M., & Jehanzeb Masud Cheema, M. (2022). Estimation of spatio-temporal groundwater storage variations in the Lower Transboundary Indus Basin using GRACE satellite. *Journal of Hydrology*, *605*, 127315. https://doi.org/10.1016/j.jhydrol.2021.127315

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., et al. (2011). The joint UK land environment simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Bibi, S., Song, Q., Zhang, Y., Liu, Y., Kamran, M. A., Sha, L., et al. (2021). Effects of climate change on terrestrial water storage and basin discharge in the lancang River Basin. *Journal of Hydrology: Regional Studies*, *37*, 100896. https://doi.org/10.1016/j.ejrh.2021.100896

Bibi, S., Zhu, T., Rateb, A., Scanlon, B. R., Kamran, M. A., Elnashar, A., et al. (2024). Benchmarking multimodel terrestrial water storage seasonal cycle against Gravity Recovery and Climate Experiment (GRACE) observations over major global river basins. *Hydrology and Earth System Sciences*, *28*(7), 1725–1750. https://doi.org/10.5194/hess-28-1725-2024

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., et al. (2020). Development of the Community Water Model (CWatM v1.04) - a high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development*, *13*(7), 3267–3298. https://doi.org/10.5194/gmd-13-3267-2020

Chaudhari, S., Pokhrel, Y., Moran, E., & Miguez-Macho, G. (2019). Multi-decadal hydrologic change and variability in the Amazon River basin: Understanding terrestrial water storage variations and drought characteristics. *Hydrology and Earth System Sciences*, *23*(7), 2841–2862. https://doi.org/10.5194/hess-23-2841-2019

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth system models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015WR017096

Croteau, M. J., Nerem, R. S., Loomis, B. D., & Sabaka, T. J. (2020). Development of a daily GRACE mascon solution for terrestrial water storage. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018468. https://doi.org/10.1029/2019JB018468

Deng, H., & Chen, Y. (2017). Influences of recent climate change and human activities on water storage variations in Central Asia. *Journal of Hydrology*, *544*, 46–57. https://doi.org/10.1016/j.jhydrol.2016.11.006

Deng, S., Zhang, M., Hou, Y., Wang, H., Yu, E., & Xu, Y. (2022). Terrestrial water storage dynamics: Different roles of climate variability, vegetation change, and human activities across climate zones in China. *Forests*, *13*(10), 1541. https://doi.org/10.3390/f13101541

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., & Wada, Y. (2016). Modelling freshwater resources at the global scale: Challenges and prospects. In *Surveys in Geophysics*. Springer. https://doi.org/10.1007/s10712-015-9343-1

Döll, P., Fritsche, M., Eicker, A., & Müller Schmied, H. (2014). Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations. *Surveys in Geophysics*, *35*(6), 1311–1331. https://doi.org/10.1007/s10712-014-9282-2

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, *50*(7), 5698–5720. https://doi.org/10.1002/2014WR015595

Dong, Y., & Zhang, J. (2012). Trend extraction in vibration signal based on EMD. In *Advanced Materials Research* (Vol. 459, pp. 377–380). https://doi.org/10.4028/www.scientific.net/AMR.459.377

Elkouk, A., Pokhrel, Y., Livneh, B., Payton, E., Luo, L., Cheng, Y., et al. (2024). Toward understanding parametric controls on runoff sensitivity to climate in the community land model: A case study over the Colorado river headwaters. *Water Resources Research*, *60*(12), e2024WR037718. https://doi.org/10.1029/2024WR037718

Ferreira, V. G., Montecino, H. D. C., Yakubu, C. I., & Heck, B. (2016). Uncertainties of the Gravity Recovery and Climate Experiment time-variable gravity-field solutions based on three-cornered hat method. *Journal of Applied Remote Sensing*, *10*(1), 015015. https://doi.org/10.1117/1.jrs.10.015015

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., Del Rocío Rivas López, M., et al. (2024). Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3a). *Geoscientific Model Development*, *17*(1), 1–51. https://doi.org/10.5194/gmd-17-1-2024

Gao, S., Hao, W., Fan, Y., Li, F., & Wang, J. (2023). A multi-source GRACE fusion solution via uncertainty quantification of GRACE-derived terrestrial water storage (TWS) change. *Journal of Geophysical Research: Solid Earth*, *128*(11), e2023JB026908. https://doi.org/10.1029/2023jb026908

Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., et al. (2017). A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3°C. *Climatic Change*, *141*(3), 577–595. https://doi.org/10.1007/s10584-016-1773-3

Gray, J. E., & Allan, D. W. (1974). A method for estimating the frequency stability of an individual oscillator. In *28th annual symposium on frequency control* (pp. 243–246). IEEE.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-N d lec, S., et al. (2018). ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: Model description and validation. *Geoscientific Model Development*, *11*(1), 121–163. https://doi.org/10.5194/gmd-11-121-2018

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., et al. (2008). Hydrology and Earth System Sciences an integrated model for the assessment of global water resources-Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, *12*(4), 1007–1025. https://doi.org/10.5194/hess-12-1007-2008

Hanasaki, N., Yoshikawa, S., Pokhrel, Y., & Kanae, S. (2018). A global hydrological simulation to specify the sources of water used by humans. *Hydrology and Earth System Sciences*, *22*(1), 789–817. https://doi.org/10.5194/hess-22-789-2018

Hasan, E., Tarhule, A., & Kirstetter, P. E. (2021). Twentieth and twenty-first century water storage changes in the nile river basin from grace/grace-fo and modeling. *Remote Sensing*, *13*(5), 1–30. https://doi.org/10.3390/rs13050953

Hattermann, F. F., Krysanova, V., Gosling, S. N., Dankers, R., Daggupati, P., Donnelly, C., et al. (2017). Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Climatic Change*, *141*(3), 561–576. https://doi.org/10.1007/s10584-016-1829-4

Heinicke, S., Volkholz, J., Schewe, J., Gosling, S. N., Müller Schmied, H., Zimmermann, S., et al. (2024). Global hydrological models continue to overestimate river discharge. *Environmental Research Letters*, *19*(7), 074005. https://doi.org/10.1088/1748-9326/ad52b0

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7), W07525. https://doi.org/10.1029/2011WR011291

Huang, R., Huang, J., Zhang, C., Zhuo, W., & Zhu, D. (2018). Drought monitoring over the Northeast China Using GRACE satellite data from 2002 to 2016. In *2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics)* (pp. 1–5). https://doi.org/10.1109/Agro-Geoinformatics.2018.8476013

Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: A reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data*, *11*(3), 1153–1170. https://doi.org/10.5194/essd-11-1153-2019

IPCC. (2021). Climate change 2021: The Physical science basis. Contribution of working group I to the Sixth assessment Report of the Intergovernmental Panel on climate change.

Jia, B., Cai, X., Zhao, F., Liu, J., Chen, S., Luo, X., et al. (2020). Potential future changes of terrestrial water storage based on climate projections by ensemble model simulations. *Advances in Water Resources*, *142*, 103635. https://doi.org/10.1016/j.advwatres.2020.103635

Jin, S., & Zhang, T. (2016). Terrestrial water storage anomalies associated with drought in Southwestern USA from GPS observations. In *Surveys in Geophysics*. Springer. https://doi.org/10.1007/s10712-016-9385-z

Kenea, T. T., Kusche, J., Kebede, S., & Güntner, A. (2020). Forecasting terrestrial water storage for drought management in Ethiopia. *Hydrological Sciences Journal*, *65*(13), 2210–2223. https://doi.org/10.1080/02626667.2020.1790564

Khorrami, B., Ali, S., Abadi, L. H., & Jehanzaib, M. (2023). Spatio-temporal variations in characteristics of terrestrial water storage and associated drought over different geographic regions of Türkiye. *Earth Science Informatics*, *16*(1), 717–731. https://doi.org/10.1007/s12145-022-00907-3

Konapala, G., Mishra, A. K., Wada, Y., & Mann, M. E. (2020). Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications*, *11*(1), 3044. https://doi.org/10.1038/s41467-020-16757-w

Kumar, A., Gosling, S. N., Johnson, M. F., Jones, M. D., Zaherpour, J., Kumar, R., et al. (2022). Multi-model evaluation of catchment- and global-scale hydrological model simulations of drought characteristics across eight large river catchments. *Advances in Water Resources*, *165*, 104212. https://doi.org/10.1016/j.advwatres.2022.104212

Lange, S., Mengel, M., Treu, S., & Büchner, M. (2022). ISIMIP3a atmospheric climate input data (v1.1). https://doi.org/10.48364/ISIMIP.982724.2

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the Community land model. *Journal of Advances in Modeling Earth Systems*, *3*(3), M03001. https://doi.org/10.1029/2011ms000045

Leelaruban, N., Oduor, P. G., Akyuz, A., Shaik, S., & Padmanabhan, G. (2012). Leveraging a spatio-temporal drought severity and coverage index with crop yield modelled as a stochastic process. *International Journal of Horticultural Science and Technology*, *2*(3), 219–236. https://doi.org/10.1504/IJHST.2012.049184

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Li, B., & Rodell, M. (2024). Terrestrial water storage in 2023. In *Nature Reviews Earth and Environment*. Springer Nature. https://doi.org/10.1038/s43017-024-00545-x

Li, F., Kusche, J., Chao, N., Wang, Z., & Löcher, A. (2021). Long-term (1979-present) total water storage anomalies over the global land derived by reconstructing GRACE data. *Geophysical Research Letters*, *48*(e2021GL093492), e2021GL093492. https://doi.org/10.1029/2021GL093492

Li, X., Long, D., Scanlon, B. R., Mann, M. E., Li, X., Tian, F., et al. (2022). Climate change threatens terrestrial water storage over the Tibetan Plateau. *Nature Climate Change*, *12*(9), 801–807. https://doi.org/10.1038/s41558-022-01443-0

Liu, X., Feng, X., Ciais, P., & Fu, B. (2020). Widespread decline in terrestrial water storage and its link to teleconnections across Asia and eastern Europe. *Hydrology and Earth System Sciences*, *24*(7), 3663–3676. https://doi.org/10.5194/hess-24-3663-2020

Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., et al. (2017). Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models. *Remote Sensing of Environment*, *192*, 198–216. https://doi.org/10.1016/j.rse.2017.02.011

Lu, J., Kong, D., Zhang, Y., Xie, Y., Gu, X., & Gulakhmadov, A. (2025). Hotspots of global water resource changes and their causes. *Earth's Future*, *13*(2), e2024EF005461. https://doi.org/10.1029/2024EF005461

Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J., & Camp, J. (2013). Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology*, *59*(216), 613–631. https://doi.org/10.3189/2013JoG12J147

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J. O., Ferguson, I. M., Ivanov, V., et al. (2014). Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *50*(2), 1531–1549. https://doi.org/10.1002/2013WR013725

Müller Schmied, H., Caceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., et al. (2021). The global water resources and use model WaterGAP v2.2d: Model description and evaluation. *Geoscientific Model Development*, *14*(2), 1037–1079. https://doi.org/10.5194/gmd-14-1037-2021

Müller Schmied, H., Gosling, S. N., Garnsworthy, M., Müller, L., Telteu, C.-E., Ahmed, A. K., et al. (2024). Graphical representation of global water models [Preprint]. *EGUsphere*. https://doi.org/10.5194/egusphere-2024-1303

Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., et al. (2024). The global water resources and use model WaterGAP v2.2e: Description and evaluation of modifications and new features. *Geoscientific Model Development*, *17*(23), 8817–8852. https://doi.org/10.5194/gmd-17-8817-2024

Oberhagemann, L., Billing, M., Von Bloh, W., Drüke, M., Forrest, M., Bowring, S. P. K., et al. (2024). Sources of uncertainty in the global fire model SPITFIRE: Development of LPJmL-SPITFIRE1.9 and directions for future improvements [Preprint]. *Geoscientific Model Development*. https://doi.org/10.5194/egusphere-2024-1914

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., et al. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, *11*(3), 226–233. https://doi.org/10.1038/s41558-020-00972-w

Premoli, A., & Tavella, P. (1993). A revisited three-comered hat method for estimating frequency standard instability. *IEEE Transactions on Instrumentation and Measurement*, *42*(1), 7–13. https://doi.org/10.1109/19.206671

Reinecke, R., Stein, L., Gnann, S., Andersson, J. C. M., Arheimer, B., Bierkens, M., et al. (2025). Uncertainties as a guide for global water model advancement. In *Wiley Interdisciplinary Reviews: Water*. John Wiley and Sons Inc. https://doi.org/10.1002/wat2.70025

Rodell, M., & Famiglietti, J. S. (2001). An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, *37*(5), 1327–1339. https://doi.org/10.1029/2000WR900306

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Saemian, P., Tourian, M. J., Elmi, O., Sneeuw, N., & AghaKouchak, A. (2024). A probabilistic approach to characterizing drought using satellite gravimetry. *Water Resources Research*, *60*(8), e2023WR036873. https://doi.org/10.1029/2023WR036873

Sakumura, C., Bettadpur, S., & Bruinsma, S. (2014). Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, *41*(5), 1389–1397. https://doi.org/10.1002/2013GL058632

Satish Kumar, K., Venkata Rathnam, E., & Sridhar, V. (2021). Tracking seasonal and monthly drought with GRACE-based terrestrial water storage assessments over major river basins in South India. *Science of the Total Environment*, *763*, 142994. https://doi.org/10.1016/j.scitotenv.2020.142994

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547–7569. https://doi.org/10.1002/2016JB013007

Scanlon, B. R., Fakhreddine, S., Rateb, A., de Graaf, I., Famiglietti, J., Gleeson, T., et al. (2023). Global water resources and the role of groundwater in a resilient water future. In *Nature Reviews Earth and Environment*. Springer Nature. https://doi.org/10.1038/s43017-022-00378-6

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., Van Beek, L. P. H., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Schumacher, M., Forootan, E., van Dijk, A. I. J. M., Müller Schmied, H., Crosbie, R. S., Kusche, J., & Döll, P. (2018). Improving drought simulations within the Murray-Darling basin by combined calibration/assimilation of GRACE data into the WaterGAP global hydrology model. *Remote Sensing of Environment*, *204*, 212–228. https://doi.org/10.1016/j.rse.2017.10.029

Seyoum, W. M., & Milewski, A. M. (2016). Monitoring and comparison of terrestrial water storage changes in the northern high plains using GRACE and in-situ based integrated hydrologic model estimates. *Advances in Water Resources*, *94*, 31–44. https://doi.org/10.1016/j.advwatres.2016.04.014

Shi, X., Wang, Y., Mao, J., Thornton, P. E., Riccuito, D. M., Hoffman, F. M., & Hao, Y. (2024). Quantifying the long-term changes of terrestrial water storage and their driving factors. *Journal of Hydrology*, *635*, 131096. https://doi.org/10.1016/j.jhydrol.2024.131096

Sinha, D., Syed, T. H., Famiglietti, J. S., Reager, J. T., & Thomas, R. C. (2017). Characterizing drought in India using GRACE observations of terrestrial water storage deficit. *Journal of Hydrometeorology*, *18*(2), 381–396. https://doi.org/10.1175/JHM-D-16-0047.s1

Stacey, J., Betts, R., Hartley, A., Mercado, L., & Gedney, N. (2025). Future global water scarcity partially alleviated by vegetation responses to atmospheric CO2 and climate change [Preprint]. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/egusphere-2025-51

Stacke, T., & Hagemann, S. (2021). HydroPy (v1.0): A new global hydrology model written in Python. *Geoscientific Model Development*, *14*(12), 7795–7816. https://doi.org/10.5194/gmd-14-7795-2021

Svoboda, M., Lecomte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., et al. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181–1190. https://doi.org/10.1175/1520-0477-83.8.1181

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, *44*(2), W02433. https://doi.org/10.1029/2006WR005779

Tang, Q., Gao, H., Yeh, P., Oki, T., Su, F., & Lettenmaier, D. P. (2010). Dynamics of terrestrial water storage change from satellite and surface observations and modeling. *Journal of Hydrometeorology*, *11*(1), 156–170. https://doi.org/10.1175/2009JHM1152.1

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., et al. (2019). Contributions of GRACE to understanding climate change. In *Nature Climate Change*. Nature Publishing Group. https://doi.org/10.1038/s41558-019-0456-2

Telteu, C. E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., et al. (2021). Understanding each other's models an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. In *Geoscientific model Development*. Copernicus GmbH. https://doi.org/10.5194/gmd-14-3843-2021

Tiwari, A. D. (2025). Data supporting Tiwari et al. (2025) "Underestimation of historical terrestrial water storage droughts in global water models" [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.17113661

Tiwari, A. D., Kushwaha, A. P., Sahai, A. K., & Mishra, V. (2025). Development of multimodel-based hydrologic outlook for India. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-23-0186.1

Tiwari, A. D., & Mishra, V. (2022). Sub-seasonal prediction of drought and Streamflow anomalies for water management in India. *Journal of Geophysical Research: Atmospheres*, *127*(3), e2021JD035737. https://doi.org/10.1029/2021JD035737

Tiwari, A. D., Mukhopadhyay, P., & Mishra, V. (2021). Influence of bias correction of meteorological and Streamflow Forecast on hydrological prediction in India. *Journal of Hydrometeorology*, *23*(7), 1171–1192. https://doi.org/10.1175/JHM-D-20

Tiwari, A. D., Pokhrel, Y., Boulange, J., Burek, P., Guillaumot, L., Gosling, S. N., et al. (2025). Similarities and divergent patterns in hydrologic fluxes and storages simulated by global water models. *Nature Water*, *3*(5), 550–560. https://doi.org/10.1038/s44221-025-00435-6

Van Dijk, A. I. J. M., Renzullo, L. J., & Rodell, M. (2011). Use of Gravity Recovery and Climate Experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system. *Water Resources Research*, *47*(11), W11524. https://doi.org/10.1029/2011WR010714

Vergnes, J. P., & Decharme, B. (2012). A simple groundwater scheme in the TRIP river routing model: Global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges. *Hydrology and Earth System Sciences*, *16*(10), 3889–3908. https://doi.org/10.5194/hess-16-3889-2012

Wada, Y., Flörke, M., Hanasaki, N., Eisner, S., Fischer, G., Tramberend, S., et al. (2016). Modeling global water use for the 21st century: The Water Futures and Solutions (WFaS) initiative and its approaches. *Geoscientific Model Development*, *9*(1), 175–222. https://doi.org/10.5194/gmd-9-175-2016

Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., et al. (2018). Recent global decline in endorheic basin water storages. *Nature Geoscience*, *11*(12), 926–932. https://doi.org/10.1038/s41561-018-0265-7

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2648–2671. https://doi.org/10.1002/2014JB011547

Wei, H., Zhang, Y., Huang, Q., Liu, C., & Wagener, T. (2025). Functional relationships reveal large differences in Streamflow Response to eCO2-vegetation in global water models. *Geophysical Research Letters*, *52*(9), e2024GL113685. https://doi.org/10.1029/2024GL113685

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences. Meteorological Applications*. Academic Press. https://doi.org/10.1002/met.16

Wirth, S. B., Braun, J., Heinke, J., Ostberg, S., Rolinski, S., Schaphoff, S., et al. (2024). Biological nitrogen fixation of natural and agricultural vegetation simulated with LPJmL 5.7.9 [Preprint]. EGUsphere. https://doi.org/10.5194/egusphere-2023-2946

Wu, W. Y., Lo, M. H., Wada, Y., Famiglietti, J. S., Reager, J. T., Yeh, P. J. F., et al. (2020). Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. *Nature Communications*, *11*(1), 3710. https://doi.org/10.1038/s41467-020-17581-y

Xu, G., Wu, Y., Liu, S., Cheng, S., Zhang, Y., Pan, Y., et al. (2023). How 2022 extreme drought influences the spatiotemporal variations of terrestrial water storage in the Yangtze River Catchment: Insights from GRACE-based drought severity index and in-situ measurements. *Journal of Hydrology*, *626*, 130245. https://doi.org/10.1016/j.jhydrol.2023.130245

Xu, T., Guo, Z., Xia, Y., Ferreira, V. G., Liu, S., Wang, K., et al. (2019). Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. *Journal of Hydrology*, *578*, 124105. https://doi.org/10.1016/j.jhydrol.2019.124105

Yan, X., Zhang, B., Yao, Y., Yang, Y., Li, J., & Ran, Q. (2021). GRACE and land surface models reveal severe drought in eastern China in 2019. *Journal of Hydrology*, *601*, 126640. https://doi.org/10.1016/j.jhydrol.2021.126640

Yang, X., Wang, N., Liang, Q., Chen, A., & Wu, Y. (2021). Impacts of human activities on the variations in terrestrial water storage of the aral sea basin. *Remote Sensing*, *13*(15), 2923. https://doi.org/10.3390/rs13152923

Yang, Y., Yin, J., Guo, S., Gu, L., He, S., & Wang, J. (2023). Projection of terrestrial drought evolution and its eco-hydrological effects in China. *Kexue Tongbao/Chinese Science Bulletin*, *68*(7), 817–829. https://doi.org/10.1360/TB-2022-0566

Yi, S., & Sneeuw, N. (2021). Filling the data gaps within GRACE missions using singular Spectrum analysis. *Journal of Geophysical Research: Solid Earth*, *126*(5), e2020JB021227. https://doi.org/10.1029/2020JB021227

Yin, J., Guo, S., Yang, Y., Chen, J., Gu, L., Wang, J., et al. (2022). Projection of droughts and their socioeconomic exposures based on terrestrial water storage anomaly over China. *Science China Earth Sciences*, *65*(9), 1772–1787. https://doi.org/10.1007/s11430-021-9927-x

Yin, J., Slater, L. J., Khouakhi, A., Yu, L., Liu, P., Li, F., et al. (2023). GTWS-MLrec: Global terrestrial water storage reconstruction by machine learning from 1940 to present. *Earth System Science Data*, *15*(12), 5597–5615. https://doi.org/10.5194/essd-15-5597-2023

Yokohata, T., Kinoshita, T., Sakurai, G., Pokhrel, Y., Ito, A., Okada, M., et al. (2020). MIROC-INTEG-LAND version 1: A global biogeochemical land surface model with human water management, crop growth, and land-use change. *Geoscientific Model Development*, *13*(10), 4713–4747. https://doi.org/10.5194/gmd-13-4713-2020

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Dankers, R., et al. (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters*, *13*(6), 065015. https://doi.org/10.1088/1748-9326/aac547

Zhang, Q., Shen, Z., Pokhrel, Y., Farinotti, D., Singh, V. P., Xu, C. Y., et al. (2023). Oceanic climate changes threaten the sustainability of Asia's water tower. *Nature*, *615*(7950), 87–93. https://doi.org/10.1038/s41586-022-05643-8

Zhang, Z., Chao, B. F., Chen, J., & Wilson, C. R. (2015). Terrestrial water storage anomalies of yangtze river basin droughts observed by GRACE and connections with ENSO. *Global and Planetary Change*, *126*, 35–45. https://doi.org/10.1016/j.gloplacha.2015.01.002

Zhao, M., Geruo, A., Velicogna, I., & Kimball, J. S. (2017). Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes. *Journal of Climate*, *30*(16), 6297–6308. https://doi.org/10.1175/JCLI-D-16-0458.s1

Zhao, Q., Zhang, B., Yao, Y., Wu, W., Meng, G., & Chen, Q. (2019). Geodetic and hydrological measurements reveal the recent acceleration of groundwater depletion in North China Plain. *Journal of Hydrology*, *575*, 1065–1072. https://doi.org/10.1016/j.jhydrol.2019.06.016